Decision Theory

João Mota

UDRC-EURASIP Summer School 2022

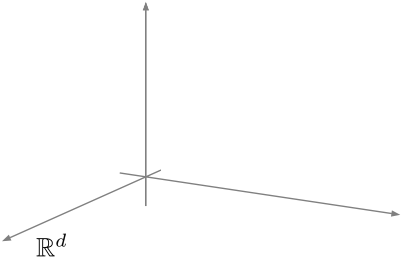
Heriot-Watt University

Decision theory

Decision theory

Problem

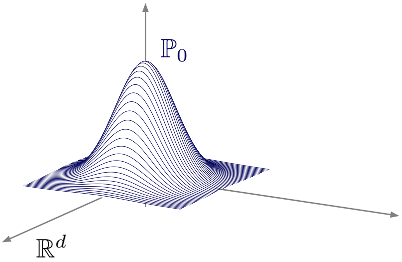
Decision theory



Problem

We observe $\boldsymbol{X} \in \mathbb{R}^d$

Decision theory

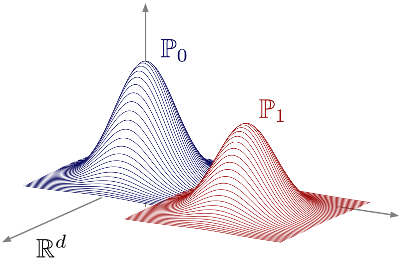


Problem

We observe $\boldsymbol{X} \in \mathbb{R}^d$

$\boldsymbol{X} \sim \mathbb{P}_0$

Decision theory

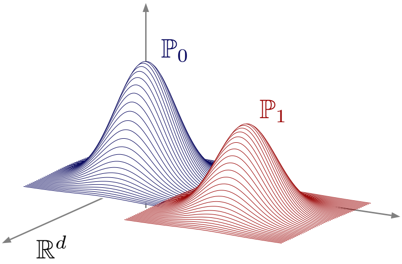


Problem

We observe $\boldsymbol{X} \in \mathbb{R}^d$

$\boldsymbol{X} \sim \mathbb{P}_0$ or $\boldsymbol{X} \sim \mathbb{P}_1$?

Decision theory



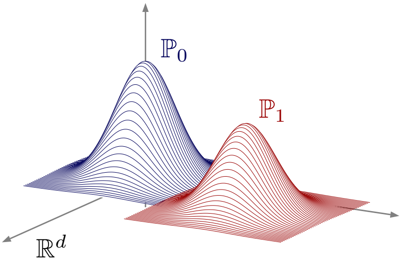
Problem

We observe $\boldsymbol{X} \in \mathbb{R}^d$

$\boldsymbol{X} \sim \mathbb{P}_0$ or $\boldsymbol{X} \sim \mathbb{P}_1$?

In classical *decision theory*, we *know* the distributions $\mathbb{P}_0$ and $\mathbb{P}_1$

Decision theory



Problem

We observe $\boldsymbol{X} \in \mathbb{R}^d$

$\boldsymbol{X} \sim \mathbb{P}_0$ or $\boldsymbol{X} \sim \mathbb{P}_1$?

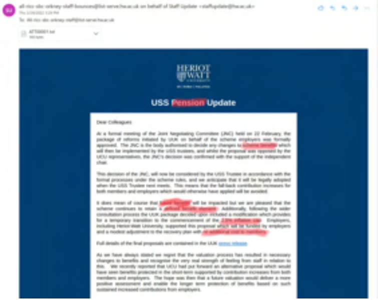
In classical *decision theory*, we *know* the distributions $\mathbb{P}_0$ and $\mathbb{P}_1$

In *machine learning*, we *have to estimate* $\mathbb{P}_0$ and $\mathbb{P}_1$ from data

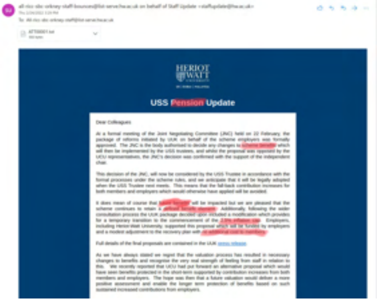
Example in 1D: Spam

Example in 1D: Spam

Example in 1D: Spam

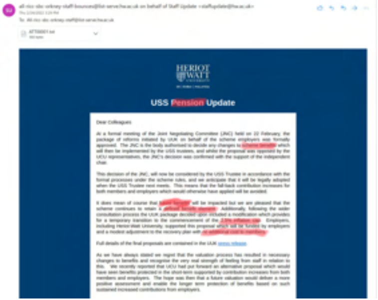


Example in 1D: Spam

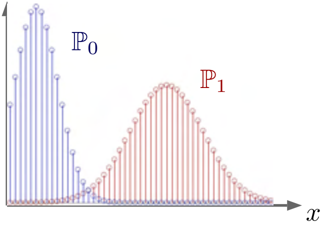


$X \in \mathbb{R}$: number of **spam words** in a message

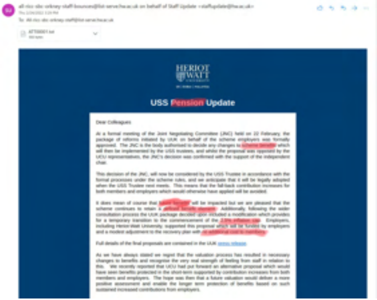
Example in 1D: Spam



$X \in \mathbb{R}$: number of **spam words** in a message



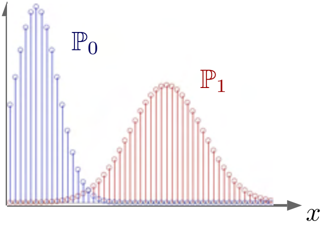
Example in 1D: Spam



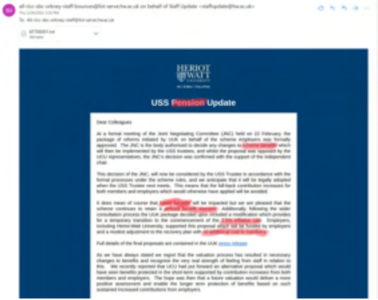
$X \in \mathbb{R}$: number of **spam words** in a message

Null Hypothesis

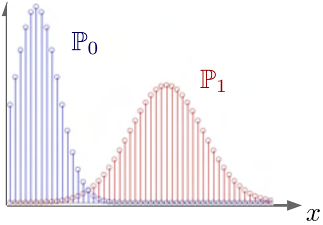
H_0: message isn't spam



Example in 1D: Spam



$X \in \mathbb{R}$: number of **spam words** in a message



Null Hypothesis

H_0: message isn't spam

Alternative Hypothesis

H_1: message isn't spam

Signal vs Noise

Signal vs Noise

Consider a test for detecting:

if given email is spam

Signal vs Noise

Consider a test for detecting:

if given email is spam   presence of aircraft in radar

Signal vs Noise

Consider a test for detecting:

if given email is spam	presence of aircraft in radar
if defendant is guilty	presence of tumor in an image

Signal vs Noise

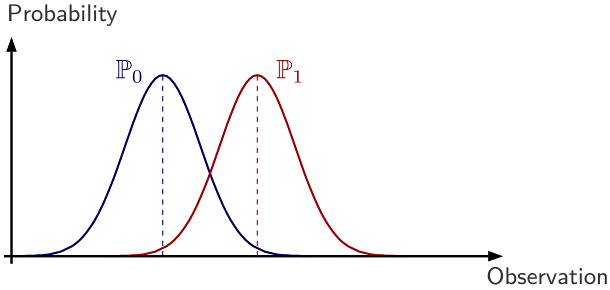
Consider a test for detecting:

if given email is spam	presence of aircraft in radar
if defendant is guilty	presence of tumor in an image

Signal vs Noise

Consider a test for detecting:

if given email is spam	presence of aircraft in radar
if defendant is guilty	presence of tumor in an image

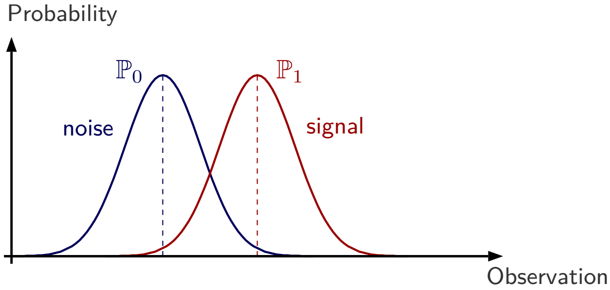


Signal vs Noise

Consider a test for detecting:

if given email is spam — presence of aircraft in radar

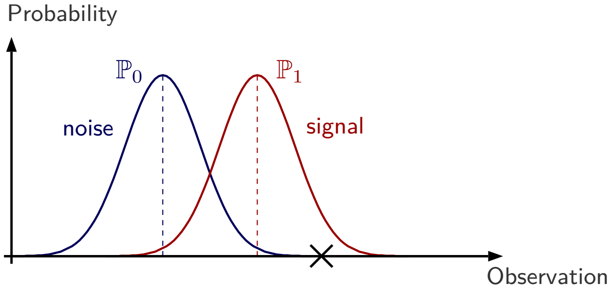
if defendant is guilty — presence of tumor in an image



Signal vs Noise

Consider a test for detecting:

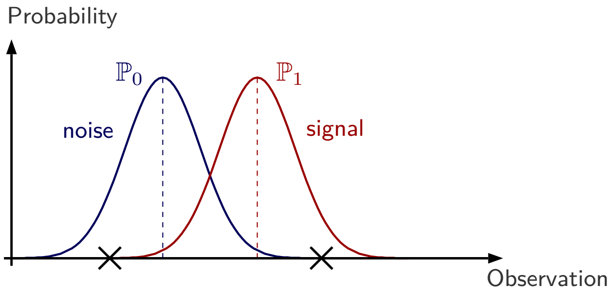
if given email is spam	presence of aircraft in radar
if defendant is guilty	presence of tumor in an image



Signal vs Noise

Consider a test for detecting:

if given email is spam	presence of aircraft in radar
if defendant is guilty	presence of tumor in an image

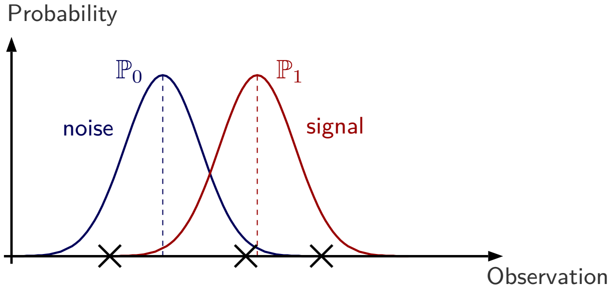


Signal vs Noise

Consider a test for detecting:

if given email is spam   presence of aircraft in radar

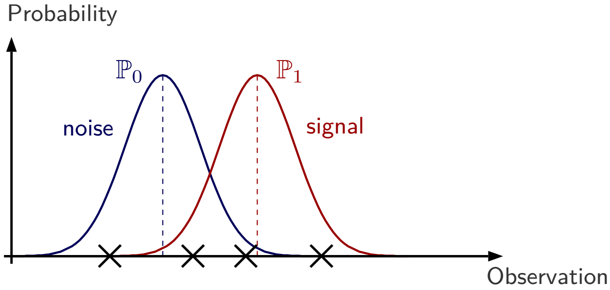
if defendant is guilty   presence of tumor in an image



Signal vs Noise

Consider a test for detecting:

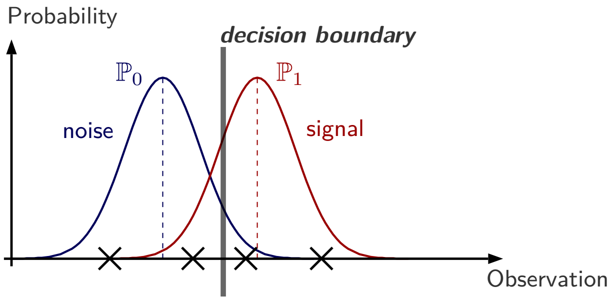
if given email is spam	presence of aircraft in radar
if defendant is guilty	presence of tumor in an image



Signal vs Noise

Consider a test for detecting:

if given email is spam	presence of aircraft in radar
if defendant is guilty	presence of tumor in an image

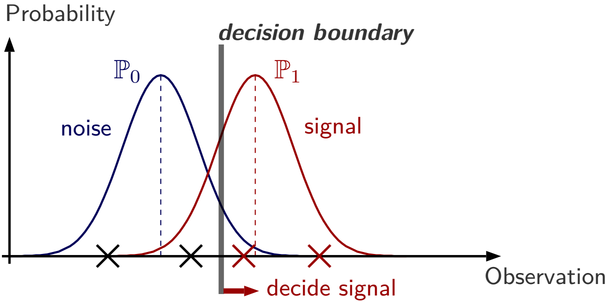


Signal vs Noise

Consider a test for detecting:

if given email is spam | presence of aircraft in radar

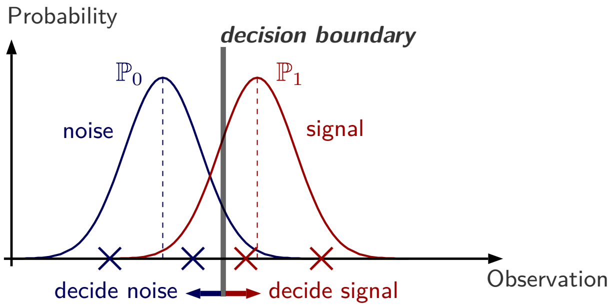
if defendant is guilty | presence of tumor in an image



Signal vs Noise

Consider a test for detecting:

if given email is spam	presence of aircraft in radar
if defendant is guilty	presence of tumor in an image



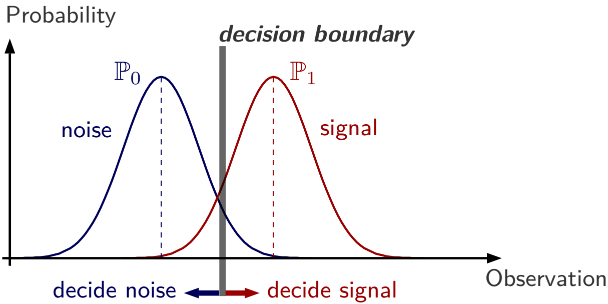
The Decision Tradeoff

The Decision Tradeoff

True hypothesis	Decide noise	Decide signal
noise		
signal		

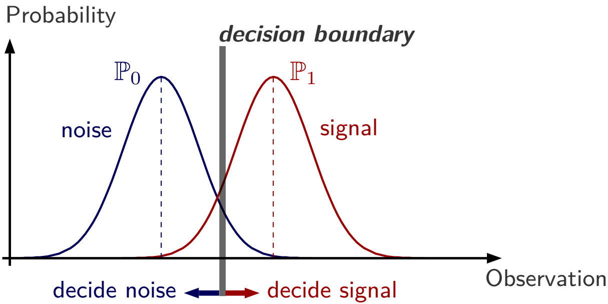
The Decision Tradeoff

True hypothesis	Decide noise	Decide signal
noise		
signal		



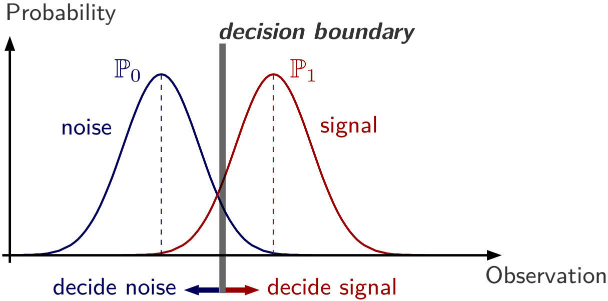
The Decision Tradeoff

True hypothesis	Decide noise	Decide signal
noise	✓	
signal		



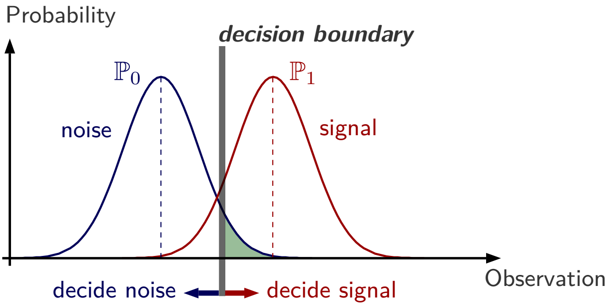
The Decision Tradeoff

True hypothesis	Decide noise	Decide signal
noise	✓	
signal		✓



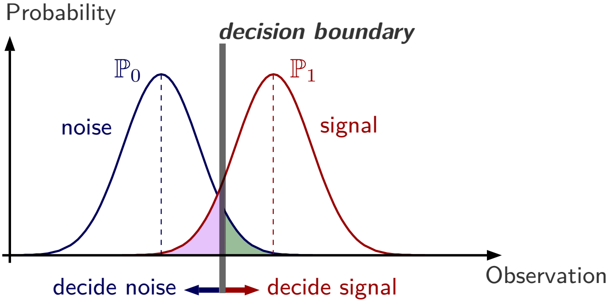
The Decision Tradeoff

True hypothesis	Decide noise	Decide signal
noise	✓	false alarm
signal		✓



The Decision Tradeoff

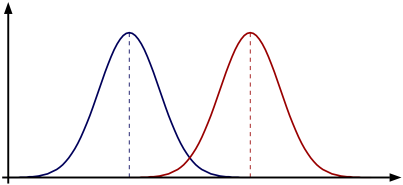
True hypothesis	Decide noise	Decide signal
noise	✓	false alarm
signal	missed detection	✓



Improving the Tradeoff

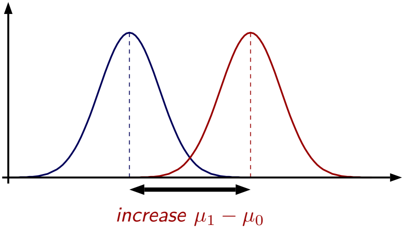
Improving the Tradeoff

Larger effect size



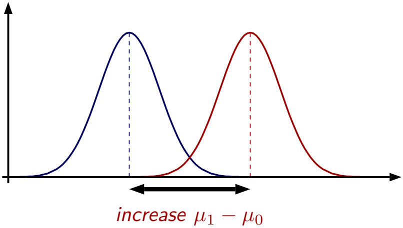
Improving the Tradeoff

Larger effect size

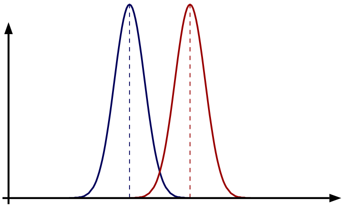


Improving the Tradeoff

Larger effect size

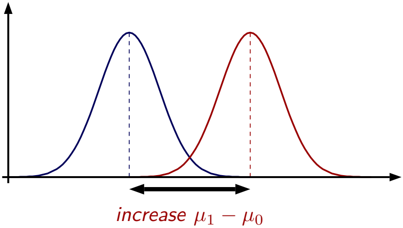


Better/more measurements

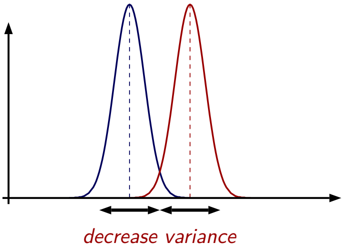


Improving the Tradeoff

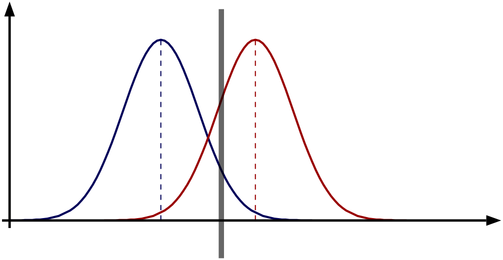
Larger effect size



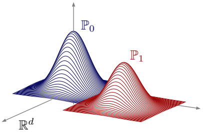
Better/more measurements



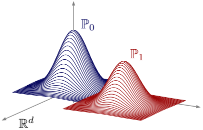
Where to place the decision boundary?



Decision and loss functions



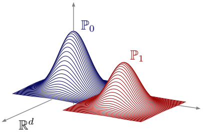
Decision and loss functions



True label

$$Y = \begin{cases} 0 & , \text{ if } H_0 \text{ is true} \\ 1 & , \text{ if } H_1 \text{ true} \end{cases}$$

Decision and loss functions



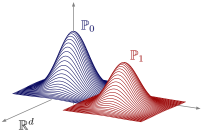
True label

$$Y = \begin{cases} 0 & , \text{ if } H_0 \text{ is true} \\ 1 & , \text{ if } H_1 \text{ true} \end{cases}$$

Decision function $f : \mathbb{R}^d \to \{0, 1\}$

$$f(\boldsymbol{X}) = \begin{cases} 0 & , \text{ if we } \underline{\textit{decide}} \ H_0 \\ 1 & , \text{ if we } \underline{\textit{decide}} \ H_1 \end{cases}$$

Decision and loss functions



True label

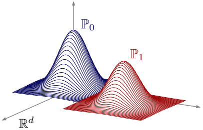
$$Y = \begin{cases} 0 & , \text{ if } H_0 \text{ is true} \\ 1 & , \text{ if } H_1 \text{ true} \end{cases}$$

Decision function $f : \mathbb{R}^d \to \{0, 1\}$

$$f(\boldsymbol{X}) = \begin{cases} 0 & , \text{ if we } \underline{\textit{decide}} \ H_0 \\ 1 & , \text{ if we } \underline{\textit{decide}} \ H_1 \end{cases}$$

Loss function $\ell : \{0, 1\} \times \{0, 1\} \to \mathbb{R}$

Decision and loss functions



True label

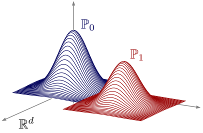
$$Y = \begin{cases} 0 & , \text{ if } H_0 \text{ is true} \\ 1 & , \text{ if } H_1 \text{ true} \end{cases}$$

Decision function $f : \mathbb{R}^d \to \{0, 1\}$

$$f(\boldsymbol{X}) = \begin{cases} 0 & , \text{ if we } \underline{\textit{decide}} \ H_0 \\ 1 & , \text{ if we } \underline{\textit{decide}} \ H_1 \end{cases}$$

Loss function $\ell : \{0, 1\} \times \{0, 1\} \to \mathbb{R}$ $\quad \ell\big(f(\boldsymbol{X}), Y\big)$

Decision and loss functions



True label

$$Y = \begin{cases} 0 & , \text{ if } H_0 \text{ is true} \\ 1 & , \text{ if } H_1 \text{ true} \end{cases}$$

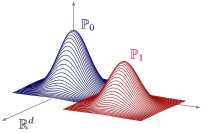
Decision function $f : \mathbb{R}^d \to \{0, 1\}$

$$f(\boldsymbol{X}) = \begin{cases} 0 & , \text{ if we } \underline{\textit{decide}} \ H_0 \\ 1 & , \text{ if we } \underline{\textit{decide}} \ H_1 \end{cases}$$

Loss function $\ell : \{0, 1\} \times \{0, 1\} \to \mathbb{R}$ $\quad \ell\big(f(\boldsymbol{X}), Y\big)$

True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true		
H_1 is true		

Decision and loss functions



True label

$$Y = \begin{cases} 0 & , \text{ if } H_0 \text{ is true} \\ 1 & , \text{ if } H_1 \text{ true} \end{cases}$$

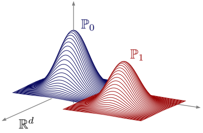
Decision function $f : \mathbb{R}^d \to \{0, 1\}$

$$f(\boldsymbol{X}) = \begin{cases} 0 & , \text{ if we } \underline{\textit{decide}}\ H_0 \\ 1 & , \text{ if we } \underline{\textit{decide}}\ H_1 \end{cases}$$

Loss function $\ell : \{0,1\} \times \{0,1\} \to \mathbb{R}$ $\quad \ell\big(f(\boldsymbol{X}), Y\big)$

True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true	$\ell(0,\, 0)$	
H_1 is true		

Decision and loss functions



True label

$$Y = \begin{cases} 0 & , \text{ if } H_0 \text{ is true} \\ 1 & , \text{ if } H_1 \text{ true} \end{cases}$$

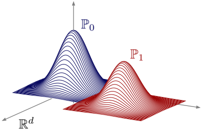
Decision function $f : \mathbb{R}^d \to \{0, 1\}$

$$f(\boldsymbol{X}) = \begin{cases} 0 & , \text{ if we } \underline{\textit{decide}} \ H_0 \\ 1 & , \text{ if we } \underline{\textit{decide}} \ H_1 \end{cases}$$

Loss function $\ell : \{0, 1\} \times \{0, 1\} \to \mathbb{R}$ $\quad \ell\big(f(\boldsymbol{X}), Y\big)$

True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true	$\ell(0, 0)$	$\ell(1, 0)$
H_1 is true		

Decision and loss functions



True label

$$Y = \begin{cases} 0 & , \text{ if } H_0 \text{ is true} \\ 1 & , \text{ if } H_1 \text{ true} \end{cases}$$

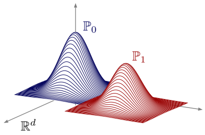
Decision function $f : \mathbb{R}^d \to \{0, 1\}$

$$f(\boldsymbol{X}) = \begin{cases} 0 & , \text{ if we } \underline{\textit{decide}} \ H_0 \\ 1 & , \text{ if we } \underline{\textit{decide}} \ H_1 \end{cases}$$

Loss function $\ell : \{0, 1\} \times \{0, 1\} \to \mathbb{R}$ $\quad \ell\big(f(\boldsymbol{X}), Y\big)$

True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true	$\ell(0, 0)$	$\ell(1, 0)$
H_1 is true	$\ell(0, 1)$	

Decision and loss functions



True label

$$Y = \begin{cases} 0 & , \text{ if } H_0 \text{ is true} \\ 1 & , \text{ if } H_1 \text{ true} \end{cases}$$

Decision function $f : \mathbb{R}^d \to \{0, 1\}$

$$f(\boldsymbol{X}) = \begin{cases} 0 & , \text{ if we } \underline{\textit{decide}} \ H_0 \\ 1 & , \text{ if we } \underline{\textit{decide}} \ H_1 \end{cases}$$

Loss function $\ell : \{0, 1\} \times \{0, 1\} \to \mathbb{R}$ $\quad \ell\big(f(\boldsymbol{X}), Y\big)$

True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true	$\ell(0,\, 0)$	$\ell(1, 0)$
H_1 is true	$\ell(0, 1)$	$\ell(1, 1)$

Risk and Optimal Decision

Risk and Optimal Decision

Given *decision function* $f : \mathbb{R}^d \to \{0, 1\}$ and *loss* $\ell : \{0, 1\}^2 \to \mathbb{R}$,

Risk and Optimal Decision

Given *decision function* $f : \mathbb{R}^d \to \{0, 1\}$ and *loss* $\ell : \{0, 1\}^2 \to \mathbb{R}$,

Risk:

Risk and Optimal Decision

Given *decision function* $f : \mathbb{R}^d \to \{0,1\}$ and *loss* $\ell : \{0,1\}^2 \to \mathbb{R}$,

Risk:

$$R[f] := \mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}), Y\big)\Big]$$

Risk and Optimal Decision

Given *decision function* $f : \mathbb{R}^d \to \{0, 1\}$ and *loss* $\ell : \{0, 1\}^2 \to \mathbb{R}$,

Risk:

$$\boxed{R[f] := \mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}), Y\big)\Big]}$$

where $\mathbb{E}_{\boldsymbol{X}Y}[\cdot]$ is the expectation with respect to $\boldsymbol{X}$ and Y

Risk and Optimal Decision

Given *decision function* $f : \mathbb{R}^d \to \{0, 1\}$ and *loss* $\ell : \{0, 1\}^2 \to \mathbb{R}$,

Risk:

$$R[f] := \mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}), Y\big)\Big]$$

where $\mathbb{E}_{\boldsymbol{X}Y}[\cdot]$ is the expectation with respect to $\boldsymbol{X}$ and Y

Optimal decision problem:

Risk and Optimal Decision

Given *decision function* $f : \mathbb{R}^d \to \{0,1\}$ and *loss* $\ell : \{0,1\}^2 \to \mathbb{R}$,

Risk:

$$\boxed{R[f] := \mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}), Y\big)\Big]}$$

where $\mathbb{E}_{\boldsymbol{X}Y}[\cdot]$ is the expectation with respect to $\boldsymbol{X}$ and Y

Optimal decision problem: Given ℓ, find f that minimizes the risk:

Risk and Optimal Decision

Given *decision function* $f : \mathbb{R}^d \to \{0, 1\}$ and *loss* $\ell : \{0, 1\}^2 \to \mathbb{R}$,

Risk:

$$R[f] := \mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}), Y\big)\Big]$$

where $\mathbb{E}_{\boldsymbol{X}Y}[\cdot]$ is the expectation with respect to $\boldsymbol{X}$ and Y

Optimal decision problem: Given ℓ, find f that minimizes the risk:

$$\underset{f:\mathbb{R}^d \to \{0,1\}}{\text{minimize}} \; \mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}), Y\big)\Big]$$

Risk and Optimal Decision

Given *decision function* $f : \mathbb{R}^d \to \{0, 1\}$ and *loss* $\ell : \{0, 1\}^2 \to \mathbb{R}$,

Risk:

$$R[f] := \mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}), Y\big)\Big]$$

where $\mathbb{E}_{\boldsymbol{X}Y}[\cdot]$ is the expectation with respect to $\boldsymbol{X}$ and Y

Optimal decision problem: Given ℓ, find f that minimizes the risk:

$$\underset{f:\mathbb{R}^d \to \{0,1\}}{\text{minimize}} \ \mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}), Y\big)\Big]$$

. . . infinite-dimensional problem

$$\underset{f:\mathbb{R}^d \to \{0,1\}}{\text{minimize}} \ \mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}), Y\big)\Big]$$

$$\underset{f:\mathbb{R}^d \to \{0,1\}}{\text{minimize}} \ \mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}), Y\big)\Big]$$

Recall that $f(\boldsymbol{X})$ and Y are binary

$$\underset{f:\mathbb{R}^d \to \{0,1\}}{\text{minimize}} \ \mathbb{E}_{\boldsymbol{X}Y}\left[\ell\big(f(\boldsymbol{X}), Y\big)\right]$$

Recall that $f(\boldsymbol{X})$ and Y are binary

Conditioning on $\boldsymbol{X}$,

$$\underset{f:\mathbb{R}^d \to \{0,1\}}{\text{minimize}} \;\; \mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}), Y\big)\Big]$$

Recall that $f(\boldsymbol{X})$ and Y are binary

Conditioning on $\boldsymbol{X}$,

$$\mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}, Y)\big)\Big] = \mathbb{E}_{\boldsymbol{X}}\left[\mathbb{E}_{Y}\left[\ell\big(f(\boldsymbol{X}),\, Y\big) \,\Big|\, \boldsymbol{X}\right]\right]$$

$$\underset{f:\mathbb{R}^d\to\{0,1\}}{\text{minimize}} \ \mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}), Y\big)\Big]$$

Recall that $f(\boldsymbol{X})$ and Y are binary

Conditioning on $\boldsymbol{X}$,

$$\begin{aligned}\mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X},Y)\big)\Big] &= \mathbb{E}_{\boldsymbol{X}}\left[\mathbb{E}_Y\left[\ell\big(f(\boldsymbol{X}),\, Y\big)\,\Big|\, \boldsymbol{X}\right]\right]\\ &= \int_{\mathbb{R}^d} \mathbb{E}_Y\left[\ell\big(f(\boldsymbol{X}),\, Y\big)\,\Big|\, \boldsymbol{X} = \boldsymbol{x}\right] f_{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\end{aligned}$$

$$\underset{f:\mathbb{R}^d\to\{0,1\}}{\text{minimize}} \;\; \mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}), Y\big)\Big]$$

Recall that $f(\boldsymbol{X})$ and Y are binary

Conditioning on $\boldsymbol{X}$,

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}, Y)\big)\Big] &= \mathbb{E}_{\boldsymbol{X}}\left[\mathbb{E}_Y\left[\ell\big(f(\boldsymbol{X}),\, Y\big)\,\Big|\, \boldsymbol{X}\right]\right] \\
&= \int_{\mathbb{R}^d} \mathbb{E}_Y\left[\ell\big(f(\boldsymbol{X}),\, Y\big)\,\Big|\, \boldsymbol{X} = \boldsymbol{x}\right] f_{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}
\end{aligned}$$

If $f(\boldsymbol{x}) = 0$,

$$\underset{f:\mathbb{R}^d\to\{0,1\}}{\text{minimize}}\;\mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}),Y\big)\Big]$$

Recall that $f(\boldsymbol{X})$ and Y are binary

Conditioning on $\boldsymbol{X}$,

$$\begin{aligned}\mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X},Y)\big)\Big] &= \mathbb{E}_{\boldsymbol{X}}\left[\mathbb{E}_Y\left[\ell\big(f(\boldsymbol{X}),\,Y\big)\,\Big|\,\boldsymbol{X}\right]\right]\\ &= \int_{\mathbb{R}^d}\mathbb{E}_Y\left[\ell\big(f(\boldsymbol{X}),\,Y\big)\,\Big|\,\boldsymbol{X}=\boldsymbol{x}\right]f_{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\end{aligned}$$

If $f(\boldsymbol{x})=0$,

$$\mathbb{E}_Y\left[\ell\big(0,\,Y\big)\,\Big|\,\boldsymbol{X}=\boldsymbol{x}\right] = \ell(0,0)\,\mathbb{P}\big(Y=0\,|\,\boldsymbol{X}=\boldsymbol{x}\big)+\ell(0,1)\,\mathbb{P}\big(Y=1\,|\,\boldsymbol{X}=\boldsymbol{x}\big)$$

$$\underset{f:\mathbb{R}^d\to\{0,1\}}{\text{minimize}} \ \mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}), Y\big)\Big]$$

Recall that $f(\boldsymbol{X})$ and Y are binary

Conditioning on $\boldsymbol{X}$,

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}, Y)\big)\Big] &= \mathbb{E}_{\boldsymbol{X}}\bigg[\mathbb{E}_Y\Big[\ell\big(f(\boldsymbol{X}),\, Y\big)\,\Big|\,\boldsymbol{X}\Big]\bigg] \\
&= \int_{\mathbb{R}^d} \mathbb{E}_Y\Big[\ell\big(f(\boldsymbol{X}),\, Y\big)\,\Big|\,\boldsymbol{X}=\boldsymbol{x}\Big] f_{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}
\end{aligned}$$

If $f(\boldsymbol{x}) = 0$,

$$\mathbb{E}_Y\Big[\ell\big(0,\, Y\big)\,\Big|\,\boldsymbol{X}=\boldsymbol{x}\Big] = \ell(0,0)\,\mathbb{P}\big(Y=0\,|\,\boldsymbol{X}=\boldsymbol{x}\big) + \ell(0,1)\,\mathbb{P}\big(Y=1\,|\,\boldsymbol{X}=\boldsymbol{x}\big)$$

If $f(\boldsymbol{x}) = 1$,

$$\underset{f:\mathbb{R}^d \to \{0,1\}}{\text{minimize}} \;\; \mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}), Y\big)\Big]$$

Recall that $f(\boldsymbol{X})$ and Y are binary

Conditioning on $\boldsymbol{X}$,

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}, Y)\big)\Big] &= \mathbb{E}_{\boldsymbol{X}}\Bigg[\mathbb{E}_Y\Big[\ell\big(f(\boldsymbol{X}),\, Y\big) \,\Big|\, \boldsymbol{X}\Big]\Bigg] \\
&= \int_{\mathbb{R}^d} \mathbb{E}_Y\Big[\ell\big(f(\boldsymbol{X}),\, Y\big) \,\Big|\, \boldsymbol{X} = \boldsymbol{x}\Big] f_{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}
\end{aligned}$$

If $f(\boldsymbol{x}) = 0$,

$$\mathbb{E}_Y\Big[\ell\big(0,\, Y\big) \,\Big|\, \boldsymbol{X} = \boldsymbol{x}\Big] = \ell(0,0)\,\mathbb{P}\big(Y = 0 \,|\, \boldsymbol{X} = \boldsymbol{x}\big) + \ell(0,1)\,\mathbb{P}\big(Y = 1 \,|\, \boldsymbol{X} = \boldsymbol{x}\big)$$

If $f(\boldsymbol{x}) = 1$,

$$\mathbb{E}_Y\Big[\ell\big(1,\, Y\big) \,\Big|\, \boldsymbol{X} = \boldsymbol{x}\Big] = \ell(1,0)\,\mathbb{P}\big(Y = 0 \,|\, \boldsymbol{X} = \boldsymbol{x}\big) + \ell(1,1)\,\mathbb{P}\big(Y = 1 \,|\, \boldsymbol{X} = \boldsymbol{x}\big)$$

Optimal decision

Optimal decision

$$f(\boldsymbol{x}) = 0 \qquad \text{if} \qquad \mathbb{E}_Y\left[\ell\big(0,\, Y\big) \,\Big|\, \boldsymbol{X} = \boldsymbol{x}\right] < \mathbb{E}_Y\left[\ell\big(1,\, Y\big) \,\Big|\, \boldsymbol{X} = \boldsymbol{x}\right]$$

Optimal decision

$$f(\boldsymbol{x}) = 0 \qquad \text{if} \qquad \mathbb{E}_Y\left[\ell\big(0,\, Y\big) \,\middle|\, \boldsymbol{X} = \boldsymbol{x}\right] < \mathbb{E}_Y\left[\ell\big(1,\, Y\big) \,\middle|\, \boldsymbol{X} = \boldsymbol{x}\right]$$

$$f(\boldsymbol{x}) = 1 \qquad \text{if} \qquad \mathbb{E}_Y\left[\ell\big(0,\, Y\big) \,\middle|\, \boldsymbol{X} = \boldsymbol{x}\right] \geq \mathbb{E}_Y\left[\ell\big(1,\, Y\big) \,\middle|\, \boldsymbol{X} = \boldsymbol{x}\right]$$

Optimal decision

$$f(\boldsymbol{x}) = 0 \qquad \text{if} \qquad \mathbb{E}_Y\left[\ell\big(0,\,Y\big)\,\middle|\,\boldsymbol{X}=\boldsymbol{x}\right] < \mathbb{E}_Y\left[\ell\big(1,\,Y\big)\,\middle|\,\boldsymbol{X}=\boldsymbol{x}\right]$$

$$f(\boldsymbol{x}) = 1 \qquad \text{if} \qquad \mathbb{E}_Y\left[\ell\big(0,\,Y\big)\,\middle|\,\boldsymbol{X}=\boldsymbol{x}\right] \geq \mathbb{E}_Y\left[\ell\big(1,\,Y\big)\,\middle|\,\boldsymbol{X}=\boldsymbol{x}\right]$$

Rearranging, the optimal decision is

Optimal decision

$$f(\boldsymbol{x}) = 0 \qquad \text{if} \qquad \mathbb{E}_Y\left[\ell\big(0,\, Y\big)\,\Big|\,\boldsymbol{X} = \boldsymbol{x}\right] < \mathbb{E}_Y\left[\ell\big(1,\, Y\big)\,\Big|\,\boldsymbol{X} = \boldsymbol{x}\right]$$

$$f(\boldsymbol{x}) = 1 \qquad \text{if} \qquad \mathbb{E}_Y\left[\ell\big(0,\, Y\big)\,\Big|\,\boldsymbol{X} = \boldsymbol{x}\right] \geq \mathbb{E}_Y\left[\ell\big(1,\, Y\big)\,\Big|\,\boldsymbol{X} = \boldsymbol{x}\right]$$

Rearranging, the optimal decision is

$$f(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \mathbb{P}\big(Y = 1 \,\big|\, \boldsymbol{X} = \boldsymbol{x}\big) \geq \frac{\ell(0,0) - \ell(1,0)}{\ell(1,1) - \ell(0,1)}\, \mathbb{P}\big(Y = 0 \,\big|\, \boldsymbol{X} = \boldsymbol{x}\big) \\ 0 & \text{otherwise} \end{cases}$$

Optimal decision

$$
\begin{aligned}
f(\boldsymbol{x}) &= 0 \qquad \text{if} \qquad \mathbb{E}_Y\left[\ell\big(0,\, Y\big) \,\middle|\, \boldsymbol{X} = \boldsymbol{x}\right] < \mathbb{E}_Y\left[\ell\big(1,\, Y\big) \,\middle|\, \boldsymbol{X} = \boldsymbol{x}\right] \\
f(\boldsymbol{x}) &= 1 \qquad \text{if} \qquad \mathbb{E}_Y\left[\ell\big(0,\, Y\big) \,\middle|\, \boldsymbol{X} = \boldsymbol{x}\right] \geq \mathbb{E}_Y\left[\ell\big(1,\, Y\big) \,\middle|\, \boldsymbol{X} = \boldsymbol{x}\right]
\end{aligned}
$$

Rearranging, the optimal decision is

$$
f(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \mathbb{P}\big(\overset{H_1}{Y = 1} \,\big|\, \boldsymbol{X} = \boldsymbol{x}\big) \geq \frac{\ell(0,0) - \ell(1,0)}{\ell(1,1) - \ell(0,1)}\, \mathbb{P}\big(Y = 0 \,\big|\, \boldsymbol{X} = \boldsymbol{x}\big) \\ 0 & \text{otherwise} \end{cases}
$$

Optimal decision

$$f(\boldsymbol{x}) = 0 \qquad \text{if} \qquad \mathbb{E}_Y\left[\ell\big(0,\, Y\big) \,\middle|\, \boldsymbol{X} = \boldsymbol{x}\right] < \mathbb{E}_Y\left[\ell\big(1,\, Y\big) \,\middle|\, \boldsymbol{X} = \boldsymbol{x}\right]$$

$$f(\boldsymbol{x}) = 1 \qquad \text{if} \qquad \mathbb{E}_Y\left[\ell\big(0,\, Y\big) \,\middle|\, \boldsymbol{X} = \boldsymbol{x}\right] \geq \mathbb{E}_Y\left[\ell\big(1,\, Y\big) \,\middle|\, \boldsymbol{X} = \boldsymbol{x}\right]$$

Rearranging, the optimal decision is

$$f(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \mathbb{P}\big(\overset{H_1}{Y = 1} \,\big|\, \boldsymbol{X} = \boldsymbol{x}\big) \geq \frac{\ell(0,0)-\ell(1,0)}{\ell(1,1)-\ell(0,1)}\, \mathbb{P}\big(\overset{H_0}{Y = 0} \,\big|\, \boldsymbol{X} = \boldsymbol{x}\big) \\ 0 & \text{otherwise} \end{cases}$$

Optimal decision

$$
\begin{aligned}
f(\boldsymbol{x}) = 0 \qquad &\text{if} \qquad \mathbb{E}_Y\left[\ell\big(0,\, Y\big) \,\middle|\, \boldsymbol{X} = \boldsymbol{x}\right] < \mathbb{E}_Y\left[\ell\big(1,\, Y\big) \,\middle|\, \boldsymbol{X} = \boldsymbol{x}\right] \\
f(\boldsymbol{x}) = 1 \qquad &\text{if} \qquad \mathbb{E}_Y\left[\ell\big(0,\, Y\big) \,\middle|\, \boldsymbol{X} = \boldsymbol{x}\right] \geq \mathbb{E}_Y\left[\ell\big(1,\, Y\big) \,\middle|\, \boldsymbol{X} = \boldsymbol{x}\right]
\end{aligned}
$$

Rearranging, the optimal decision is

$$
f(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \mathbb{P}\big(\overset{H_1}{Y = 1} \,\big|\, \boldsymbol{X} = \boldsymbol{x}\big) \geq \frac{\ell(0,0) - \ell(1,0)}{\ell(1,1) - \ell(0,1)}\, \mathbb{P}\big(\overset{H_0}{Y = 0} \,\big|\, \boldsymbol{X} = \boldsymbol{x}\big) \\ 0 & \text{otherwise} \end{cases}
$$

Bayes rule

Optimal decision

$$f(\boldsymbol{x}) = 0 \quad \text{if} \quad \mathbb{E}_Y\left[\ell\big(0,\, Y\big) \,\middle|\, \boldsymbol{X} = \boldsymbol{x}\right] < \mathbb{E}_Y\left[\ell\big(1,\, Y\big) \,\middle|\, \boldsymbol{X} = \boldsymbol{x}\right]$$

$$f(\boldsymbol{x}) = 1 \quad \text{if} \quad \mathbb{E}_Y\left[\ell\big(0,\, Y\big) \,\middle|\, \boldsymbol{X} = \boldsymbol{x}\right] \geq \mathbb{E}_Y\left[\ell\big(1,\, Y\big) \,\middle|\, \boldsymbol{X} = \boldsymbol{x}\right]$$

Rearranging, the optimal decision is

$$f(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \mathbb{P}\big(\overbrace{Y = 1}^{H_1} \,\big|\, \boldsymbol{X} = \boldsymbol{x}\big) \geq \frac{\ell(0,0) - \ell(1,0)}{\ell(1,1) - \ell(0,1)}\, \mathbb{P}\big(\overbrace{Y = 0}^{H_0} \,\big|\, \boldsymbol{X} = \boldsymbol{x}\big) \\ 0 & \text{otherwise} \end{cases}$$

Bayes rule

$$\frac{f_{\boldsymbol{X} \mid H_1}(\boldsymbol{x} \mid H_1)\, \mathbb{P}(H_1)}{f_{\boldsymbol{X}}(\boldsymbol{x})}$$

Optimal decision

$$
\begin{aligned}
f(\boldsymbol{x}) = 0 \qquad & \text{if} \qquad \mathbb{E}_Y\left[\ell\big(0,\, Y\big)\,\middle|\,\boldsymbol{X} = \boldsymbol{x}\right] < \mathbb{E}_Y\left[\ell\big(1,\, Y\big)\,\middle|\,\boldsymbol{X} = \boldsymbol{x}\right] \\
f(\boldsymbol{x}) = 1 \qquad & \text{if} \qquad \mathbb{E}_Y\left[\ell\big(0,\, Y\big)\,\middle|\,\boldsymbol{X} = \boldsymbol{x}\right] \geq \mathbb{E}_Y\left[\ell\big(1,\, Y\big)\,\middle|\,\boldsymbol{X} = \boldsymbol{x}\right]
\end{aligned}
$$

Rearranging, the optimal decision is

$$
f(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \mathbb{P}\big(\overset{H_1}{Y = 1}\,\big|\,\boldsymbol{X} = \boldsymbol{x}\big) \geq \frac{\ell(0,0)-\ell(1,0)}{\ell(1,1)-\ell(0,1)}\,\mathbb{P}\big(\overset{H_0}{Y = 0}\,\big|\,\boldsymbol{X} = \boldsymbol{x}\big) \\ 0 & \text{otherwise} \end{cases}
$$

Bayes rule

$$
\frac{f_{\boldsymbol{X}|H_1}(\boldsymbol{x}\,|\,H_1)\,\mathbb{P}(H_1)}{f_{\boldsymbol{X}}(\boldsymbol{x})} \qquad\qquad \frac{f_{\boldsymbol{X}|H_0}(\boldsymbol{x}\,|\,H_0)\,\mathbb{P}(H_0)}{f_{\boldsymbol{X}}(\boldsymbol{x})}
$$

Optimal decision

$$f(\boldsymbol{x}) = 0 \quad \text{if} \quad \mathbb{E}_Y\left[\ell\big(0,\, Y\big) \,\Big|\, \boldsymbol{X} = \boldsymbol{x}\right] < \mathbb{E}_Y\left[\ell\big(1,\, Y\big) \,\Big|\, \boldsymbol{X} = \boldsymbol{x}\right]$$

$$f(\boldsymbol{x}) = 1 \quad \text{if} \quad \mathbb{E}_Y\left[\ell\big(0,\, Y\big) \,\Big|\, \boldsymbol{X} = \boldsymbol{x}\right] \geq \mathbb{E}_Y\left[\ell\big(1,\, Y\big) \,\Big|\, \boldsymbol{X} = \boldsymbol{x}\right]$$

Rearranging, the optimal decision is

$$f(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \mathbb{P}\big(\overset{H_1}{Y = 1} \,\big|\, \boldsymbol{X} = \boldsymbol{x}\big) \geq \frac{\ell(0,0) - \ell(1,0)}{\ell(1,1) - \ell(0,1)}\, \mathbb{P}\big(\overset{H_0}{Y = 0} \,\big|\, \boldsymbol{X} = \boldsymbol{x}\big) \\ 0 & \text{otherwise} \end{cases}$$

Bayes rule

$$\frac{f_{\boldsymbol{X}|H_1}(\boldsymbol{x} \,|\, H_1)\,\mathbb{P}(H_1)}{f_{\boldsymbol{X}}(\boldsymbol{x})} \qquad\qquad \frac{f_{\boldsymbol{X}|H_0}(\boldsymbol{x} \,|\, H_0)\,\mathbb{P}(H_0)}{f_{\boldsymbol{X}}(\boldsymbol{x})}$$

$$f(\boldsymbol{x}) = 1 \quad \text{if} \quad \frac{f_{\boldsymbol{X}|H_1}(\boldsymbol{x} \,|\, H_1)}{f_{\boldsymbol{X}|H_0}(\boldsymbol{x} \,|\, H_0)} \geq \frac{\ell(0,0) - \ell(1,0)}{\ell(1,1) - \ell(0,1)} \cdot \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)}$$

Optimal decision

$$f(\boldsymbol{x}) = 0 \quad \text{if} \quad \mathbb{E}_Y\left[\ell\big(0, Y\big) \,\middle|\, \boldsymbol{X} = \boldsymbol{x}\right] < \mathbb{E}_Y\left[\ell\big(1, Y\big) \,\middle|\, \boldsymbol{X} = \boldsymbol{x}\right]$$

$$f(\boldsymbol{x}) = 1 \quad \text{if} \quad \mathbb{E}_Y\left[\ell\big(0, Y\big) \,\middle|\, \boldsymbol{X} = \boldsymbol{x}\right] \geq \mathbb{E}_Y\left[\ell\big(1, Y\big) \,\middle|\, \boldsymbol{X} = \boldsymbol{x}\right]$$

Rearranging, the optimal decision is

$$f(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \mathbb{P}\big(\overset{H_1}{Y = 1} \,|\, \boldsymbol{X} = \boldsymbol{x}\big) \geq \frac{\ell(0,0)-\ell(1,0)}{\ell(1,1)-\ell(0,1)}\, \mathbb{P}\big(\overset{H_0}{Y = 0} \,|\, \boldsymbol{X} = \boldsymbol{x}\big) \\ 0 & \text{otherwise} \end{cases}$$

Bayes rule

$$\frac{f_{\boldsymbol{X}|H_1}(\boldsymbol{x} \,|\, H_1)\, \mathbb{P}(H_1)}{f_{\boldsymbol{X}}(\boldsymbol{x})} \qquad \frac{f_{\boldsymbol{X}|H_0}(\boldsymbol{x} \,|\, H_0)\, \mathbb{P}(H_0)}{f_{\boldsymbol{X}}(\boldsymbol{x})}$$

likelihood ratio test

$$f(\boldsymbol{x}) = 1 \quad \text{if} \quad \boxed{\frac{f_{\boldsymbol{X}|H_1}(\boldsymbol{x} \,|\, H_1)}{f_{\boldsymbol{X}|H_0}(\boldsymbol{x} \,|\, H_0)} \geq \frac{\ell(0,0) - \ell(1,0)}{\ell(1,1) - \ell(0,1)} \cdot \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)}}$$

Optimal decision

$$
\begin{aligned}
f(\boldsymbol{x}) = 0 \quad &\text{if} \quad \mathbb{E}_Y\left[\ell(0,\,Y)\,\middle|\,\boldsymbol{X}=\boldsymbol{x}\right] < \mathbb{E}_Y\left[\ell(1,\,Y)\,\middle|\,\boldsymbol{X}=\boldsymbol{x}\right] \\
f(\boldsymbol{x}) = 1 \quad &\text{if} \quad \mathbb{E}_Y\left[\ell(0,\,Y)\,\middle|\,\boldsymbol{X}=\boldsymbol{x}\right] \geq \mathbb{E}_Y\left[\ell(1,\,Y)\,\middle|\,\boldsymbol{X}=\boldsymbol{x}\right]
\end{aligned}
$$

Rearranging, the optimal decision is

$$
f(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \underbrace{\mathbb{P}(\overset{H_1}{Y=1}\mid \boldsymbol{X}=\boldsymbol{x})}_{\dfrac{f_{\boldsymbol{X}\mid H_1}(\boldsymbol{x}\mid H_1)\,\mathbb{P}(H_1)}{f_{\boldsymbol{X}}(\boldsymbol{x})}} \geq \frac{\ell(0,0)-\ell(1,0)}{\ell(1,1)-\ell(0,1)}\,\underbrace{\mathbb{P}(\overset{H_0}{Y=0}\mid \boldsymbol{X}=\boldsymbol{x})}_{\dfrac{f_{\boldsymbol{X}\mid H_0}(\boldsymbol{x}\mid H_0)\,\mathbb{P}(H_0)}{f_{\boldsymbol{X}}(\boldsymbol{x})}} \\ 0 & \text{otherwise} \end{cases}
$$

Bayes rule

likelihood ratio test

$$
f(\boldsymbol{x}) = 1 \quad \text{if} \quad \boxed{\underbrace{\frac{f_{\boldsymbol{X}\mid H_1}(\boldsymbol{x}\mid H_1)}{f_{\boldsymbol{X}\mid H_0}(\boldsymbol{x}\mid H_0)}}_{\mathcal{L}(\boldsymbol{x})} \geq \frac{\ell(0,0)-\ell(1,0)}{\ell(1,1)-\ell(0,1)} \cdot \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)}}
$$

Optimal decision

$$f(\boldsymbol{x}) = 0 \quad \text{if} \quad \mathbb{E}_Y\left[\ell(0, Y) \,\middle|\, \boldsymbol{X} = \boldsymbol{x}\right] < \mathbb{E}_Y\left[\ell(1, Y) \,\middle|\, \boldsymbol{X} = \boldsymbol{x}\right]$$

$$f(\boldsymbol{x}) = 1 \quad \text{if} \quad \mathbb{E}_Y\left[\ell(0, Y) \,\middle|\, \boldsymbol{X} = \boldsymbol{x}\right] \geq \mathbb{E}_Y\left[\ell(1, Y) \,\middle|\, \boldsymbol{X} = \boldsymbol{x}\right]$$

Rearranging, the optimal decision is

$$f(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \mathbb{P}\big(\overset{H_1}{Y = 1} \,\big|\, \boldsymbol{X} = \boldsymbol{x}\big) \geq \frac{\ell(0,0)-\ell(1,0)}{\ell(1,1)-\ell(0,1)} \, \mathbb{P}\big(\overset{H_0}{Y = 0} \,\big|\, \boldsymbol{X} = \boldsymbol{x}\big) \\ 0 & \text{otherwise} \end{cases}$$

Bayes rule

$$\frac{f_{\boldsymbol{X}|H_1}(\boldsymbol{x} \mid H_1)\,\mathbb{P}(H_1)}{f_{\boldsymbol{X}}(\boldsymbol{x})} \qquad \frac{f_{\boldsymbol{X}|H_0}(\boldsymbol{x} \mid H_0)\,\mathbb{P}(H_0)}{f_{\boldsymbol{X}}(\boldsymbol{x})}$$

likelihood ratio test

$$f(\boldsymbol{x}) = 1 \quad \text{if} \quad \boxed{\frac{f_{\boldsymbol{X}|H_1}(\boldsymbol{x} \mid H_1)}{f_{\boldsymbol{X}|H_0}(\boldsymbol{x} \mid H_0)} \geq \frac{\ell(0,0)-\ell(1,0)}{\ell(1,1)-\ell(0,1)} \cdot \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)}}$$

$\mathcal{L}(\boldsymbol{x})$: likelihood ratio

Likelihood ratio test

$$\underset{f:\mathbb{R}^d\to\{0,1\}}{\text{minimize}} \; \mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}),Y\big)\Big]$$

Likelihood ratio test

$$\underset{f:\mathbb{R}^d \to \{0,1\}}{\text{minimize}} \ \mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}), Y\big)\Big]$$

The decision that minimizes the risk in a binary hypothesis test is

Likelihood ratio test

$$\underset{f:\mathbb{R}^d \to \{0,1\}}{\text{minimize}} \; \mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}), Y\big)\Big]$$

The decision that minimizes the risk in a binary hypothesis test is

$$f(\boldsymbol{x}) = \mathbb{1}_{\{\mathcal{L}(\boldsymbol{x}) \geq \eta\}}(\boldsymbol{x})$$

Likelihood ratio test

$$\underset{f:\mathbb{R}^d \to \{0,1\}}{\text{minimize}} \;\; \mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}), Y\big)\Big]$$

The decision that minimizes the risk in a binary hypothesis test is

$$f(\boldsymbol{x}) = \mathbb{1}_{\{\mathcal{L}(\boldsymbol{x}) \geq \eta\}}(\boldsymbol{x})$$

- Indicator function of set $\mathcal{S}$: $\quad \mathbb{1}_{\mathcal{S}}(s) = \begin{cases} 1 & , \text{ if } s \in \mathcal{S} \\ 0 & , \text{ if } s \notin \mathcal{S} \end{cases}$

Likelihood ratio test

$$\underset{f:\mathbb{R}^d \to \{0,1\}}{\text{minimize}} \ \mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}), Y\big)\Big]$$

The decision that minimizes the risk in a binary hypothesis test is

$$f(\boldsymbol{x}) = \mathbb{1}_{\{\mathcal{L}(\boldsymbol{x}) \geq \eta\}}(\boldsymbol{x})$$

- Indicator function of set $\mathcal{S}$: $\quad \mathbb{1}_{\mathcal{S}}(s) = \begin{cases} 1 & , \text{ if } s \in \mathcal{S} \\ 0 & , \text{ if } s \notin \mathcal{S} \end{cases}$

- *Likelihood ratio:* $\quad \mathcal{L}(\boldsymbol{x}) = \dfrac{f_{\boldsymbol{X}|H_1}\big(\boldsymbol{x} \,\big|\, H_1\big)}{f_{\boldsymbol{X}|H_0}\big(\boldsymbol{x} \,\big|\, H_0\big)}$

Likelihood ratio test

$$\underset{f:\mathbb{R}^d\to\{0,1\}}{\text{minimize}} \ \mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}),Y\big)\Big]$$

The decision that minimizes the risk in a binary hypothesis test is

$$f(\boldsymbol{x}) = \mathbb{1}_{\{\mathcal{L}(\boldsymbol{x})\geq\eta\}}(\boldsymbol{x})$$

- Indicator function of set $\mathcal{S}$: $\mathbb{1}_{\mathcal{S}}(s) = \begin{cases} 1 & , \text{ if } s\in\mathcal{S} \\ 0 & , \text{ if } s\notin\mathcal{S} \end{cases}$
- *Likelihood ratio:* $\mathcal{L}(\boldsymbol{x}) = \dfrac{f_{\boldsymbol{X}|H_1}\big(\boldsymbol{x}\,\big|\,H_1\big)}{f_{\boldsymbol{X}|H_0}\big(\boldsymbol{x}\,\big|\,H_0\big)}$
- *Decision threshold:* $\eta = \dfrac{\ell(1,0)-\ell(0,0)}{\ell(0,1)-\ell(1,1)}\cdot\dfrac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)}$

Example in $\mathbb{R}$

Example in $\mathbb{R}$

$H_0 \,:\, X = W$

$H_1 \,:\, X = c + W$

Example in $\mathbb{R}$

$H_0 : X = W$

no aircraft/tumor/spam

innocent defendant

$H_1 : X = c + W$

Example in $\mathbb{R}$

$H_0 : X = W$

no aircraft/tumor/spam

innocent defendant

$H_1 : X = c + W$

aircraft/tumor/spam

guilty defendant

Example in $\mathbb{R}$

$H_0 : X = W$

no aircraft/tumor/spam

innocent defendant

$H_1 : X = c + W$

aircraft/tumor/spam

guilty defendant

$W \sim \mathcal{N}(0, 1)$

Example in $\mathbb{R}$

$H_0 \,:\, X = W$

no aircraft/tumor/spam

innocent defendant

$H_1 \,:\, X = c + W$

aircraft/tumor/spam

guilty defendant

$W \sim \mathcal{N}(0,1)$

$$f_W(w) = \frac{1}{\sqrt{2\pi}} e^{-\frac{w^2}{2}}$$

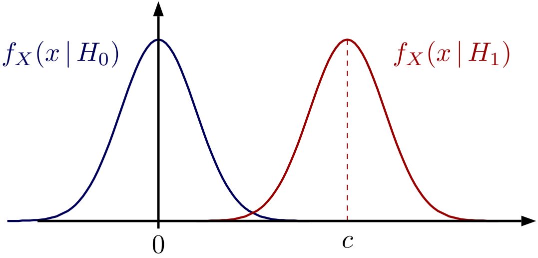
Example in $\mathbb{R}$

$H_0 : X = W$
no aircraft/tumor/spam
innocent defendant

$H_1 : X = c + W$
aircraft/tumor/spam
guilty defendant

$W \sim \mathcal{N}(0, 1)$

$$f_W(w) = \frac{1}{\sqrt{2\pi}} e^{-\frac{w^2}{2}}$$



Example in $\mathbb{R}$

Assume

Example in $\mathbb{R}$

Assume

- $c = 1$

Example in $\mathbb{R}$

Assume

- $c = 1$
- Loss values

True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true	0	1
H_1 is true	25	0

Example in $\mathbb{R}$

Assume

- $c = 1$
- Loss values

True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true	0	1
H_1 is true	25	0

- Base rates: $\mathbb{P}(H_0) = 0.95$ $\mathbb{P}(H_1) = 0.05$

Example in $\mathbb{R}$

Assume

- $c = 1$
- Loss values

True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true	0	1
H_1 is true	25	0

- Base rates: $\mathbb{P}(H_0) = 0.95$ $\mathbb{P}(H_1) = 0.05$

Compute the decision threshold

Example in $\mathbb{R}$

Decision threshold occurs for

Example in $\mathbb{R}$

Decision threshold occurs for

$$\mathcal{L}(x) = \eta$$

Example in $\mathbb{R}$

Decision threshold occurs for

$$\mathcal{L}(x) = \eta \qquad \Longleftrightarrow \qquad \log\, \mathcal{L}(x) = \log\, \eta$$

Example in $\mathbb{R}$

Decision threshold occurs for

$$\mathcal{L}(x) = \eta \qquad \Longleftrightarrow \qquad \log\, \mathcal{L}(x) = \log\, \eta$$

with

$$\mathcal{L}(x)$$

Example in $\mathbb{R}$

Decision threshold occurs for

$$\mathcal{L}(x) = \eta \qquad \Longleftrightarrow \qquad \log\, \mathcal{L}(x) = \log\, \eta$$

with

$$\mathcal{L}(x) \;=\; \frac{f_{X\,|\,H_1}\big(x \,\big|\, H_1\big)}{f_{X\,|\,H_0}\big(x \,\big|\, H_0\big)}$$

Example in $\mathbb{R}$

Decision threshold occurs for

$$\mathcal{L}(x) = \eta \qquad \Longleftrightarrow \qquad \log\, \mathcal{L}(x) = \log\, \eta$$

with

$$\mathcal{L}(x) \;=\; \frac{f_{X\,|\,H_1}\big(x \,\big|\, H_1\big)}{f_{X\,|\,H_0}\big(x \,\big|\, H_0\big)} \;=\; \frac{\exp\Big(-\frac{(x-1)^2}{2}\Big)}{\exp\Big(-\frac{x^2}{2}\Big)}$$

Example in $\mathbb{R}$

Decision threshold occurs for

$$\mathcal{L}(x) = \eta \qquad \iff \qquad \log\, \mathcal{L}(x) = \log\, \eta$$

with

$$\mathcal{L}(x) = \frac{f_{X\,|\,H_1}\big(x \,\big|\, H_1\big)}{f_{X\,|\,H_0}\big(x \,\big|\, H_0\big)} = \frac{\exp\left(-\frac{(x-1)^2}{2}\right)}{\exp\left(-\frac{x^2}{2}\right)} = \exp\left(x - \frac{1}{2}\right)$$

Example in $\mathbb{R}$

Decision threshold occurs for

$$\mathcal{L}(x) = \eta \qquad \Longleftrightarrow \qquad \log\,\mathcal{L}(x) = \log\,\eta$$

with

$$\mathcal{L}(x) = \frac{f_{X\,|\,H_1}\big(x \,\big|\, H_1\big)}{f_{X\,|\,H_0}\big(x \,\big|\, H_0\big)} = \frac{\exp\Big(-\frac{(x-1)^2}{2}\Big)}{\exp\Big(-\frac{x^2}{2}\Big)} = \exp\Big(x - \frac{1}{2}\Big)$$

$$\eta = \frac{\ell(0,0) - \ell(1,0)}{\ell(1,1) - \ell(0,1)} \cdot \frac{\mathbb{P}\big(H_0\big)}{\mathbb{P}\big(H_1\big)}$$

Example in $\mathbb{R}$

Decision threshold occurs for

$$\mathcal{L}(x) = \eta \qquad \Longleftrightarrow \qquad \log\, \mathcal{L}(x) = \log\, \eta$$

with

$$\mathcal{L}(x) \;=\; \frac{f_{X\,|\,H_1}\big(x \,\big|\, H_1\big)}{f_{X\,|\,H_0}\big(x \,\big|\, H_0\big)} \;=\; \frac{\exp\Big(-\frac{(x-1)^2}{2}\Big)}{\exp\Big(-\frac{x^2}{2}\Big)} \;=\; \exp\Big(x - \frac{1}{2}\Big)$$

$$\eta \;=\; \frac{\ell(0,0)-\ell(1,0)}{\ell(1,1)-\ell(0,1)} \cdot \frac{\mathbb{P}\big(H_0\big)}{\mathbb{P}\big(H_1\big)} \;=\; \frac{0-1}{0-25} \cdot \frac{0.95}{0.05}$$

Example in $\mathbb{R}$

Decision threshold occurs for

$$\mathcal{L}(x) = \eta \qquad \iff \qquad \log\, \mathcal{L}(x) = \log\, \eta$$

with

$$\mathcal{L}(x) \;=\; \frac{f_{X \mid H_1}\big(x \,\big|\, H_1\big)}{f_{X \mid H_0}\big(x \,\big|\, H_0\big)} \;=\; \frac{\exp\Big(-\frac{(x-1)^2}{2}\Big)}{\exp\Big(-\frac{x^2}{2}\Big)} \;=\; \exp\Big(x - \frac{1}{2}\Big)$$

$$\eta \;=\; \frac{\ell(0,0) - \ell(1,0)}{\ell(1,1) - \ell(0,1)} \cdot \frac{\mathbb{P}\big(H_0\big)}{\mathbb{P}\big(H_1\big)} \;=\; \frac{0-1}{0-25} \cdot \frac{0.95}{0.05} \;\simeq\; 0.76$$

Example in $\mathbb{R}$

Decision threshold occurs for

$$\mathcal{L}(x) = \eta \qquad \Longleftrightarrow \qquad \log\,\mathcal{L}(x) = \log\,\eta$$

with

$$\mathcal{L}(x) \;=\; \frac{f_{X\,|\,H_1}\big(x \,\big|\, H_1\big)}{f_{X\,|\,H_0}\big(x \,\big|\, H_0\big)} \;=\; \frac{\exp\Big(-\frac{(x-1)^2}{2}\Big)}{\exp\Big(-\frac{x^2}{2}\Big)} \;=\; \exp\Big(x - \frac{1}{2}\Big)$$

$$\eta \;=\; \frac{\ell(0,0) - \ell(1,0)}{\ell(1,1) - \ell(0,1)} \cdot \frac{\mathbb{P}\big(H_0\big)}{\mathbb{P}\big(H_1\big)} \;=\; \frac{0-1}{0-25} \cdot \frac{0.95}{0.05} \;\simeq\; 0.76$$

The decision threshold is then

Example in $\mathbb{R}$

Decision threshold occurs for

$$\mathcal{L}(x) = \eta \qquad \Longleftrightarrow \qquad \log\, \mathcal{L}(x) = \log\, \eta$$

with

$$\mathcal{L}(x) \;=\; \frac{f_{X\,|\,H_1}\big(x \,\big|\, H_1\big)}{f_{X\,|\,H_0}\big(x \,\big|\, H_0\big)} \;=\; \frac{\exp\Big(-\frac{(x-1)^2}{2}\Big)}{\exp\Big(-\frac{x^2}{2}\Big)} \;=\; \exp\Big(x - \frac{1}{2}\Big)$$

$$\eta \;=\; \frac{\ell(0,0) - \ell(1,0)}{\ell(1,1) - \ell(0,1)} \cdot \frac{\mathbb{P}\big(H_0\big)}{\mathbb{P}\big(H_1\big)} \;=\; \frac{0-1}{0-25} \cdot \frac{0.95}{0.05} \;\simeq\; 0.76$$

The decision threshold is then

$$x - \frac{1}{2} = \log\, 0.76$$

Example in $\mathbb{R}$

Decision threshold occurs for

$$\mathcal{L}(x) = \eta \qquad \Longleftrightarrow \qquad \log\, \mathcal{L}(x) = \log\, \eta$$

with

$$\mathcal{L}(x) \;=\; \frac{f_{X\,|\,H_1}\big(x \,\big|\, H_1\big)}{f_{X\,|\,H_0}\big(x \,\big|\, H_0\big)} \;=\; \frac{\exp\Big(-\frac{(x-1)^2}{2}\Big)}{\exp\Big(-\frac{x^2}{2}\Big)} \;=\; \exp\Big(x - \frac{1}{2}\Big)$$

$$\eta \;=\; \frac{\ell(0,0) - \ell(1,0)}{\ell(1,1) - \ell(0,1)} \cdot \frac{\mathbb{P}\big(H_0\big)}{\mathbb{P}\big(H_1\big)} \;=\; \frac{0-1}{0-25} \cdot \frac{0.95}{0.05} \;\simeq\; 0.76$$

The decision threshold is then

$$x - \frac{1}{2} = \log\, 0.76 \qquad \Longleftrightarrow \qquad x \simeq 0.23$$

Example in $\mathbb{R}$

$c = 1$

True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true	0	1
H_1 is true	25	0

$\mathbb{P}(H_0) = 0.95$

$\mathbb{P}(H_1) = 0.05$

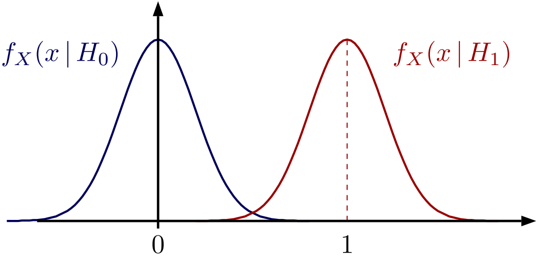
Example in $\mathbb{R}$

$c = 1$

True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true	0	1
H_1 is true	25	0

$\mathbb{P}(H_0) = 0.95$

$\mathbb{P}(H_1) = 0.05$



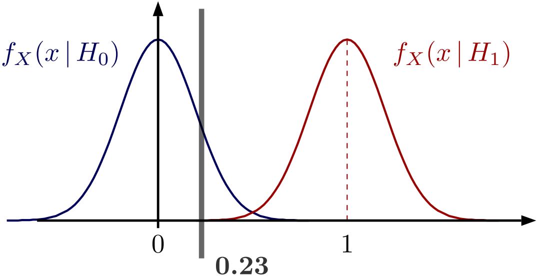
Example in $\mathbb{R}$

$c = 1$

True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true	0	1
H_1 is true	25	0

$\mathbb{P}(H_0) = 0.95$

$\mathbb{P}(H_1) = 0.05$



Particular cases

Particular cases

Recall the problem:

$$\underset{f:\mathbb{R}^d \to \{0,1\}}{\text{minimize}} \;\; \mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}), Y\big)\Big]$$

Particular cases

Recall the problem:

$$\underset{f:\mathbb{R}^d \to \{0,1\}}{\text{minimize}} \ \mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}), Y\big)\Big]$$

Expected value is w.r.t. *joint distribution* $\mathbb{P}_{\boldsymbol{X}Y}$

Particular cases

Recall the problem:

$$\underset{f:\mathbb{R}^d \to \{0,1\}}{\text{minimize}} \ \mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}), Y\big)\Big]$$

Expected value is w.r.t. *joint distribution* $\mathbb{P}_{\boldsymbol{X}Y}$

When class $Y \in \{0,1\}$ is viewed as a parameter of $\mathbb{P}_{\boldsymbol{X}Y}$ to estimate,

Particular cases

Recall the problem:

$$\underset{f:\mathbb{R}^d \to \{0,1\}}{\text{minimize}} \;\; \mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}), Y\big)\Big]$$

Expected value is w.r.t. *joint distribution* $\mathbb{P}_{\boldsymbol{X}Y}$

When class $Y \in \{0,1\}$ is viewed as a parameter of $\mathbb{P}_{\boldsymbol{X}Y}$ to estimate,

- Maximum a posteriori (MAP)
- Maximum likelihood (ML)

can be seen as *likelihood ratio tests*

Maximum a posteriori (MAP)

Maximum a posteriori (MAP)

Consider $\ell(0,\, 0) = \ell(1,\, 1) = 0$ and $\ell(1,\, 0) = \ell(0,\, 1) = 1$.

Maximum a posteriori (MAP)

Consider $\ell(0,\,0) = \ell(1,\,1) = 0$ and $\ell(1,\,0) = \ell(0,\,1) = 1$. Then,

$$\mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}), Y\big)\Big]$$

Maximum a posteriori (MAP)

Consider $\ell(0,\, 0) = \ell(1,\, 1) = 0$ and $\ell(1,\, 0) = \ell(0,\, 1) = 1$. Then,

$$\mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}), Y\big)\Big] = \int_{\mathbb{R}^d} \mathbb{E}_Y\left[\ell\big(f(\boldsymbol{X}),\, Y\big) \,\Big|\, \boldsymbol{X} = \boldsymbol{x}\right] f_{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$$

Maximum a posteriori (MAP)

Consider $\ell(0,\,0) = \ell(1,\,1) = 0$ and $\ell(1,\,0) = \ell(0,\,1) = 1$. Then,

$$\mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}), Y\big)\Big] = \int_{\mathbb{R}^d} \mathbb{E}_Y\left[\ell\big(f(\boldsymbol{X}),\, Y\big)\,\Big|\,\boldsymbol{X} = \boldsymbol{x}\right] f_{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$$

If $f(\boldsymbol{x}) = 0$,

Maximum a posteriori (MAP)

Consider $\ell(0,\,0) = \ell(1,\,1) = 0$ and $\ell(1,\,0) = \ell(0,\,1) = 1$. Then,

$$\mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}), Y\big)\Big] = \int_{\mathbb{R}^d} \mathbb{E}_Y\left[\ell\big(f(\boldsymbol{X}),\, Y\big)\,\Big|\,\boldsymbol{X} = \boldsymbol{x}\right] f_{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$$

If $f(\boldsymbol{x}) = 0$, $\qquad \mathbb{E}_Y\Big[\ell\big(0,\,Y\big)\,\big|\,\boldsymbol{X} = \boldsymbol{x}\Big] = \mathbb{P}\big(Y = 1\,|\,\boldsymbol{X} = \boldsymbol{x}\big)$

Maximum a posteriori (MAP)

Consider $\ell(0,\,0) = \ell(1,\,1) = 0$ and $\ell(1,\,0) = \ell(0,\,1) = 1$. Then,

$$\mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}), Y\big)\Big] = \int_{\mathbb{R}^d} \mathbb{E}_Y\left[\ell\big(f(\boldsymbol{X}),\, Y\big)\;\Big|\; \boldsymbol{X} = \boldsymbol{x}\right] f_{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$$

If $f(\boldsymbol{x}) = 0$, $\qquad \mathbb{E}_Y\Big[\ell\big(0,\,Y\big)\,\big|\,\boldsymbol{X} = \boldsymbol{x}\Big] = \mathbb{P}\big(Y = 1 \,|\, \boldsymbol{X} = \boldsymbol{x}\big)$

If $f(\boldsymbol{x}) = 1$,

Maximum a posteriori (MAP)

Consider $\ell(0,\,0) = \ell(1,\,1) = 0$ and $\ell(1,\,0) = \ell(0,\,1) = 1$. Then,

$$\mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}), Y\big)\Big] = \int_{\mathbb{R}^d} \mathbb{E}_Y\left[\ell\big(f(\boldsymbol{X}),\, Y\big)\,\Big|\,\boldsymbol{X} = \boldsymbol{x}\right] f_{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$$

If $f(\boldsymbol{x}) = 0$, $\qquad \mathbb{E}_Y\Big[\ell\big(0,\, Y\big)\,\big|\,\boldsymbol{X} = \boldsymbol{x}\Big] = \mathbb{P}\big(Y = 1\,|\,\boldsymbol{X} = \boldsymbol{x}\big)$

If $f(\boldsymbol{x}) = 1$, $\qquad \mathbb{E}_Y\Big[\ell\big(1,\, Y\big)\,\big|\,\boldsymbol{X} = \boldsymbol{x}\Big] = \mathbb{P}\big(Y = 0\,|\,\boldsymbol{X} = \boldsymbol{x}\big)$

Maximum a posteriori (MAP)

Consider $\ell(0,\,0) = \ell(1,\,1) = 0$ and $\ell(1,\,0) = \ell(0,\,1) = 1$. Then,

$$\mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}), Y\big)\Big] = \int_{\mathbb{R}^d} \mathbb{E}_Y\Big[\ell\big(f(\boldsymbol{X}),\,Y\big)\,\Big|\,\boldsymbol{X} = \boldsymbol{x}\Big] f_{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$$

If $f(\boldsymbol{x}) = 0$, $\qquad \mathbb{E}_Y\Big[\ell\big(0,\,Y\big)\,\big|\,\boldsymbol{X} = \boldsymbol{x}\Big] = \mathbb{P}\big(Y = 1\,|\,\boldsymbol{X} = \boldsymbol{x}\big)$

If $f(\boldsymbol{x}) = 1$, $\qquad \mathbb{E}_Y\Big[\ell\big(1,\,Y\big)\,\big|\,\boldsymbol{X} = \boldsymbol{x}\Big] = \mathbb{P}\big(Y = 0\,|\,\boldsymbol{X} = \boldsymbol{x}\big)$

So, select $f(\boldsymbol{x}) = 1$ if

Maximum a posteriori (MAP)

Consider $\ell(0,\,0) = \ell(1,\,1) = 0$ and $\ell(1,\,0) = \ell(0,\,1) = 1$. Then,

$$\mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}),Y\big)\Big] = \int_{\mathbb{R}^d} \mathbb{E}_Y\Big[\ell\big(f(\boldsymbol{X}),\,Y\big)\,\Big|\,\boldsymbol{X}=\boldsymbol{x}\Big] f_{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$$

If $f(\boldsymbol{x}) = 0$, $\qquad \mathbb{E}_Y\Big[\ell\big(0,\,Y\big)\,\big|\,\boldsymbol{X}=\boldsymbol{x}\Big] = \mathbb{P}\big(Y=1\,|\,\boldsymbol{X}=\boldsymbol{x}\big)$

If $f(\boldsymbol{x}) = 1$, $\qquad \mathbb{E}_Y\Big[\ell\big(1,\,Y\big)\,\big|\,\boldsymbol{X}=\boldsymbol{x}\Big] = \mathbb{P}\big(Y=0\,|\,\boldsymbol{X}=\boldsymbol{x}\big)$

So, select $f(\boldsymbol{x}) = 1$ if $\mathbb{P}\big(Y=1\,|\,\boldsymbol{X}=\boldsymbol{x}\big) \geq \mathbb{P}\big(Y=0\,|\,\boldsymbol{X}=\boldsymbol{x}\big)$

So, select $f(\boldsymbol{x}) = 1$ if $\mathbb{P}\big(Y = 1 \,|\, \boldsymbol{X} = \boldsymbol{x}\big) \geq \mathbb{P}\big(Y = 0 \,|\, \boldsymbol{X} = \boldsymbol{x}\big)$

So, select $f(\boldsymbol{x}) = 1$ if $\mathbb{P}\big(Y = 1 \,|\, \boldsymbol{X} = \boldsymbol{x}\big) \geq \mathbb{P}\big(Y = 0 \,|\, \boldsymbol{X} = \boldsymbol{x}\big)$

That is,

So, select $f(\boldsymbol{x}) = 1$ if $\mathbb{P}\big(Y = 1 \,|\, \boldsymbol{X} = \boldsymbol{x}\big) \geq \mathbb{P}\big(Y = 0 \,|\, \boldsymbol{X} = \boldsymbol{x}\big)$

That is, $$f(\boldsymbol{x}) = \arg\max_{i} \; \mathbb{P}\big(Y = i \,\big|\, \boldsymbol{X} = \boldsymbol{x}\big)$$

So, select $f(\boldsymbol{x}) = 1$ if $\mathbb{P}\big(Y = 1 \,|\, \boldsymbol{X} = \boldsymbol{x}\big) \geq \mathbb{P}\big(Y = 0 \,|\, \boldsymbol{X} = \boldsymbol{x}\big)$

That is,
$$f(\boldsymbol{x}) = \arg\max_i \; \underbrace{\boxed{\mathbb{P}\big(Y = i \,\big|\, \boldsymbol{X} = \boldsymbol{x}\big)}}_{\textit{posterior}}$$

So, select $f(\boldsymbol{x}) = 1$ if $\mathbb{P}\big(Y = 1 \,|\, \boldsymbol{X} = \boldsymbol{x}\big) \geq \mathbb{P}\big(Y = 0 \,|\, \boldsymbol{X} = \boldsymbol{x}\big)$

That is,
$$f(\boldsymbol{x}) = \arg\max_i \; \underbrace{\mathbb{P}\big(Y = i \,\big|\, \boldsymbol{X} = \boldsymbol{x}\big)}_{\textbf{\textit{posterior}}}$$

This is a likelihood ratio test, because

So, select $f(\boldsymbol{x}) = 1$ if $\mathbb{P}\big(Y = 1 \,|\, \boldsymbol{X} = \boldsymbol{x}\big) \geq \mathbb{P}\big(Y = 0 \,|\, \boldsymbol{X} = \boldsymbol{x}\big)$

That is, $$f(\boldsymbol{x}) = \arg\max_i \; \underbrace{\mathbb{P}\big(Y = i \,\big|\, \boldsymbol{X} = \boldsymbol{x}\big)}_{\textbf{\textit{posterior}}}$$

This is a likelihood ratio test, because

$$\mathbb{P}\big(Y = 1 \,|\, \boldsymbol{X} = \boldsymbol{x}\big) \geq \mathbb{P}\big(Y = 0 \,|\, \boldsymbol{X} = \boldsymbol{x}\big)$$

So, select $f(\boldsymbol{x}) = 1$ if $\mathbb{P}\big(Y = 1 \,|\, \boldsymbol{X} = \boldsymbol{x}\big) \geq \mathbb{P}\big(Y = 0 \,|\, \boldsymbol{X} = \boldsymbol{x}\big)$

That is, $$f(\boldsymbol{x}) = \arg\max_i \; \underbrace{\boxed{\mathbb{P}\big(Y = i \,\big|\, \boldsymbol{X} = \boldsymbol{x}\big)}}_{\textbf{\textit{posterior}}}$$

This is a likelihood ratio test, because

$$\mathbb{P}\big(Y = 1 \,|\, \boldsymbol{X} = \boldsymbol{x}\big) \geq \mathbb{P}\big(Y = 0 \,|\, \boldsymbol{X} = \boldsymbol{x}\big)$$

$$\iff \qquad f_{\boldsymbol{X}|H_1}(x \,|\, H_1) \cdot \mathbb{P}(H_1) \geq f_{\boldsymbol{X}|H_0}(x \,|\, H_0) \cdot \mathbb{P}(H_0)$$

So, select $f(\boldsymbol{x}) = 1$ if $\mathbb{P}\big(Y = 1 \,|\, \boldsymbol{X} = \boldsymbol{x}\big) \geq \mathbb{P}\big(Y = 0 \,|\, \boldsymbol{X} = \boldsymbol{x}\big)$

That is,

$$f(\boldsymbol{x}) = \arg\max_i \underbrace{\boxed{\mathbb{P}\big(Y = i \,\big|\, \boldsymbol{X} = \boldsymbol{x}\big)}}_{\textbf{\textit{posterior}}}$$

This is a likelihood ratio test, because

$$\mathbb{P}\big(Y = 1 \,|\, \boldsymbol{X} = \boldsymbol{x}\big) \geq \mathbb{P}\big(Y = 0 \,|\, \boldsymbol{X} = \boldsymbol{x}\big)$$

$$\iff \quad f_{\boldsymbol{X}|H_1}(x \,|\, H_1) \cdot \mathbb{P}(H_1) \geq f_{\boldsymbol{X}|H_0}(x \,|\, H_0) \cdot \mathbb{P}(H_0)$$

$$\iff \quad \frac{f_{\boldsymbol{X}|H_1}(x \,|\, H_1)}{f_{\boldsymbol{X}|H_0}(x \,|\, H_0)} \geq \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)}$$

So, select $f(\boldsymbol{x}) = 1$ if $\mathbb{P}\big(Y = 1 \,|\, \boldsymbol{X} = \boldsymbol{x}\big) \ge \mathbb{P}\big(Y = 0 \,|\, \boldsymbol{X} = \boldsymbol{x}\big)$

That is, $$f(\boldsymbol{x}) = \arg\max_i \underbrace{\mathbb{P}\big(Y = i \,\big|\, \boldsymbol{X} = \boldsymbol{x}\big)}_{\textbf{\textit{posterior}}}$$

This is a likelihood ratio test, because

$$\mathbb{P}\big(Y = 1 \,|\, \boldsymbol{X} = \boldsymbol{x}\big) \ge \mathbb{P}\big(Y = 0 \,|\, \boldsymbol{X} = \boldsymbol{x}\big)$$

$$\iff \quad f_{\boldsymbol{X}|H_1}(x \,|\, H_1) \cdot \mathbb{P}(H_1) \ge f_{\boldsymbol{X}|H_0}(x \,|\, H_0) \cdot \mathbb{P}(H_0)$$

$$\iff \quad \mathcal{L}(\boldsymbol{x}) = \frac{f_{\boldsymbol{X}|H_1}(x \,|\, H_1)}{f_{\boldsymbol{X}|H_0}(x \,|\, H_0)} \ge \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)}$$

So, select $f(\boldsymbol{x}) = 1$ if $\mathbb{P}\big(Y = 1 \,|\, \boldsymbol{X} = \boldsymbol{x}\big) \geq \mathbb{P}\big(Y = 0 \,|\, \boldsymbol{X} = \boldsymbol{x}\big)$

That is, $$f(\boldsymbol{x}) = \arg\max_i \; \underbrace{\mathbb{P}\big(Y = i \,\big|\, \boldsymbol{X} = \boldsymbol{x}\big)}_{\textbf{\textit{posterior}}}$$

This is a likelihood ratio test, because

$$\mathbb{P}\big(Y = 1 \,|\, \boldsymbol{X} = \boldsymbol{x}\big) \geq \mathbb{P}\big(Y = 0 \,|\, \boldsymbol{X} = \boldsymbol{x}\big)$$

$$\iff \quad f_{\boldsymbol{X}|H_1}(x \,|\, H_1) \cdot \mathbb{P}(H_1) \geq f_{\boldsymbol{X}|H_0}(x \,|\, H_0) \cdot \mathbb{P}(H_0)$$

$$\iff \quad \mathcal{L}(\boldsymbol{x}) = \frac{f_{\boldsymbol{X}|H_1}(x \,|\, H_1)}{f_{\boldsymbol{X}|H_0}(x \,|\, H_0)} \geq \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)} = \eta$$

So, select $f(\boldsymbol{x}) = 1$ if $\mathbb{P}\big(Y = 1 \,|\, \boldsymbol{X} = \boldsymbol{x}\big) \geq \mathbb{P}\big(Y = 0 \,|\, \boldsymbol{X} = \boldsymbol{x}\big)$

That is, $$f(\boldsymbol{x}) = \arg\max_i \underbrace{\mathbb{P}\big(Y = i \,\big|\, \boldsymbol{X} = \boldsymbol{x}\big)}_{\textit{posterior}}$$

This is a likelihood ratio test, because

$$\mathbb{P}\big(Y = 1 \,|\, \boldsymbol{X} = \boldsymbol{x}\big) \geq \mathbb{P}\big(Y = 0 \,|\, \boldsymbol{X} = \boldsymbol{x}\big)$$

$$\iff \quad f_{\boldsymbol{X}|H_1}(x \,|\, H_1) \cdot \mathbb{P}(H_1) \geq f_{\boldsymbol{X}|H_0}(x \,|\, H_0) \cdot \mathbb{P}(H_0)$$

$$\iff \quad \mathcal{L}(\boldsymbol{x}) = \frac{f_{\boldsymbol{X}|H_1}(x \,|\, H_1)}{f_{\boldsymbol{X}|H_0}(x \,|\, H_0)} \geq \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)} = \eta$$

Recall that MAP rule minimizes probability of incorrect decision:

So, select $f(\boldsymbol{x}) = 1$ if $\mathbb{P}\big(Y = 1 \,|\, \boldsymbol{X} = \boldsymbol{x}\big) \geq \mathbb{P}\big(Y = 0 \,|\, \boldsymbol{X} = \boldsymbol{x}\big)$

That is, $$f(\boldsymbol{x}) = \arg\max_i \underbrace{\boxed{\mathbb{P}\big(Y = i \,\big|\, \boldsymbol{X} = \boldsymbol{x}\big)}}_{\textbf{\textit{posterior}}}$$

This is a likelihood ratio test, because

$$\mathbb{P}\big(Y = 1 \,|\, \boldsymbol{X} = \boldsymbol{x}\big) \geq \mathbb{P}\big(Y = 0 \,|\, \boldsymbol{X} = \boldsymbol{x}\big)$$

$$\iff \quad f_{\boldsymbol{X}|H_1}(x \,|\, H_1) \cdot \mathbb{P}(H_1) \geq f_{\boldsymbol{X}|H_0}(x \,|\, H_0) \cdot \mathbb{P}(H_0)$$

$$\iff \quad \mathcal{L}(\boldsymbol{x}) = \frac{f_{\boldsymbol{X}|H_1}(x \,|\, H_1)}{f_{\boldsymbol{X}|H_0}(x \,|\, H_0)} \geq \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)} = \eta$$

Recall that MAP rule minimizes probability of incorrect decision:

$$\mathbb{P}\big(\text{error}\big) = \mathbb{P}\big(f(\boldsymbol{X}) = 1\,,\ H_0\big) + \mathbb{P}\big(f(\boldsymbol{X}) = 0\,,\ H_1\big)$$

Maximum Likelihood (ML)

Maximum Likelihood (ML)

Consider $\ell(0,\,0) = \ell(1,\,1) = 0$ and $\ell(1,\,0) = \ell(0,\,1) = 1$

Maximum Likelihood (ML)

Consider $\ell(0,\, 0) = \ell(1,\, 1) = 0$ and $\ell(1,\, 0) = \ell(0,\, 1) = 1$

And $\mathbb{P}(H_0) = \mathbb{P}(H_1) = \frac{1}{2}$

Maximum Likelihood (ML)

Consider $\ell(0, 0) = \ell(1, 1) = 0$ and $\ell(1, 0) = \ell(0, 1) = 1$

And $\mathbb{P}(H_0) = \mathbb{P}(H_1) = \frac{1}{2}$

The optimal decision (MAP) is

Maximum Likelihood (ML)

Consider $\ell(0, 0) = \ell(1, 1) = 0$ and $\ell(1, 0) = \ell(0, 1) = 1$

And $\mathbb{P}(H_0) = \mathbb{P}(H_1) = \frac{1}{2}$

The optimal decision (MAP) is

$$f(\boldsymbol{x}) = \arg\max_i \; \mathbb{P}\big(Y = i \,\big|\, \boldsymbol{X} = \boldsymbol{x}\big)$$

Maximum Likelihood (ML)

Consider $\ell(0,\,0) = \ell(1,\,1) = 0$ and $\ell(1,\,0) = \ell(0,\,1) = 1$

And $\mathbb{P}(H_0) = \mathbb{P}(H_1) = \frac{1}{2}$

The optimal decision (MAP) is

$$f(\boldsymbol{x}) \;=\; \arg\max_i \;\; \mathbb{P}\big(Y = i \,\big|\, \boldsymbol{X} = \boldsymbol{x}\big)$$

$$=\; \arg\max_i \;\; \frac{f_{\boldsymbol{X}|Y}(\boldsymbol{x} \,|\, Y = i) \cdot \mathbb{P}(Y = i)}{f_{\boldsymbol{X}}(\boldsymbol{x})} \qquad \Big(\text{Bayes rule}\Big)$$

Maximum Likelihood (ML)

Consider $\ell(0,\,0) = \ell(1,\,1) = 0$ and $\ell(1,\,0) = \ell(0,\,1) = 1$

And $\mathbb{P}(H_0) = \mathbb{P}(H_1) = \frac{1}{2}$

The optimal decision (MAP) is

$$
\begin{aligned}
f(\boldsymbol{x}) &= \arg\max_i \ \mathbb{P}\big(Y = i \,\big|\, \boldsymbol{X} = \boldsymbol{x}\big) \\
&= \arg\max_i \ \frac{f_{\boldsymbol{X}|Y}(\boldsymbol{x} \,|\, Y = i) \cdot \mathbb{P}(Y = i)}{f_{\boldsymbol{X}}(\boldsymbol{x})} && \Big(\text{Bayes rule}\Big) \\
&= \arg\max_i \ f_{\boldsymbol{X}|Y}\Big(\boldsymbol{x} \,\Big|\, Y = i\Big) && \left(\mathbb{P}(Y = i) = \frac{1}{2}\right)
\end{aligned}
$$

Maximum Likelihood (ML)

Consider $\ell(0,\, 0) = \ell(1,\, 1) = 0$ and $\ell(1,\, 0) = \ell(0,\, 1) = 1$

And $\mathbb{P}(H_0) = \mathbb{P}(H_1) = \frac{1}{2}$

The optimal decision (MAP) is

$$f(\boldsymbol{x}) = \arg\max_i \; \mathbb{P}\big(Y = i \,\big|\, \boldsymbol{X} = \boldsymbol{x}\big)$$

$$= \arg\max_i \; \frac{f_{\boldsymbol{X}|Y}(\boldsymbol{x} \,|\, Y = i) \cdot \mathbb{P}(Y = i)}{f_{\boldsymbol{X}}(\boldsymbol{x})} \qquad \Big(\text{Bayes rule}\Big)$$

$$= \underbrace{\arg\max_i \; f_{\boldsymbol{X}|Y}\Big(\boldsymbol{x} \,\Big|\, Y = i\Big)}_{\textit{maximum likelihood}} \qquad \left(\mathbb{P}(Y = i) = \frac{1}{2}\right)$$

Maximum Likelihood (ML)

Consider $\ell(0,\,0) = \ell(1,\,1) = 0$ and $\ell(1,\,0) = \ell(0,\,1) = 1$

And $\mathbb{P}(H_0) = \mathbb{P}(H_1) = \frac{1}{2}$

The optimal decision (MAP) is

$$f(\boldsymbol{x}) = \arg\max_i \ \mathbb{P}\big(Y = i \,\big|\, \boldsymbol{X} = \boldsymbol{x}\big)$$

$$= \arg\max_i \ \frac{f_{\boldsymbol{X}|Y}(\boldsymbol{x} \,|\, Y = i) \cdot \mathbb{P}(Y = i)}{f_{\boldsymbol{X}}(\boldsymbol{x})} \qquad \Big(\text{Bayes rule}\Big)$$

$$= \underbrace{\boxed{\arg\max_i \ f_{\boldsymbol{X}|Y}\Big(\boldsymbol{x} \,\Big|\, Y = i\Big)}}_{\textit{maximum likelihood}} \qquad \left(\mathbb{P}\big(Y = i\big) = \frac{1}{2}\right)$$

This corresponds to a likelihood ratio test with $\eta = 1$

Types of errors and successes

Types of errors and successes

Table of probabilities

True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true		
H_1 is true		

Types of errors and successes

True Positive Rate (TPR)

Table of probabilities

True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true		
H_1 is true		

Types of errors and successes

True Positive Rate (TPR)
power, sensitivity, recall

Table of probabilities

True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true		
H_1 is true		

Types of errors and successes

True Positive Rate (TPR)
power, sensitivity, recall

Table of probabilities		
True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true		
H_1 is true		TPR

Types of errors and successes

True Positive Rate (TPR)
power, sensitivity, recall

$$\text{TPR} = \mathbb{P}\big(f(\boldsymbol{X}) = 1 \,\big|\, H_1\big)$$

Table of probabilities

True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true		
H_1 is true		TPR

Types of errors and successes

True Positive Rate (TPR)
power, sensitivity, recall

$$\text{TPR} = \mathbb{P}\left(f(\boldsymbol{X}) = 1 \,\middle|\, H_1\right)$$

False Positive Rate (FPR)
type I error, false alarm

Table of probabilities

True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true		
H_1 is true		TPR

Types of errors and successes

True Positive Rate (TPR)
power, sensitivity, recall

$$\text{TPR} = \mathbb{P}\big(f(\boldsymbol{X}) = 1 \,\big|\, H_1\big)$$

False Positive Rate (FPR)
type I error, false alarm

Table of probabilities

True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true		FPR
H_1 is true		TPR

Types of errors and successes

True Positive Rate (TPR) $\quad \text{TPR} = \mathbb{P}\big(f(\boldsymbol{X}) = 1 \,\big|\, H_1\big)$

power, sensitivity, recall

False Positive Rate (FPR) $\quad \text{FPR} = \mathbb{P}\big(f(\boldsymbol{X}) = 1 \,\big|\, H_0\big)$

type I error, false alarm

Table of probabilities

True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true		FPR
H_1 is true		TPR

Types of errors and successes

True Positive Rate (TPR) $\quad \text{TPR} = \mathbb{P}\big(f(\boldsymbol{X}) = 1 \,\big|\, H_1\big)$

power, sensitivity, recall

False Positive Rate (FPR) $\quad \text{FPR} = \mathbb{P}\big(f(\boldsymbol{X}) = 1 \,\big|\, H_0\big)$

type I error, false alarm

True Negative Rate (TNR)

specificity

Table of probabilities

True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true		FPR
H_1 is true		TPR

Types of errors and successes

True Positive Rate (TPR) $\qquad \text{TPR} = \mathbb{P}\big(f(\boldsymbol{X}) = 1 \,\big|\, H_1\big)$

power, sensitivity, recall

False Positive Rate (FPR) $\qquad \text{FPR} = \mathbb{P}\big(f(\boldsymbol{X}) = 1 \,\big|\, H_0\big)$

type I error, false alarm

True Negative Rate (TNR)

specificity

Table of probabilities

True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true	TNR	FPR
H_1 is true		TPR

Types of errors and successes

True Positive Rate (TPR) $\quad \mathrm{TPR} = \mathbb{P}\big(f(\boldsymbol{X}) = 1 \,\big|\, H_1\big)$

power, sensitivity, recall

False Positive Rate (FPR) $\quad \mathrm{FPR} = \mathbb{P}\big(f(\boldsymbol{X}) = 1 \,\big|\, H_0\big)$

type I error, false alarm

True Negative Rate (TNR) $\quad \mathrm{TNR} = \mathbb{P}\big(f(\boldsymbol{X}) = 0 \,\big|\, H_0\big)$

specificity

Table of probabilities

True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true	TNR	FPR
H_1 is true		TPR

Types of errors and successes

True Positive Rate (TPR) $\quad \text{TPR} = \mathbb{P}\big(f(\boldsymbol{X}) = 1 \,\big|\, H_1\big)$

power, sensitivity, recall

False Positive Rate (FPR) $\quad \text{FPR} = \mathbb{P}\big(f(\boldsymbol{X}) = 1 \,\big|\, H_0\big)$

type I error, false alarm

True Negative Rate (TNR) $\quad \text{TNR} = \mathbb{P}\big(f(\boldsymbol{X}) = 0 \,\big|\, H_0\big)$

specificity

False Negative Rate (FNR)

type II error, missed detection

Table of probabilities

True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true	TNR	FPR
H_1 is true		TPR

Types of errors and successes

True Positive Rate (TPR) $\quad \text{TPR} = \mathbb{P}\big(f(\boldsymbol{X}) = 1 \,\big|\, H_1\big)$

power, sensitivity, recall

False Positive Rate (FPR) $\quad \text{FPR} = \mathbb{P}\big(f(\boldsymbol{X}) = 1 \,\big|\, H_0\big)$

type I error, false alarm

True Negative Rate (TNR) $\quad \text{TNR} = \mathbb{P}\big(f(\boldsymbol{X}) = 0 \,\big|\, H_0\big)$

specificity

False Negative Rate (FNR)

type II error, missed detection

Table of probabilities

True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true	TNR	FPR
H_1 is true	FNR	TPR

Types of errors and successes

True Positive Rate (TPR) $\quad \text{TPR} = \mathbb{P}\big(f(\boldsymbol{X}) = 1 \,\big|\, H_1\big)$

power, sensitivity, recall

False Positive Rate (FPR) $\quad \text{FPR} = \mathbb{P}\big(f(\boldsymbol{X}) = 1 \,\big|\, H_0\big)$

type I error, false alarm

True Negative Rate (TNR) $\quad \text{TNR} = \mathbb{P}\big(f(\boldsymbol{X}) = 0 \,\big|\, H_0\big)$

specificity

False Negative Rate (FNR) $\quad \text{FNR} = \mathbb{P}\big(f(\boldsymbol{X}) = 0 \,\big|\, H_1\big)$

type II error, missed detection

Table of probabilities

True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true	TNR	FPR
H_1 is true	FNR	TPR

Types of errors and successes

True Positive Rate (TPR) $\quad \text{TPR} = \mathbb{P}\big(f(\boldsymbol{X}) = 1 \,\big|\, H_1\big)$

power, sensitivity, recall

False Positive Rate (FPR) $\quad \text{FPR} = \mathbb{P}\big(f(\boldsymbol{X}) = 1 \,\big|\, H_0\big)$

type I error, false alarm

True Negative Rate (TNR) $\quad \text{TNR} = \mathbb{P}\big(f(\boldsymbol{X}) = 0 \,\big|\, H_0\big)$

specificity

False Negative Rate (FNR) $\quad \text{FNR} = \mathbb{P}\big(f(\boldsymbol{X}) = 0 \,\big|\, H_1\big)$

type II error, missed detection

Table of probabilities

True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true	TNR	FPR $\boldsymbol{\alpha}$
H_1 is true	FNR	TPR

Types of errors and successes

True Positive Rate (TPR)
power, sensitivity, recall

$$\text{TPR} = \mathbb{P}\big(f(\boldsymbol{X}) = 1 \,\big|\, H_1\big)$$

False Positive Rate (FPR)
type I error, false alarm

$$\text{FPR} = \mathbb{P}\big(f(\boldsymbol{X}) = 1 \,\big|\, H_0\big)$$

True Negative Rate (TNR)
specificity

$$\text{TNR} = \mathbb{P}\big(f(\boldsymbol{X}) = 0 \,\big|\, H_0\big)$$

False Negative Rate (FNR)
type II error, missed detection

$$\text{FNR} = \mathbb{P}\big(f(\boldsymbol{X}) = 0 \,\big|\, H_1\big)$$

Table of probabilities

True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true	TNR	FPR $\boldsymbol{\alpha}$
H_1 is true	FNR $\boldsymbol{\beta}$	TPR

Other measures

Other measures

Previous measures don't account for *base rates*: $\mathbb{P}(H_0)$, $\mathbb{P}(H_1)$

Other measures

Previous measures don't account for *base rates*: $\mathbb{P}(H_0)$, $\mathbb{P}(H_1)$

That is, they are suitable when $\mathbb{P}(H_0) \simeq \mathbb{P}(H_1)$

Other measures

Previous measures don't account for *base rates*: $\mathbb{P}(H_0)$, $\mathbb{P}(H_1)$

That is, they are suitable when $\mathbb{P}(H_0) \simeq \mathbb{P}(H_1)$

Alternatives

Other measures

Previous measures don't account for *base rates*: $\mathbb{P}(H_0)$, $\mathbb{P}(H_1)$

That is, they are suitable when $\mathbb{P}(H_0) \simeq \mathbb{P}(H_1)$

Alternatives

Precision: $\mathbb{P}\big(H_1 \,\big|\, f(\boldsymbol{X}) = 1\big)$

Other measures

Previous measures don't account for *base rates*: $\mathbb{P}(H_0)$, $\mathbb{P}(H_1)$

That is, they are suitable when $\mathbb{P}(H_0) \simeq \mathbb{P}(H_1)$

Alternatives

Precision: $$\mathbb{P}\big(H_1 \,\big|\, f(\boldsymbol{X}) = 1\big) = \frac{\mathsf{TPR} \cdot \mathbb{P}(H_1)}{\mathsf{TPR} \cdot \mathbb{P}(H_1) + \mathsf{FPR} \cdot \mathbb{P}(H_0)}$$

Other measures

Previous measures don't account for *base rates*: $\mathbb{P}(H_0)$, $\mathbb{P}(H_1)$

That is, they are suitable when $\mathbb{P}(H_0) \simeq \mathbb{P}(H_1)$

Alternatives

Precision: $$\mathbb{P}\big(H_1 \,\big|\, f(\boldsymbol{X}) = 1\big) = \frac{\mathsf{TPR} \cdot \mathbb{P}(H_1)}{\mathsf{TPR} \cdot \mathbb{P}(H_1) + \mathsf{FPR} \cdot \mathbb{P}(H_0)}$$

$\boldsymbol{F_1}$**-score:** harmonic mean between precision and recall (TPR):

Other measures

Previous measures don't account for *base rates*: $\mathbb{P}(H_0)$, $\mathbb{P}(H_1)$

That is, they are suitable when $\mathbb{P}(H_0) \simeq \mathbb{P}(H_1)$

Alternatives

Precision: $\mathbb{P}\big(H_1 \,\big|\, f(\boldsymbol{X}) = 1\big) = \dfrac{\mathsf{TPR} \cdot \mathbb{P}(H_1)}{\mathsf{TPR} \cdot \mathbb{P}(H_1) + \mathsf{FPR} \cdot \mathbb{P}(H_0)}$

***F_1*-score:** harmonic mean between precision and recall (TPR):

$$F_1 = \frac{2}{\frac{1}{\mathbb{P}(H_1 \mid f(\boldsymbol{X})=1)} + \frac{1}{\mathsf{TPR}}}$$

Table of probabilities

True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true	TNR	FPR $\boldsymbol{\alpha}$
H_1 is true	FNR $\boldsymbol{\beta}$	TPR

Table of probabilities

True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true	TNR	FPR $\boldsymbol{\alpha}$
H_1 is true	FNR $\boldsymbol{\beta}$	TPR

α and β are in *conflict:*

Table of probabilities

True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true	TNR	FPR $\boldsymbol{\alpha}$
H_1 is true	FNR $\boldsymbol{\beta}$	TPR

α and β are in *conflict:*

- $\alpha \downarrow \implies \beta \uparrow$: and vice-versa

Table of probabilities

True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true	TNR	FPR $\boldsymbol{\alpha}$
H_1 is true	FNR $\boldsymbol{\beta}$	TPR

α and β are in *conflict:*

- $\alpha \downarrow \implies \beta \uparrow$: and vice-versa
- Both can decrease only by observing more data

Table of probabilities

True hypothesis	$f(\boldsymbol{X}) = 0$	$f(\boldsymbol{X}) = 1$
H_0 is true	TNR	FPR $\boldsymbol{\alpha}$
H_1 is true	FNR $\boldsymbol{\beta}$	TPR

α and β are in *conflict:*

- $\alpha \downarrow \implies \beta \uparrow$: and vice-versa
- Both can decrease only by observing more data

It turns out that likelihood ratio tests are *Pareto optimal*

Neyman-Pearson Lemma

Neyman-Pearson Lemma

Theorem

Let $f_{\mathsf{LRT}}(\boldsymbol{x})$ be a likelihood ratio decision rule

Neyman-Pearson Lemma

Theorem

Let $f_{\mathsf{LRT}}(\boldsymbol{x})$ be a likelihood ratio decision rule with FPR and FNR

$$\alpha_{\mathsf{LRT}} = \mathbb{P}\Big(f_{\mathsf{LRT}}(\boldsymbol{x}) = 1 \,\Big|\, H_0\Big) \qquad \beta_{\mathsf{LRT}} = \mathbb{P}\Big(f_{\mathsf{LRT}}(\boldsymbol{x}) = 0 \,\Big|\, H_1\Big)$$

Neyman-Pearson Lemma

Theorem

Let $f_{\mathsf{LRT}}(\boldsymbol{x})$ be a likelihood ratio decision rule with FPR and FNR

$$\alpha_{\mathsf{LRT}} = \mathbb{P}\Big(f_{\mathsf{LRT}}(\boldsymbol{x}) = 1 \,\Big|\, H_0\Big) \qquad \beta_{\mathsf{LRT}} = \mathbb{P}\Big(f_{\mathsf{LRT}}(\boldsymbol{x}) = 0 \,\Big|\, H_1\Big)$$

Let $f(\boldsymbol{x})$ be another (deterministic or probabilistic) decision rule with

Neyman-Pearson Lemma

Theorem

Let $f_{\mathsf{LRT}}(\boldsymbol{x})$ be a likelihood ratio decision rule with FPR and FNR

$$\alpha_{\mathsf{LRT}} = \mathbb{P}\Big(f_{\mathsf{LRT}}(\boldsymbol{x}) = 1 \,\Big|\, H_0\Big) \qquad \beta_{\mathsf{LRT}} = \mathbb{P}\Big(f_{\mathsf{LRT}}(\boldsymbol{x}) = 0 \,\Big|\, H_1\Big)$$

Let $f(\boldsymbol{x})$ be another (deterministic or probabilistic) decision rule with

$$\alpha = \mathbb{P}\Big(f(\boldsymbol{x}) = 1 \,\Big|\, H_0\Big) \qquad \beta = \mathbb{P}\Big(f(\boldsymbol{x}) = 0 \,\Big|\, H_1\Big)$$

Neyman-Pearson Lemma

Theorem

Let $f_{\mathsf{LRT}}(\boldsymbol{x})$ be a likelihood ratio decision rule with FPR and FNR

$$\alpha_{\mathsf{LRT}} = \mathbb{P}\Big(f_{\mathsf{LRT}}(\boldsymbol{x}) = 1 \,\Big|\, H_0\Big) \qquad \beta_{\mathsf{LRT}} = \mathbb{P}\Big(f_{\mathsf{LRT}}(\boldsymbol{x}) = 0 \,\Big|\, H_1\Big)$$

Let $f(\boldsymbol{x})$ be another (deterministic or probabilistic) decision rule with

$$\alpha = \mathbb{P}\Big(f(\boldsymbol{x}) = 1 \,\Big|\, H_0\Big) \qquad \beta = \mathbb{P}\Big(f(\boldsymbol{x}) = 0 \,\Big|\, H_1\Big)$$

Then,

Neyman-Pearson Lemma

Theorem

Let $f_{\mathsf{LRT}}(\boldsymbol{x})$ be a likelihood ratio decision rule with FPR and FNR

$$\alpha_{\mathsf{LRT}} = \mathbb{P}\Big(f_{\mathsf{LRT}}(\boldsymbol{x}) = 1 \,\Big|\, H_0\Big) \qquad \beta_{\mathsf{LRT}} = \mathbb{P}\Big(f_{\mathsf{LRT}}(\boldsymbol{x}) = 0 \,\Big|\, H_1\Big)$$

Let $f(\boldsymbol{x})$ be another (deterministic or probabilistic) decision rule with

$$\alpha = \mathbb{P}\Big(f(\boldsymbol{x}) = 1 \,\Big|\, H_0\Big) \qquad \beta = \mathbb{P}\Big(f(\boldsymbol{x}) = 0 \,\Big|\, H_1\Big)$$

Then,

$$\alpha \le \alpha_{\mathsf{LRT}} \qquad \Longrightarrow \qquad \beta \ge \beta_{\mathsf{LRT}}$$

Neyman-Pearson Lemma

Theorem

Let $f_{\mathsf{LRT}}(\boldsymbol{x})$ be a likelihood ratio decision rule with FPR and FNR

$$\alpha_{\mathsf{LRT}} = \mathbb{P}\Big(f_{\mathsf{LRT}}(\boldsymbol{x}) = 1 \,\Big|\, H_0\Big) \qquad \beta_{\mathsf{LRT}} = \mathbb{P}\Big(f_{\mathsf{LRT}}(\boldsymbol{x}) = 0 \,\Big|\, H_1\Big)$$

Let $f(\boldsymbol{x})$ be another (deterministic or probabilistic) decision rule with

$$\alpha = \mathbb{P}\Big(f(\boldsymbol{x}) = 1 \,\Big|\, H_0\Big) \qquad \beta = \mathbb{P}\Big(f(\boldsymbol{x}) = 0 \,\Big|\, H_1\Big)$$

Then,

$$\alpha \leq \alpha_{\mathsf{LRT}} \quad \Longrightarrow \quad \beta \geq \beta_{\mathsf{LRT}}$$

$$\beta \leq \beta_{\mathsf{LRT}} \quad \Longrightarrow \quad \alpha \geq \alpha_{\mathsf{LRT}}$$

Neyman-Pearson Lemma

Theorem

Let $f_{\mathsf{LRT}}(\boldsymbol{x})$ be a likelihood ratio decision rule with FPR and FNR

$$\alpha_{\mathsf{LRT}} = \mathbb{P}\Big(f_{\mathsf{LRT}}(\boldsymbol{x}) = 1 \,\Big|\, H_0\Big) \qquad \beta_{\mathsf{LRT}} = \mathbb{P}\Big(f_{\mathsf{LRT}}(\boldsymbol{x}) = 0 \,\Big|\, H_1\Big)$$

Let $f(\boldsymbol{x})$ be another (deterministic or probabilistic) decision rule with

$$\alpha = \mathbb{P}\Big(f(\boldsymbol{x}) = 1 \,\Big|\, H_0\Big) \qquad \beta = \mathbb{P}\Big(f(\boldsymbol{x}) = 0 \,\Big|\, H_1\Big)$$

Then,

$$\alpha \le \alpha_{\mathsf{LRT}} \qquad \Longrightarrow \qquad \beta \ge \beta_{\mathsf{LRT}}$$

$$\beta \le \beta_{\mathsf{LRT}} \qquad \Longrightarrow \qquad \alpha \ge \alpha_{\mathsf{LRT}}$$

And the same relations hold with strict inequalities $(<, >)$

Proof

Proof

- $f_{\text{MAP}}(\boldsymbol{x})$ minimizes probability of error over all rules (det. and prob.):

Proof

- $f_{\mathsf{MAP}}(\boldsymbol{x})$ minimizes probability of error over all rules (det. and prob.):

$$\mathbb{P}\Big(\mathsf{error}_{\mathsf{MAP}}\Big)$$

$$= \mathbb{P}\Big(f_{\mathsf{MAP}}(\boldsymbol{X}) = 1,\ H_0\Big) + \mathbb{P}\Big(f_{\mathsf{MAP}}(\boldsymbol{X}) = 1,\ H_1\Big)$$

Proof

- $f_{\mathsf{MAP}}(\boldsymbol{x})$ minimizes probability of error over all rules (det. and prob.):

$$
\begin{aligned}
&\mathbb{P}\Big(\mathsf{error}_{\mathsf{MAP}}\Big) \\
&= \mathbb{P}\Big(f_{\mathsf{MAP}}(\boldsymbol{X}) = 1,\ H_0\Big) + \mathbb{P}\Big(f_{\mathsf{MAP}}(\boldsymbol{X}) = 1,\ H_1\Big) \\
&= \mathbb{P}\Big(f_{\mathsf{MAP}}(\boldsymbol{X}) = 1 \,\Big|\, H_0\Big)\ \mathbb{P}(H_0) + \mathbb{P}\Big(f_{\mathsf{MAP}}(\boldsymbol{X}) = 1 \,\Big|\, H_1\Big)\ \mathbb{P}(H_1)
\end{aligned}
$$

Proof

- $f_{\mathsf{MAP}}(\boldsymbol{x})$ minimizes probability of error over all rules (det. and prob.):

$$
\begin{aligned}
&\mathbb{P}\Big(\mathsf{error}_{\mathsf{MAP}}\Big) \\
&= \mathbb{P}\Big(f_{\mathsf{MAP}}(\boldsymbol{X}) = 1,\ H_0\Big) + \mathbb{P}\Big(f_{\mathsf{MAP}}(\boldsymbol{X}) = 1,\ H_1\Big) \\
&= \underbrace{\mathbb{P}\Big(f_{\mathsf{MAP}}(\boldsymbol{X}) = 1 \,\Big|\, H_0\Big)}_{\alpha_{\mathsf{MAP}}}\ \mathbb{P}(H_0) + \mathbb{P}\Big(f_{\mathsf{MAP}}(\boldsymbol{X}) = 1 \,\Big|\, H_1\Big)\ \mathbb{P}(H_1)
\end{aligned}
$$

Proof

- $f_{\mathsf{MAP}}(\boldsymbol{x})$ minimizes probability of error over all rules (det. and prob.):

$$
\begin{aligned}
&\mathbb{P}\Big(\mathsf{error}_{\mathsf{MAP}}\Big) \\
&= \mathbb{P}\Big(f_{\mathsf{MAP}}(\boldsymbol{X}) = 1,\ H_0\Big) + \mathbb{P}\Big(f_{\mathsf{MAP}}(\boldsymbol{X}) = 1,\ H_1\Big) \\
&= \underbrace{\mathbb{P}\Big(f_{\mathsf{MAP}}(\boldsymbol{X}) = 1 \,\Big|\, H_0\Big)}_{\alpha_{\mathsf{MAP}}} \mathbb{P}(H_0) + \underbrace{\mathbb{P}\Big(f_{\mathsf{MAP}}(\boldsymbol{X}) = 1 \,\Big|\, H_1\Big)}_{\beta_{\mathsf{MAP}}} \mathbb{P}(H_1)
\end{aligned}
$$

Proof

- $f_{\mathsf{MAP}}(\boldsymbol{x})$ minimizes probability of error over all rules (det. and prob.):

$$
\begin{aligned}
&\mathbb{P}\Big(\mathsf{error}_{\mathsf{MAP}}\Big)\\
&= \mathbb{P}\Big(f_{\mathsf{MAP}}(\boldsymbol{X}) = 1,\ H_0\Big) + \mathbb{P}\Big(f_{\mathsf{MAP}}(\boldsymbol{X}) = 1,\ H_1\Big)\\
&= \underbrace{\mathbb{P}\Big(f_{\mathsf{MAP}}(\boldsymbol{X}) = 1 \,\Big|\, H_0\Big)}_{\alpha_{\mathsf{MAP}}} \mathbb{P}(H_0) + \underbrace{\mathbb{P}\Big(f_{\mathsf{MAP}}(\boldsymbol{X}) = 1 \,\Big|\, H_1\Big)}_{\beta_{\mathsf{MAP}}} \mathbb{P}(H_1)\\
&= \alpha_{\mathsf{MAP}}\, \mathbb{P}(H_0) + \beta_{\mathsf{MAP}}\, \mathbb{P}(H_1)
\end{aligned}
$$

Proof

- $f_{\mathsf{MAP}}(\boldsymbol{x})$ minimizes probability of error over all rules (det. and prob.):

$$\begin{aligned}
&\mathbb{P}\Big(\mathsf{error}_{\mathsf{MAP}}\Big)\\
&= \mathbb{P}\Big(f_{\mathsf{MAP}}(\boldsymbol{X}) = 1,\ H_0\Big) + \mathbb{P}\Big(f_{\mathsf{MAP}}(\boldsymbol{X}) = 1,\ H_1\Big)\\
&= \underbrace{\mathbb{P}\Big(f_{\mathsf{MAP}}(\boldsymbol{X}) = 1 \,\Big|\, H_0\Big)}_{\alpha_{\mathsf{MAP}}} \mathbb{P}(H_0) + \underbrace{\mathbb{P}\Big(f_{\mathsf{MAP}}(\boldsymbol{X}) = 1 \,\Big|\, H_1\Big)}_{\beta_{\mathsf{MAP}}} \mathbb{P}(H_1)\\
&= \alpha_{\mathsf{MAP}}\, \mathbb{P}(H_0) + \beta_{\mathsf{MAP}}\, \mathbb{P}(H_1)
\end{aligned}$$

- $f_{\mathsf{MAP}}(\boldsymbol{x})$ is characterized by

Proof

- $f_{\mathsf{MAP}}(\boldsymbol{x})$ minimizes probability of error over all rules (det. and prob.):

$$
\begin{aligned}
&\mathbb{P}\Big(\mathsf{error}_{\mathsf{MAP}}\Big) \\
&= \mathbb{P}\Big(f_{\mathsf{MAP}}(\boldsymbol{X}) = 1,\ H_0\Big) + \mathbb{P}\Big(f_{\mathsf{MAP}}(\boldsymbol{X}) = 1,\ H_1\Big) \\
&= \underbrace{\mathbb{P}\Big(f_{\mathsf{MAP}}(\boldsymbol{X}) = 1 \,\Big|\, H_0\Big)}_{\alpha_{\mathsf{MAP}}} \mathbb{P}(H_0) + \underbrace{\mathbb{P}\Big(f_{\mathsf{MAP}}(\boldsymbol{X}) = 1 \,\Big|\, H_1\Big)}_{\beta_{\mathsf{MAP}}} \mathbb{P}(H_1) \\
&= \alpha_{\mathsf{MAP}}\, \mathbb{P}(H_0) + \beta_{\mathsf{MAP}}\, \mathbb{P}(H_1)
\end{aligned}
$$

- $f_{\mathsf{MAP}}(\boldsymbol{x})$ is characterized by $\quad f_{\mathsf{MAP}}(\boldsymbol{x}) = 1 \quad \text{if} \quad \mathcal{L}(\boldsymbol{x}) \geq \eta = \dfrac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)}$

Proof

- $f_{\mathsf{MAP}}(\boldsymbol{x})$ minimizes probability of error over all rules (det. and prob.):

$$
\begin{aligned}
&\mathbb{P}\Big(\text{error}_{\mathsf{MAP}}\Big)\\
&=\mathbb{P}\Big(f_{\mathsf{MAP}}(\boldsymbol{X})=1,\ H_0\Big)+\mathbb{P}\Big(f_{\mathsf{MAP}}(\boldsymbol{X})=1,\ H_1\Big)\\
&=\underbrace{\mathbb{P}\Big(f_{\mathsf{MAP}}(\boldsymbol{X})=1 \,\Big|\, H_0\Big)}_{\alpha_{\mathsf{MAP}}}\mathbb{P}(H_0)+\underbrace{\mathbb{P}\Big(f_{\mathsf{MAP}}(\boldsymbol{X})=1 \,\Big|\, H_1\Big)}_{\beta_{\mathsf{MAP}}}\mathbb{P}(H_1)\\
&=\alpha_{\mathsf{MAP}}\,\mathbb{P}(H_0)+\beta_{\mathsf{MAP}}\,\mathbb{P}(H_1)
\end{aligned}
$$

- $f_{\mathsf{MAP}}(\boldsymbol{x})$ is characterized by $\quad f_{\mathsf{MAP}}(\boldsymbol{x})=1 \quad$ if $\quad \mathcal{L}(\boldsymbol{x})\geq\eta=\dfrac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)}$

- Select $\;0<\eta<1\;$ such that $\mathbb{P}(H_0)=\dfrac{\eta}{\eta+1}$ and $\mathbb{P}(H_1)=\dfrac{1}{\eta+1}$

Proof

- $f_{\text{MAP}}(\boldsymbol{x})$ minimizes probability of error over all rules (det. and prob.):

$$\begin{aligned}
&\mathbb{P}\Big(\text{error}_{\text{MAP}}\Big) \\
&= \mathbb{P}\Big(f_{\text{MAP}}(\boldsymbol{X}) = 1,\ H_0\Big) + \mathbb{P}\Big(f_{\text{MAP}}(\boldsymbol{X}) = 1,\ H_1\Big) \\
&= \underbrace{\mathbb{P}\Big(f_{\text{MAP}}(\boldsymbol{X}) = 1 \,\Big|\, H_0\Big)}_{\alpha_{\text{MAP}}} \mathbb{P}(H_0) + \underbrace{\mathbb{P}\Big(f_{\text{MAP}}(\boldsymbol{X}) = 1 \,\Big|\, H_1\Big)}_{\beta_{\text{MAP}}} \mathbb{P}(H_1) \\
&= \alpha_{\text{MAP}}\, \mathbb{P}(H_0) + \beta_{\text{MAP}}\, \mathbb{P}(H_1)
\end{aligned}$$

- $f_{\text{MAP}}(\boldsymbol{x})$ is characterized by $\quad f_{\text{MAP}}(\boldsymbol{x}) = 1 \quad \text{if} \quad \mathcal{L}(\boldsymbol{x}) \geq \eta = \dfrac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)}$

- Select $\;0 < \eta < 1\;$ such that $\mathbb{P}(H_0) = \dfrac{\eta}{\eta + 1}$ and $\mathbb{P}(H_1) = \dfrac{1}{\eta + 1}$

  This implies $f_{\text{LRT}}(\boldsymbol{x}) = f_{\text{MAP}}(\boldsymbol{x})$

Proof

- Thus, for any decision rule $f(\boldsymbol{x})$ with errors α and β,

Proof

- Thus, for any decision rule $f(\boldsymbol{x})$ with errors α and β,

$$\mathbb{P}\big(\mathsf{error}_f\big) \geq \mathbb{P}\big(\mathsf{error}_{\mathsf{LRT}}\big)$$

Proof

- Thus, for any decision rule $f(\boldsymbol{x})$ with errors α and β,

$$\mathbb{P}\big(\mathsf{error}_f\big) \geq \mathbb{P}\big(\mathsf{error}_{\mathsf{LRT}}\big)$$

$$\iff \qquad \alpha \frac{\eta}{\eta+1} + \beta \frac{1}{\eta+1} \; \geq \; \alpha_{\mathsf{LRT}} \, \frac{\eta}{\eta+1} + \beta_{\mathsf{LRT}} \, \frac{1}{\eta+1}$$

Proof

- Thus, for any decision rule $f(\boldsymbol{x})$ with errors α and β,

$$\mathbb{P}\big(\mathsf{error}_f\big) \geq \mathbb{P}\big(\mathsf{error}_{\mathsf{LRT}}\big)$$

$$\iff \qquad \alpha\frac{\eta}{\eta+1} + \beta\frac{1}{\eta+1} \;\geq\; \alpha_{\mathsf{LRT}}\,\frac{\eta}{\eta+1} + \beta_{\mathsf{LRT}}\,\frac{1}{\eta+1}$$

$$\iff \qquad (\alpha - \alpha_{\mathsf{LRT}})\,\frac{\eta}{\eta+1} \; + (\beta - \beta_{\mathsf{LRT}})\,\frac{1}{\eta+1} \;\; \geq \; 0$$

Proof

- Thus, for any decision rule $f(\boldsymbol{x})$ with errors α and β,

$$\mathbb{P}\big(\mathsf{error}_f\big) \geq \mathbb{P}\big(\mathsf{error}_{\mathsf{LRT}}\big)$$

$$\iff \qquad \alpha\frac{\eta}{\eta+1} + \beta\frac{1}{\eta+1} \;\geq\; \alpha_{\mathsf{LRT}}\,\frac{\eta}{\eta+1} + \beta_{\mathsf{LRT}}\,\frac{1}{\eta+1}$$

$$\iff \qquad (\alpha - \alpha_{\mathsf{LRT}})\,\underbrace{\frac{\eta}{\eta+1}}_{>0} + (\beta - \beta_{\mathsf{LRT}})\,\frac{1}{\eta+1} \;\geq\; 0$$

Proof

- Thus, for any decision rule $f(\boldsymbol{x})$ with errors α and β,

$$\mathbb{P}\big(\mathsf{error}_f\big) \geq \mathbb{P}\big(\mathsf{error}_{\mathsf{LRT}}\big)$$

$$\iff \qquad \alpha \frac{\eta}{\eta+1} + \beta \frac{1}{\eta+1} \ \geq \ \alpha_{\mathsf{LRT}} \, \frac{\eta}{\eta+1} + \beta_{\mathsf{LRT}} \, \frac{1}{\eta+1}$$

$$\iff \qquad (\alpha - \alpha_{\mathsf{LRT}}) \underbrace{\frac{\eta}{\eta+1}}_{>0} + (\beta - \beta_{\mathsf{LRT}}) \underbrace{\frac{1}{\eta+1}}_{>0} \ \geq \ 0$$

Proof

- Thus, for any decision rule $f(\boldsymbol{x})$ with errors α and β,

$$\mathbb{P}\big(\mathsf{error}_f\big) \geq \mathbb{P}\big(\mathsf{error}_{\mathsf{LRT}}\big)$$

$$\iff \qquad \alpha\frac{\eta}{\eta+1} + \beta\frac{1}{\eta+1} \;\geq\; \alpha_{\mathsf{LRT}}\,\frac{\eta}{\eta+1} + \beta_{\mathsf{LRT}}\,\frac{1}{\eta+1}$$

$$\iff \qquad (\alpha - \alpha_{\mathsf{LRT}})\underbrace{\frac{\eta}{\eta+1}}_{>0} + (\beta - \beta_{\mathsf{LRT}})\underbrace{\frac{1}{\eta+1}}_{>0} \;\geq\; 0$$

- Thus,

Proof

- Thus, for any decision rule $f(\boldsymbol{x})$ with errors α and β,

$$\mathbb{P}\big(\mathsf{error}_f\big) \geq \mathbb{P}\big(\mathsf{error}_{\mathsf{LRT}}\big)$$

$$\iff \qquad \alpha\frac{\eta}{\eta+1} + \beta\frac{1}{\eta+1} \;\geq\; \alpha_{\mathsf{LRT}}\,\frac{\eta}{\eta+1} + \beta_{\mathsf{LRT}}\,\frac{1}{\eta+1}$$

$$\iff \qquad (\alpha - \alpha_{\mathsf{LRT}})\underbrace{\frac{\eta}{\eta+1}}_{>0} + (\beta - \beta_{\mathsf{LRT}})\underbrace{\frac{1}{\eta+1}}_{>0} \;\geq\; 0$$

- Thus,

$$\alpha \;\leq\; \alpha_{\mathsf{LRT}} \qquad \Longrightarrow \qquad \beta \;\geq\; \beta_{\mathsf{LRT}}$$

Proof

- Thus, for any decision rule $f(\boldsymbol{x})$ with errors α and β,

$$\mathbb{P}\big(\mathsf{error}_f\big) \geq \mathbb{P}\big(\mathsf{error}_{\mathsf{LRT}}\big)$$

$$\iff \qquad \alpha \frac{\eta}{\eta+1} + \beta \frac{1}{\eta+1} \ \geq \ \alpha_{\mathsf{LRT}} \frac{\eta}{\eta+1} + \beta_{\mathsf{LRT}} \frac{1}{\eta+1}$$

$$\iff \qquad (\alpha - \alpha_{\mathsf{LRT}}) \underbrace{\frac{\eta}{\eta+1}}_{>0} + (\beta - \beta_{\mathsf{LRT}}) \underbrace{\frac{1}{\eta+1}}_{>0} \ \geq \ 0$$

- Thus,

$$\alpha \ \leq \ \alpha_{\mathsf{LRT}} \qquad \Longrightarrow \qquad \beta \ \geq \ \beta_{\mathsf{LRT}}$$

$$\beta \ \leq \ \beta_{\mathsf{LRT}} \qquad \Longrightarrow \qquad \alpha \ \geq \ \alpha_{\mathsf{LRT}}$$

Proof

- Thus, for any decision rule $f(\boldsymbol{x})$ with errors α and β,

$$\mathbb{P}\big(\mathsf{error}_f\big) \geq \mathbb{P}\big(\mathsf{error}_{\mathsf{LRT}}\big)$$

$$\iff \qquad \alpha\frac{\eta}{\eta+1} + \beta\frac{1}{\eta+1} \geq \alpha_{\mathsf{LRT}}\,\frac{\eta}{\eta+1} + \beta_{\mathsf{LRT}}\,\frac{1}{\eta+1}$$

$$\iff \qquad (\alpha - \alpha_{\mathsf{LRT}})\underbrace{\frac{\eta}{\eta+1}}_{>0} + (\beta - \beta_{\mathsf{LRT}})\underbrace{\frac{1}{\eta+1}}_{>0} \geq 0$$

- Thus,

$$\alpha \leq \alpha_{\mathsf{LRT}} \qquad \Longrightarrow \qquad \beta \geq \beta_{\mathsf{LRT}}$$

$$\beta \leq \beta_{\mathsf{LRT}} \qquad \Longrightarrow \qquad \alpha \geq \alpha_{\mathsf{LRT}}$$

$\square$

Receiver Operating Characteristic (ROC)

Receiver Operating Characteristic (ROC)

Consider a likelihood ratio test with threshold η:

Receiver Operating Characteristic (ROC)

Consider a likelihood ratio test with threshold η: $f_{\mathsf{LRT}}(\boldsymbol{x}\,;\,\eta)$

Receiver Operating Characteristic (ROC)

Consider a likelihood ratio test with threshold η: $f_{\mathsf{LRT}}(\boldsymbol{x}\,;\,\eta)$

For each η, there is a pair

Receiver Operating Characteristic (ROC)

Consider a likelihood ratio test with threshold η: $f_{\mathsf{LRT}}(\boldsymbol{x}\,;\,\eta)$

For each η, there is a pair $\Big(\alpha(\eta),\, 1 - \beta(\eta)\Big)$

Receiver Operating Characteristic (ROC)

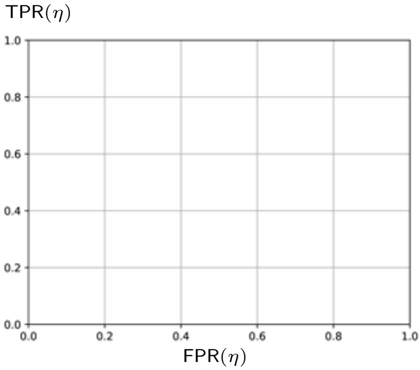
Consider a likelihood ratio test with threshold η: $f_{\mathsf{LRT}}(\boldsymbol{x}\,;\,\eta)$

For each η, there is a pair $\Big(\alpha(\eta),\, 1-\beta(\eta)\Big) = \Big(\mathsf{FPR}(\eta),\, \mathsf{TPR}(\eta)\Big)$

Receiver Operating Characteristic (ROC)

Consider a likelihood ratio test with threshold η: $f_{\mathsf{LRT}}(\boldsymbol{x}\,;\,\eta)$

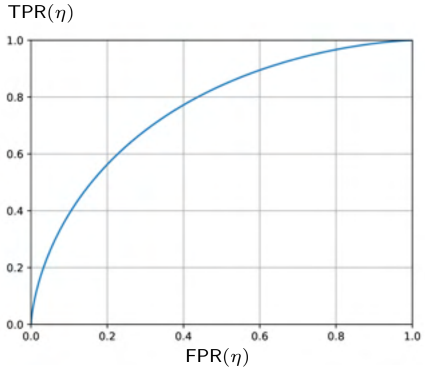
For each η, there is a pair $\Big(\alpha(\eta),\,1-\beta(\eta)\Big) = \Big(\mathsf{FPR}(\eta),\,\mathsf{TPR}(\eta)\Big)$



Receiver Operating Characteristic (ROC)

Consider a likelihood ratio test with threshold η: $f_{\text{LRT}}(\boldsymbol{x}\,;\,\eta)$

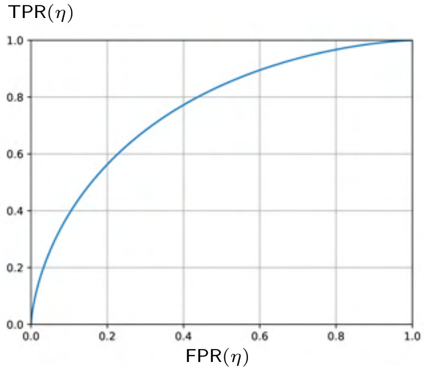
For each η, there is a pair $\Big(\alpha(\eta),\, 1-\beta(\eta)\Big) = \Big(\text{FPR}(\eta),\, \text{TPR}(\eta)\Big)$



Receiver Operating Characteristic (ROC)

Consider a likelihood ratio test with threshold η: $f_{\mathsf{LRT}}(\boldsymbol{x}\,;\,\eta)$

For each η, there is a pair $\Big(\alpha(\eta),\, 1-\beta(\eta)\Big) = \Big(\mathsf{FPR}(\eta),\, \mathsf{TPR}(\eta)\Big)$

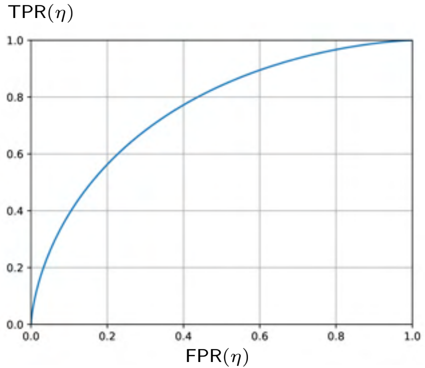


ROC curve

Receiver Operating Characteristic (ROC)

Consider a likelihood ratio test with threshold η: $f_{\text{LRT}}(\boldsymbol{x}\,;\,\eta)$

For each η, there is a pair $\Big(\alpha(\eta),\, 1-\beta(\eta)\Big) = \Big(\text{FPR}(\eta),\, \text{TPR}(\eta)\Big)$



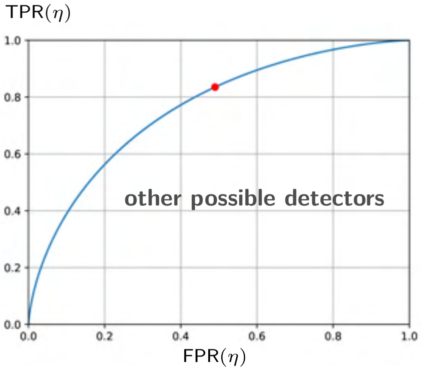
ROC curve

Neyman-Pearson implies *Pareto optimality*

Receiver Operating Characteristic (ROC)

Consider a likelihood ratio test with threshold η: $f_{\text{LRT}}(\boldsymbol{x}\,;\,\eta)$

For each η, there is a pair $\Big(\alpha(\eta),\, 1-\beta(\eta)\Big) = \Big(\text{FPR}(\eta),\, \text{TPR}(\eta)\Big)$



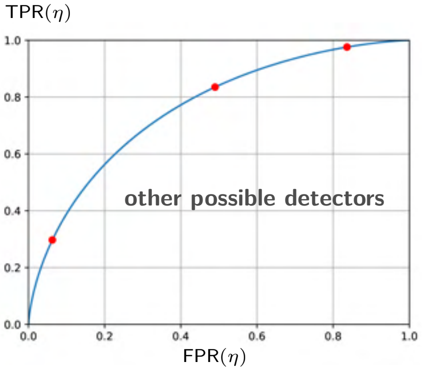
ROC curve

Neyman-Pearson implies *Pareto optimality*

Receiver Operating Characteristic (ROC)

Consider a likelihood ratio test with threshold η: $f_{\text{LRT}}(\boldsymbol{x}\,;\,\eta)$

For each η, there is a pair $\Big(\alpha(\eta),\, 1-\beta(\eta)\Big) = \Big(\text{FPR}(\eta),\, \text{TPR}(\eta)\Big)$



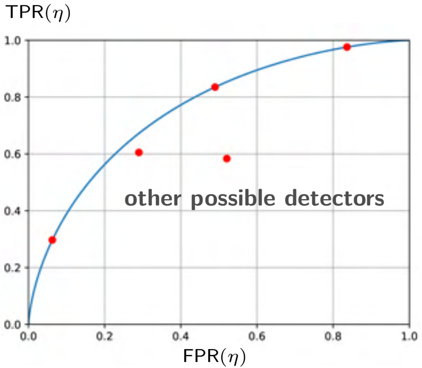
ROC curve

Neyman-Pearson implies *Pareto optimality*

Receiver Operating Characteristic (ROC)

Consider a likelihood ratio test with threshold η: $f_{\mathsf{LRT}}(\boldsymbol{x}\,;\,\eta)$

For each η, there is a pair $\Big(\alpha(\eta),\, 1-\beta(\eta)\Big) = \Big(\mathsf{FPR}(\eta),\, \mathsf{TPR}(\eta)\Big)$

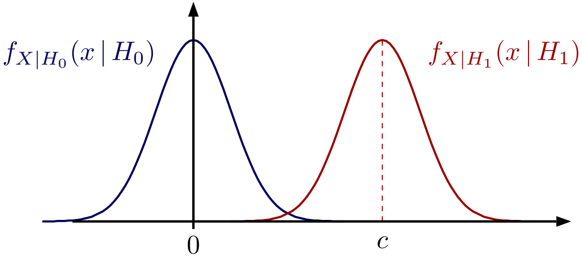


ROC curve

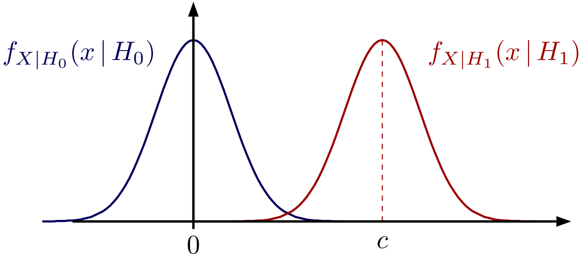
Neyman-Pearson implies *Pareto optimality*

Example

Example

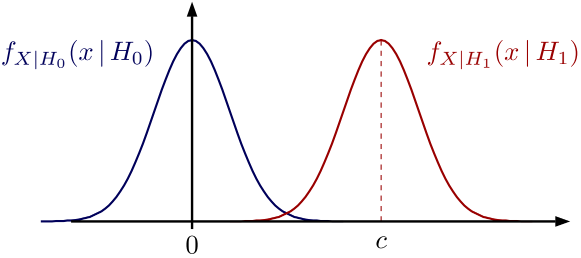


Example



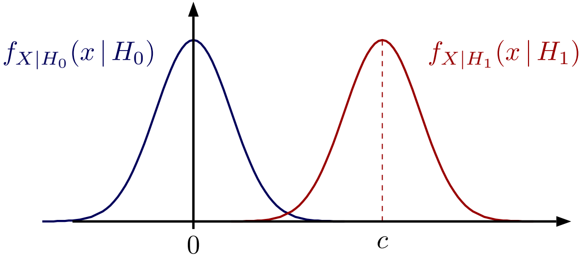
$$f_{X|H_0}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$$

Example



$$f_{X|H_0}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \qquad f_{X|H_1}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-c)^2}{2\sigma^2}}$$

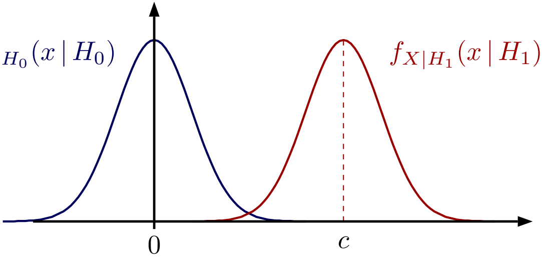
Example



$$f_{X|H_0}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \qquad f_{X|H_1}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-c)^2}{2\sigma^2}}$$

Likelihood ratio test:

Example

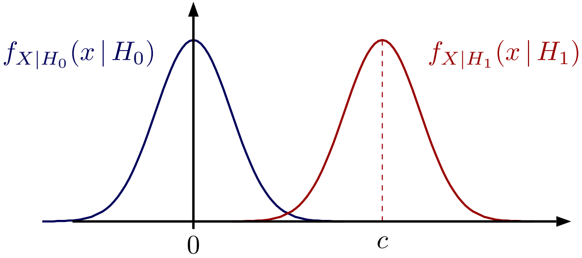


$$f_{X|H_0}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \qquad f_{X|H_1}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-c)^2}{2\sigma^2}}$$

Likelihood ratio test:

$$\mathcal{L}(x) = \frac{f_{X|H_1}(x)}{f_{X|H_0}(x)} \geq \eta$$

Example

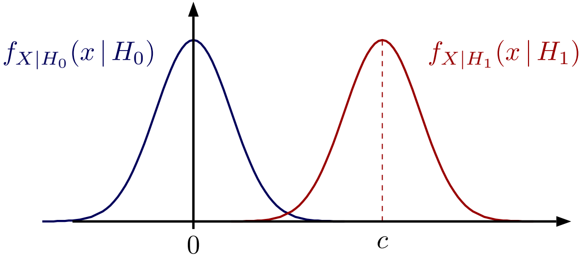


$$f_{X|H_0}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \qquad f_{X|H_1}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-c)^2}{2\sigma^2}}$$

Likelihood ratio test:

$$\mathcal{L}(x) = \frac{f_{X|H_1}(x)}{f_{X|H_0}(x)} \geq \eta \quad \Longleftrightarrow \quad x \geq \frac{c}{2} + \frac{\sigma^2}{c} \log \eta$$

Example



$$f_{X|H_0}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \qquad f_{X|H_1}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-c)^2}{2\sigma^2}}$$

Likelihood ratio test:

$$\mathcal{L}(x) = \frac{f_{X|H_1}(x)}{f_{X|H_0}(x)} \geq \eta \quad \Longleftrightarrow \quad x \geq \frac{c}{2} + \frac{\sigma^2}{c} \log \eta =: \gamma$$

Example

$x \geq \frac{c}{2} + \frac{\sigma^2}{c} \log \eta := \gamma$

Example

$x \geq \frac{c}{2} + \frac{\sigma^2}{c} \log \eta := \gamma$

$\mathsf{TPR}(\eta)$

Example

$x \geq \frac{c}{2} + \frac{\sigma^2}{c} \log \eta \; := \; \gamma$

$$\mathsf{TPR}(\eta) = \mathbb{P}\Big(X \geq \gamma \,\Big|\, H_1\Big)$$

Example

$x \geq \frac{c}{2} + \frac{\sigma^2}{c} \log \eta \; := \; \gamma$

$$\mathsf{TPR}(\eta) = \mathbb{P}\Big(X \geq \gamma \,\Big|\, H_1\Big) = \int_{\gamma}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-c)^2}{2\sigma^2}} \,\mathrm{d}\,x$$

Example

$$x \geq \frac{c}{2} + \frac{\sigma^2}{c} \log \eta \, := \, \gamma$$

$$\begin{aligned}\mathsf{TPR}(\eta) &= \mathbb{P}\Big(X \geq \gamma \,\Big|\, H_1\Big) = \int_{\gamma}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-c)^2}{2\sigma^2}} \,\mathrm{d}\,x \\ &= \int_{\frac{\gamma-c}{\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \,\mathrm{d}z\end{aligned}$$

Example

$x \geq \frac{c}{2} + \frac{\sigma^2}{c} \log \eta \ := \ \gamma$

$$
\begin{aligned}
\mathsf{TPR}(\eta) &= \mathbb{P}\Big(X \geq \gamma \,\Big|\, H_1\Big) = \int_{\gamma}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-c)^2}{2\sigma^2}} \,\mathrm{d}\,x \\
&= \int_{\frac{\gamma-c}{\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \,\mathrm{d}z \\
&= Q\left(\frac{\gamma - c}{\sigma}\right)
\end{aligned}
$$

Example

$x \geq \frac{c}{2} + \frac{\sigma^2}{c} \log \eta \ := \ \gamma$

$$\mathsf{TPR}(\eta) = \mathbb{P}\Big(X \geq \gamma \,\Big|\, H_1\Big) = \int_\gamma^\infty \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-c)^2}{2\sigma^2}} \, \mathrm{d}\, x$$

$$= \int_{\frac{\gamma-c}{\sigma}}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \, \mathrm{d}z$$

$$= Q\left(\frac{\gamma - c}{\sigma}\right)$$

$\left[Q(x) := \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{z^2}{2}} \, \mathrm{d}z\right]$

Example

$$x \geq \frac{c}{2} + \frac{\sigma^2}{c} \log \eta := \gamma$$

$$\begin{aligned}
\mathsf{TPR}(\eta) &= \mathbb{P}\Big(X \geq \gamma \,\Big|\, H_1\Big) = \int_\gamma^\infty \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-c)^2}{2\sigma^2}} \,\mathrm{d}\,x \\
&= \int_{\frac{\gamma-c}{\sigma}}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \,\mathrm{d}z \\
&= Q\left(\frac{\gamma - c}{\sigma}\right) \\
&= Q\left(\frac{\log \eta}{c/\sigma} - \frac{c/\sigma}{2}\right)
\end{aligned}$$

$$\left[Q(x) := \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{z^2}{2}} \,\mathrm{d}z\right]$$

Example

$x \geq \frac{c}{2} + \frac{\sigma^2}{c} \log \eta := \gamma$

$$\begin{aligned}
\mathsf{TPR}(\eta) &= \mathbb{P}\Big(X \geq \gamma \,\Big|\, H_1\Big) = \int_{\gamma}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-c)^2}{2\sigma^2}} \, \mathrm{d}\, x \\
&= \int_{\frac{\gamma - c}{\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \, \mathrm{d}z \\
&= Q\left(\frac{\gamma - c}{\sigma}\right) \\
&= Q\left(\frac{\log \eta}{c/\sigma} - \frac{c/\sigma}{2}\right) \\
&= Q\left(\frac{\log \eta}{\mathsf{SNR}} - \frac{\mathsf{SNR}}{2}\right)
\end{aligned}$$

$\left[Q(x) := \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-\frac{z^2}{2}} \, \mathrm{d}z\right]$

Example

$x \geq \frac{c}{2} + \frac{\sigma^2}{c} \log \eta := \gamma$

$$
\begin{aligned}
\mathsf{TPR}(\eta) &= \mathbb{P}\Big(X \geq \gamma \,\Big|\, H_1\Big) = \int_{\gamma}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-c)^2}{2\sigma^2}} \, \mathrm{d}\, x \\
&= \int_{\frac{\gamma-c}{\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \, \mathrm{d}z \\
&= Q\left(\frac{\gamma - c}{\sigma}\right) && \left[Q(x) := \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-\frac{z^2}{2}} \, \mathrm{d}z\right] \\
&= Q\left(\frac{\log \eta}{c/\sigma} - \frac{c/\sigma}{2}\right) \\
&= Q\left(\frac{\log \eta}{\mathsf{SNR}} - \frac{\mathsf{SNR}}{2}\right) && \left[\mathsf{SNR} := \frac{c}{\sigma}\right]
\end{aligned}
$$

Example

$x \geq \frac{c}{2} + \frac{\sigma^2}{c} \log \eta \;:=\; \gamma$

$$\begin{aligned}
\mathsf{TPR}(\eta) &= \mathbb{P}\Big(X \geq \gamma \,\Big|\, H_1\Big) = \int_{\gamma}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-c)^2}{2\sigma^2}} \,\mathrm{d}\,x \\
&= \int_{\frac{\gamma - c}{\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \,\mathrm{d}z \\
&= Q\left(\frac{\gamma - c}{\sigma}\right) && \left[Q(x) := \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-\frac{z^2}{2}} \,\mathrm{d}z\right] \\
&= Q\left(\frac{\log \eta}{c/\sigma} - \frac{c/\sigma}{2}\right) \\
&= Q\left(\frac{\log \eta}{\mathsf{SNR}} - \frac{\mathsf{SNR}}{2}\right) && \left[\mathsf{SNR} := \frac{c}{\sigma}\right]
\end{aligned}$$

Similarly,

$$\mathsf{FPR}(\eta) = \mathbb{P}\Big(X \geq \gamma \,\Big|\, H_0\Big) \;=\; \cdots \;=\; Q\left(\frac{\log \eta}{\mathsf{SNR}} + \frac{\mathsf{SNR}}{2}\right)$$

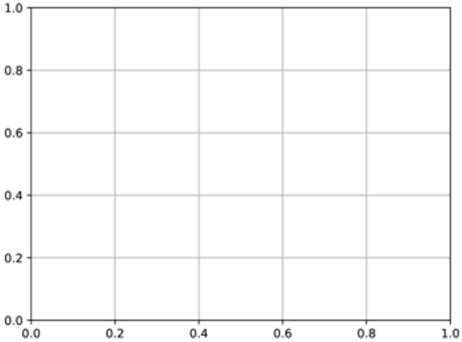
Example

ROC curve for different values of SNR

Example

ROC curve for different values of SNR

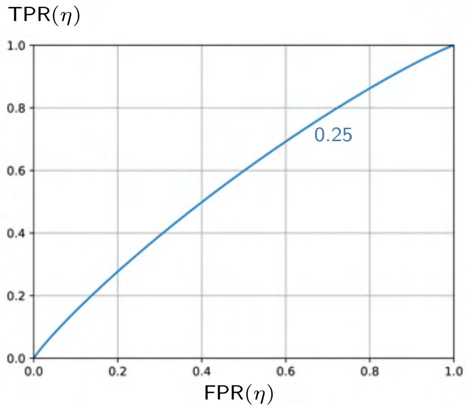
$\mathrm{TPR}(\eta)$



$\mathrm{FPR}(\eta)$

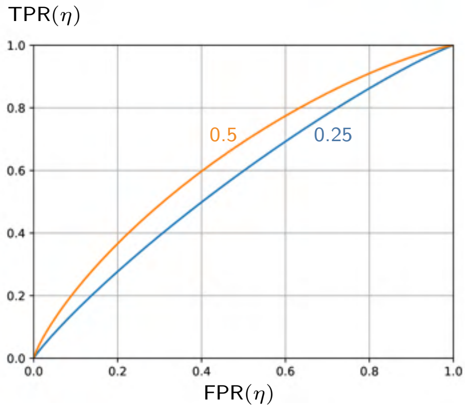
Example

ROC curve for different values of SNR



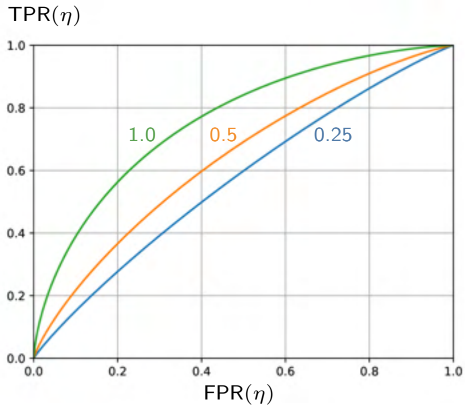
Example

ROC curve for different values of SNR



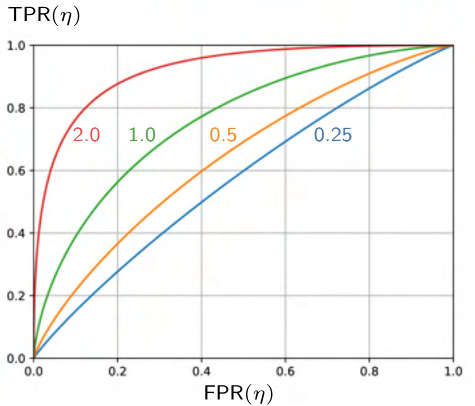
Example

ROC curve for different values of SNR



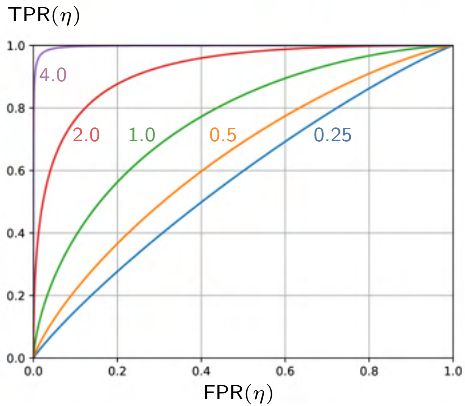
Example

ROC curve for different values of SNR



Example

ROC curve for different values of SNR



Properties of ROC curves

Properties of ROC curves

Property 1: $(0, 0)$ and $(1, 1)$ are in the ROC curve

Properties of ROC curves

Property 1: $(0, 0)$ and $(1, 1)$ are in the ROC curve

Property 2: TPR $\geq$ FPR

Properties of ROC curves

Property 1: $(0, 0)$ and $(1, 1)$ are in the ROC curve

Property 2: TPR $\geq$ FPR

Property 3: The ROC curve is concave

Proofs

Proofs

Property 1: $(0, 0)$ and $(1, 1)$ are in the ROC curve

Proof:

Proofs

Property 1: $(0,0)$ and $(1,1)$ are in the ROC curve

Proof:

- When $\eta \to +\infty$,

Proofs

Property 1: $(0, 0)$ and $(1, 1)$ are in the ROC curve

Proof:

- When $\eta \to +\infty$,

$$f(x) = \mathbb{1}_{\mathcal{L}(\boldsymbol{x}) \geq \eta}(\boldsymbol{x})$$

Proofs

Property 1: $(0, 0)$ and $(1, 1)$ are in the ROC curve

Proof:

- When $\eta \to +\infty$,

$$f(x) = \mathbb{1}_{\mathcal{L}(\boldsymbol{x}) \geq \eta}(\boldsymbol{x}) = 0$$

Proofs

Property 1: $(0,0)$ and $(1,1)$ are in the ROC curve

Proof:

- When $\eta \to +\infty$,

$$f(x) = \mathbb{1}_{\mathcal{L}(\boldsymbol{x}) \geq \eta}(\boldsymbol{x}) = 0$$

  for any $\boldsymbol{x}$

Proofs

Property 1: $(0, 0)$ and $(1, 1)$ are in the ROC curve

Proof:

- When $\eta \to +\infty$,

$$f(x) = \mathbb{1}_{\mathcal{L}(\boldsymbol{x}) \geq \eta}(\boldsymbol{x}) = 0$$

for any $\boldsymbol{x}$

Then, $\big(\text{FPR}(+\infty)\,,\,\text{TPR}(+\infty)\big) \,=\, (0, 0)$

Proofs

Property 1: $(0,0)$ and $(1,1)$ are in the ROC curve

Proof:

- When $\eta \to +\infty$,

$$f(x) = \mathbb{1}_{\mathcal{L}(\boldsymbol{x}) \geq \eta}(\boldsymbol{x}) = 0$$

  for any $\boldsymbol{x}$

  Then, $\big(\mathsf{FPR}(+\infty)\,,\, \mathsf{TPR}(+\infty)\big) \,=\, (0,0)$

- Similarly, $\big(\mathsf{FPR}(-\infty)\,,\, \mathsf{TPR}(-\infty)\big) \,=\, (1,1)$

Proofs

Property 1: $(0, 0)$ and $(1, 1)$ are in the ROC curve

Proof:

- When $\eta \to +\infty$,

$$f(x) = \mathbb{1}_{\mathcal{L}(\boldsymbol{x}) \geq \eta}(\boldsymbol{x}) = 0$$

  for any $\boldsymbol{x}$

  Then, $\big(\text{FPR}(+\infty)\,,\,\text{TPR}(+\infty)\big) \;=\; (0, 0)$

- Similarly, $\big(\text{FPR}(-\infty)\,,\,\text{TPR}(-\infty)\big) \;=\; (1, 1)$

$\square$

Proofs

Proofs

Consider two decision rules (omitting dependence on η):

$$\left(\mathsf{FPR}^{(1)}, \mathsf{TPR}^{(1)}\right) \;\; , \;\; \left(\mathsf{FPR}^{(2)}, \mathsf{TPR}^{(2)}\right)$$

Proofs

Consider two decision rules (omitting dependence on η):

$$\left(\mathsf{FPR}^{(1)}, \mathsf{TPR}^{(1)}\right) \ , \ \left(\mathsf{FPR}^{(2)}, \mathsf{TPR}^{(2)}\right)$$

Build a *randomized* decision rule by selecting 1 w.p. p, and 2 w.p. $1-p$

Proofs

Consider two decision rules (omitting dependence on η):

$$\left(\mathsf{FPR}^{(1)}, \mathsf{TPR}^{(1)}\right) \ , \ \left(\mathsf{FPR}^{(2)}, \mathsf{TPR}^{(2)}\right)$$

Build a *randomized* decision rule by selecting 1 w.p. p, and 2 w.p. $1-p$

$$\left(\mathsf{FPR}, \mathsf{TPR}\right) = \left(p\,\mathsf{FPR}^{(1)} + (1-p)\mathsf{FPR}^{(2)},\ p\,\mathsf{TPR}^{(1)} + (1-p)\mathsf{TPR}^{(2)}\right)$$

Proofs

Consider two decision rules (omitting dependence on η):

$$\left(\mathsf{FPR}^{(1)}, \mathsf{TPR}^{(1)}\right) \quad , \quad \left(\mathsf{FPR}^{(2)}, \mathsf{TPR}^{(2)}\right)$$

Build a *randomized* decision rule by selecting 1 w.p. p, and 2 w.p. $1-p$

$$\left(\mathsf{FPR}, \mathsf{TPR}\right) = \left(p\,\mathsf{FPR}^{(1)} + (1-p)\mathsf{FPR}^{(2)},\ p\,\mathsf{TPR}^{(1)} + (1-p)\mathsf{TPR}^{(2)}\right)$$

Property 2: $\mathsf{TPR} \geq \mathsf{FPR}$

Proof:

Proofs

Consider two decision rules (omitting dependence on η):

$$\left(\mathsf{FPR}^{(1)}, \mathsf{TPR}^{(1)}\right) \;,\; \left(\mathsf{FPR}^{(2)}, \mathsf{TPR}^{(2)}\right)$$

Build a *randomized* decision rule by selecting 1 w.p. p, and 2 w.p. $1-p$

$$\left(\mathsf{FPR}, \mathsf{TPR}\right) = \left(p\,\mathsf{FPR}^{(1)} + (1-p)\mathsf{FPR}^{(2)},\; p\,\mathsf{TPR}^{(1)} + (1-p)\mathsf{TPR}^{(2)}\right)$$

Property 2: $\mathsf{TPR} \geq \mathsf{FPR}$

Proof:

- For any achievable $\alpha > 0$,

Proofs

Consider two decision rules (omitting dependence on η):

$$\left(\mathsf{FPR}^{(1)}\,,\ \mathsf{TPR}^{(1)}\right)\ \ ,\ \ \left(\mathsf{FPR}^{(2)}\,,\ \mathsf{TPR}^{(2)}\right)$$

Build a *randomized* decision rule by selecting 1 w.p. p, and 2 w.p. $1-p$

$$\left(\mathsf{FPR}\,,\ \mathsf{TPR}\right) = \left(p\,\mathsf{FPR}^{(1)} + (1-p)\mathsf{FPR}^{(2)}\,,\ p\,\mathsf{TPR}^{(1)} + (1-p)\mathsf{TPR}^{(2)}\right)$$

Property 2: $\mathsf{TPR} \geq \mathsf{FPR}$

Proof:

- For any achievable $\alpha > 0$, we can always find a randomized rule s.t.

$$\mathsf{TPR} = \mathsf{FPR} = \alpha$$

Proofs

Consider two decision rules (omitting dependence on η):

$$\left(\mathsf{FPR}^{(1)}, \mathsf{TPR}^{(1)}\right) \ , \ \left(\mathsf{FPR}^{(2)}, \mathsf{TPR}^{(2)}\right)$$

Build a *randomized* decision rule by selecting 1 w.p. p, and 2 w.p. $1-p$

$$\left(\mathsf{FPR}, \mathsf{TPR}\right) = \left(p\,\mathsf{FPR}^{(1)} + (1-p)\mathsf{FPR}^{(2)},\ p\,\mathsf{TPR}^{(1)} + (1-p)\mathsf{TPR}^{(2)}\right)$$

Property 2: $\mathsf{TPR} \geq \mathsf{FPR}$

Proof:

- For any achievable $\alpha > 0$, we can always find a randomized rule s.t.

$$\mathsf{TPR} = \mathsf{FPR} = \alpha$$

- By the Neyman-Pearson lemma,

Proofs

Consider two decision rules (omitting dependence on η):

$$\left(\mathsf{FPR}^{(1)}, \mathsf{TPR}^{(1)}\right) \;,\; \left(\mathsf{FPR}^{(2)}, \mathsf{TPR}^{(2)}\right)$$

Build a *randomized* decision rule by selecting 1 w.p. p, and 2 w.p. $1-p$

$$\left(\mathsf{FPR}, \mathsf{TPR}\right) = \left(p\,\mathsf{FPR}^{(1)} + (1-p)\mathsf{FPR}^{(2)},\; p\,\mathsf{TPR}^{(1)} + (1-p)\mathsf{TPR}^{(2)}\right)$$

Property 2: $\mathsf{TPR} \geq \mathsf{FPR}$

Proof:

- For any achievable $\alpha > 0$, we can always find a randomized rule s.t.

$$\mathsf{TPR} = \mathsf{FPR} = \alpha$$

- By the Neyman-Pearson lemma,

  if an LRT has $\mathsf{FPR}^{\star} = \alpha$,

Proofs

Consider two decision rules (omitting dependence on η):

$$\Big(\mathsf{FPR}^{(1)}\,,\ \mathsf{TPR}^{(1)}\Big)\quad,\quad\Big(\mathsf{FPR}^{(2)}\,,\ \mathsf{TPR}^{(2)}\Big)$$

Build a *randomized* decision rule by selecting 1 w.p. p, and 2 w.p. $1-p$

$$\big(\mathsf{FPR}\,,\ \mathsf{TPR}\big) = \Big(p\,\mathsf{FPR}^{(1)} + (1-p)\mathsf{FPR}^{(2)}\,,\ p\,\mathsf{TPR}^{(1)} + (1-p)\mathsf{TPR}^{(2)}\Big)$$

Property 2: $\mathsf{TPR} \geq \mathsf{FPR}$

Proof:

- For any achievable $\alpha > 0$, we can always find a randomized rule s.t.

$$\mathsf{TPR} = \mathsf{FPR} = \alpha$$

- By the Neyman-Pearson lemma,

  if an LRT has $\mathsf{FPR}^{\star} = \alpha$, then its $\mathsf{TPR}^{\star} \geq \alpha$

Proofs

Consider two decision rules (omitting dependence on η):

$$\Big(\mathsf{FPR}^{(1)}\,,\ \mathsf{TPR}^{(1)}\Big)\ \ ,\ \ \Big(\mathsf{FPR}^{(2)}\,,\ \mathsf{TPR}^{(2)}\Big)$$

Build a *randomized* decision rule by selecting 1 w.p. p, and 2 w.p. $1-p$

$$\big(\mathsf{FPR}\,,\ \mathsf{TPR}\big) = \Big(p\,\mathsf{FPR}^{(1)} + (1-p)\mathsf{FPR}^{(2)}\,,\ p\,\mathsf{TPR}^{(1)} + (1-p)\mathsf{TPR}^{(2)}\Big)$$

Property 2: $\mathsf{TPR} \geq \mathsf{FPR}$

Proof:

- For any achievable $\alpha > 0$, we can always find a randomized rule s.t.

$$\mathsf{TPR} = \mathsf{FPR} = \alpha$$

- By the Neyman-Pearson lemma,

  if an LRT has $\mathsf{FPR}^\star = \alpha$, then its $\mathsf{TPR}^\star \geq \alpha$ $\square$

Proofs

Property 3: The ROC curve is concave

Proof:

Proofs

Property 3: The ROC curve is concave

Proof:

- Consider two achievable points in the ROC diagram:

$$\big(\mathsf{FPR}(\eta_1)\,,\,\mathsf{TPR}(\eta_1)\big) \quad \text{and} \quad \big(\mathsf{FPR}(\eta_2)\,,\,\mathsf{TPR}(\eta_2)\big)$$

Proofs

Property 3: The ROC curve is concave

Proof:

- Consider two achievable points in the ROC diagram:

$$\big(\mathsf{FPR}(\eta_1)\,,\,\mathsf{TPR}(\eta_1)\big) \quad \text{and} \quad \big(\mathsf{FPR}(\eta_2)\,,\,\mathsf{TPR}(\eta_2)\big)$$

- For any $0 \leq t \leq 1$, we can form a randomized rule such that

$$\Big(t\,\mathsf{FPR}(\eta_1) + (1-t)\mathsf{FPR}(\eta_2)\ ,\ t\,\mathsf{TPR}(\eta_1) + (1-t)\mathsf{TPR}(\eta_2)\Big) \quad (1)$$

Proofs

Property 3: The ROC curve is concave

Proof:

- Consider two achievable points in the ROC diagram:

$$\big(\mathsf{FPR}(\eta_1)\,,\,\mathsf{TPR}(\eta_1)\big) \quad \text{and} \quad \big(\mathsf{FPR}(\eta_2)\,,\,\mathsf{TPR}(\eta_2)\big)$$

- For any $0 \leq t \leq 1$, we can form a randomized rule such that

$$\Big(t\,\mathsf{FPR}(\eta_1) + (1-t)\mathsf{FPR}(\eta_2)\;,\;t\,\mathsf{TPR}(\eta_1) + (1-t)\mathsf{TPR}(\eta_2)\Big) \qquad (1)$$

- By the Neyman-Pearson lemma, if an LRT has

$$\mathsf{FPR}^\star = t\,\mathsf{FPR}(\eta_1) + (1-t)\mathsf{FPR}(\eta_2)\;,$$

Proofs

Property 3: The ROC curve is concave

Proof:

- Consider two achievable points in the ROC diagram:

$$\big(\mathsf{FPR}(\eta_1)\,,\,\mathsf{TPR}(\eta_1)\big) \quad \text{and} \quad \big(\mathsf{FPR}(\eta_2)\,,\,\mathsf{TPR}(\eta_2)\big)$$

- For any $0 \le t \le 1$, we can form a randomized rule such that

$$\Big(t\,\mathsf{FPR}(\eta_1) + (1-t)\mathsf{FPR}(\eta_2)\;,\;t\,\mathsf{TPR}(\eta_1) + (1-t)\mathsf{TPR}(\eta_2)\Big) \quad (1)$$

- By the Neyman-Pearson lemma, if an LRT has

$$\mathsf{FPR}^\star = t\,\mathsf{FPR}(\eta_1) + (1-t)\mathsf{FPR}(\eta_2)\;, \qquad \text{then}$$

$$\mathsf{TPR}^\star \ge t\,\mathsf{TPR}(\eta_1) + (1-t)\mathsf{TPR}(\eta_2)$$

Proofs

Property 3: The ROC curve is concave

Proof:

- Consider two achievable points in the ROC diagram:

$$\big(\mathsf{FPR}(\eta_1)\,,\,\mathsf{TPR}(\eta_1)\big) \quad \text{and} \quad \big(\mathsf{FPR}(\eta_2)\,,\,\mathsf{TPR}(\eta_2)\big)$$

- For any $0 \le t \le 1$, we can form a randomized rule such that

$$\Big(t\,\mathsf{FPR}(\eta_1) + (1-t)\mathsf{FPR}(\eta_2)\ ,\ t\,\mathsf{TPR}(\eta_1) + (1-t)\mathsf{TPR}(\eta_2)\Big) \quad (1)$$

- By the Neyman-Pearson lemma, if an LRT has

$$\mathsf{FPR}^\star = t\,\mathsf{FPR}(\eta_1) + (1-t)\mathsf{FPR}(\eta_2)\ , \qquad \text{then}$$

$$\mathsf{TPR}^\star \ge t\,\mathsf{TPR}(\eta_1) + (1-t)\mathsf{TPR}(\eta_2)$$

Thus, the ROC curve is above (1)

Proofs

Property 3: The ROC curve is concave

Proof:

- Consider two achievable points in the ROC diagram:

$$\big(\mathsf{FPR}(\eta_1)\,,\,\mathsf{TPR}(\eta_1)\big) \quad \text{and} \quad \big(\mathsf{FPR}(\eta_2)\,,\,\mathsf{TPR}(\eta_2)\big)$$

- For any $0 \le t \le 1$, we can form a randomized rule such that

$$\Big(t\,\mathsf{FPR}(\eta_1) + (1-t)\mathsf{FPR}(\eta_2)\ ,\ t\,\mathsf{TPR}(\eta_1) + (1-t)\mathsf{TPR}(\eta_2)\Big) \quad (1)$$

- By the Neyman-Pearson lemma, if an LRT has

$$\mathsf{FPR}^\star = t\,\mathsf{FPR}(\eta_1) + (1-t)\mathsf{FPR}(\eta_2)\ , \qquad \text{then}$$

$$\mathsf{TPR}^\star \ge t\,\mathsf{TPR}(\eta_1) + (1-t)\mathsf{TPR}(\eta_2)$$

Thus, the ROC curve is above (1) $\square$

Example: SARS-CoV-2 Tests

Example: SARS-CoV-2 Tests

Lab-based tests (ELISA, CLIA) and rapid tests (lateral flow)

Example: SARS-CoV-2 Tests

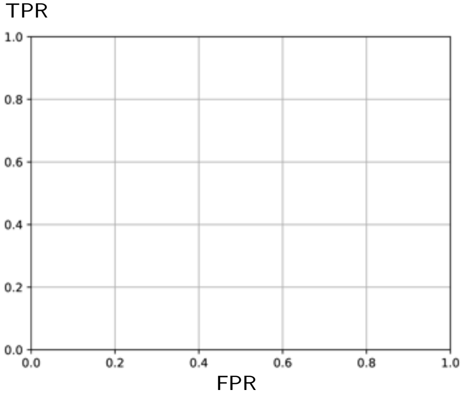
Lab-based tests (ELISA, CLIA) and rapid tests (lateral flow)

Detection of IgG, IgM, or IgG/IgM antibodies at days 8-14, 15-21, 22-35 (95% CI)

Example: SARS-CoV-2 Tests

Lab-based tests (ELISA, CLIA) and rapid tests (lateral flow)

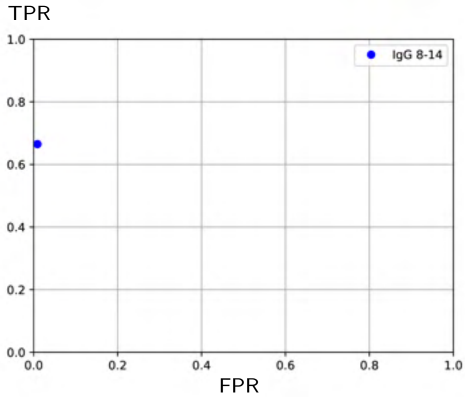
Detection of IgG, IgM, or IgG/IgM antibodies at days 8-14, 15-21, 22-35 (95% CI)



Example: SARS-CoV-2 Tests

Lab-based tests (ELISA, CLIA) and rapid tests (lateral flow)

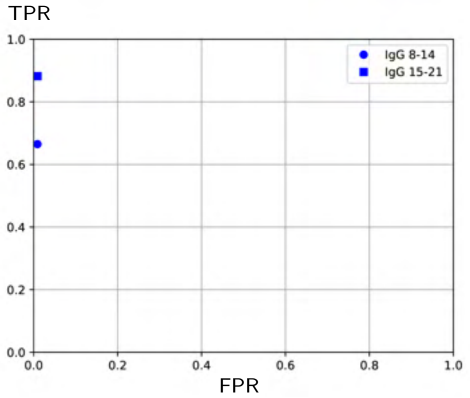
Detection of IgG, IgM, or IgG/IgM antibodies at days 8-14, 15-21, 22-35 (95% CI)



Example: SARS-CoV-2 Tests

Lab-based tests (ELISA, CLIA) and rapid tests (lateral flow)

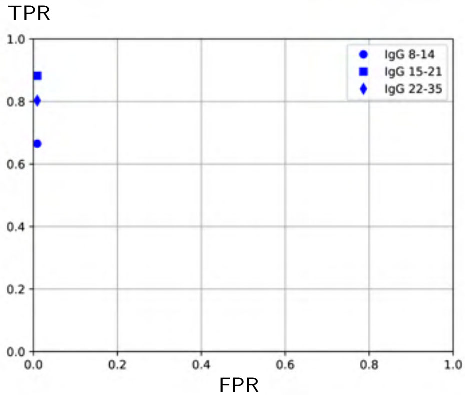
Detection of IgG, IgM, or IgG/IgM antibodies at days 8-14, 15-21, 22-35 (95% CI)



Example: SARS-CoV-2 Tests

Lab-based tests (ELISA, CLIA) and rapid tests (lateral flow)

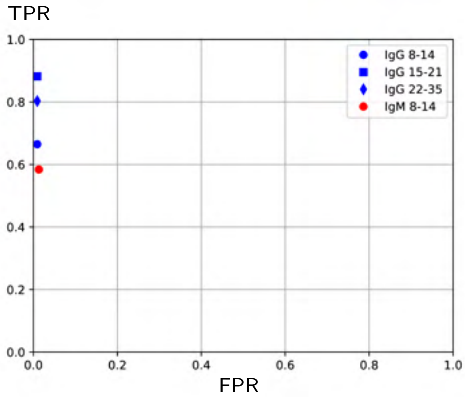
Detection of IgG, IgM, or IgG/IgM antibodies at days 8-14, 15-21, 22-35 (95% CI)



Example: SARS-CoV-2 Tests

Lab-based tests (ELISA, CLIA) and rapid tests (lateral flow)

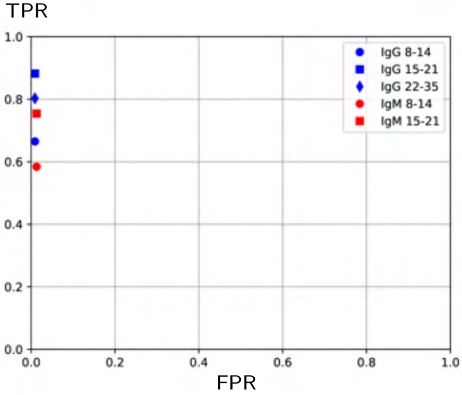
Detection of IgG, IgM, or IgG/IgM antibodies at days 8-14, 15-21, 22-35 (95% CI)



Example: SARS-CoV-2 Tests

Lab-based tests (ELISA, CLIA) and rapid tests (lateral flow)

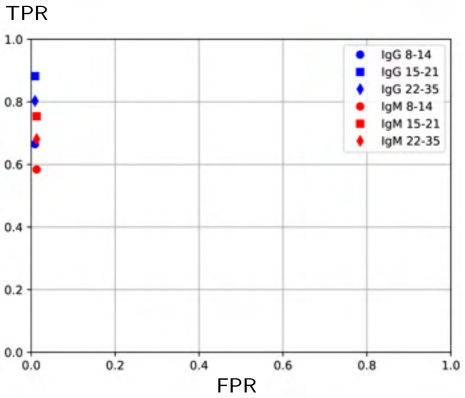
Detection of IgG, IgM, or IgG/IgM antibodies at days 8-14, 15-21, 22-35 (95% CI)



Example: SARS-CoV-2 Tests

Lab-based tests (ELISA, CLIA) and rapid tests (lateral flow)

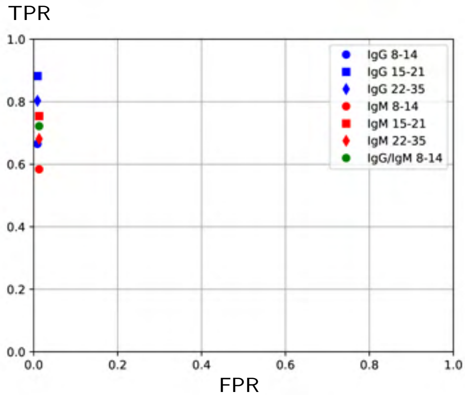
Detection of IgG, IgM, or IgG/IgM antibodies at days 8-14, 15-21, 22-35 (95% CI)



Example: SARS-CoV-2 Tests

Lab-based tests (ELISA, CLIA) and rapid tests (lateral flow)

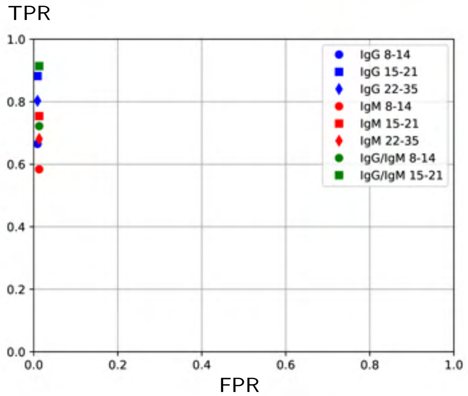
Detection of IgG, IgM, or IgG/IgM antibodies at days 8-14, 15-21, 22-35 (95% CI)



Example: SARS-CoV-2 Tests

Lab-based tests (ELISA, CLIA) and rapid tests (lateral flow)

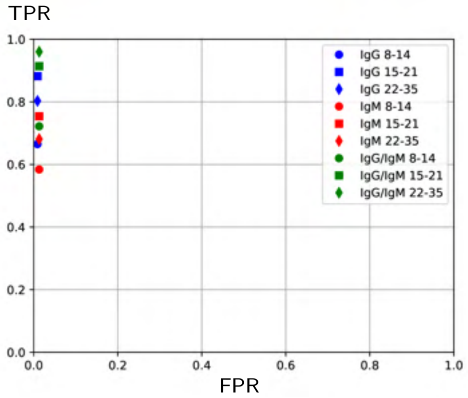
Detection of IgG, IgM, or IgG/IgM antibodies at days 8-14, 15-21, 22-35 (95% CI)



Example: SARS-CoV-2 Tests

Lab-based tests (ELISA, CLIA) and rapid tests (lateral flow)

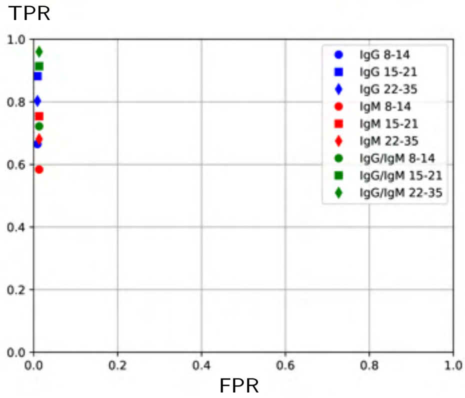
Detection of IgG, IgM, or IgG/IgM antibodies at days 8-14, 15-21, 22-35 (95% CI)



Example: SARS-CoV-2 Tests

Lab-based tests (ELISA, CLIA) and rapid tests (lateral flow)

Detection of IgG, IgM, or IgG/IgM antibodies at days 8-14, 15-21, 22-35 (95% CI)



Deeks et al, **Antibody tests for identification of current and past infection with SARSCoV2**, Cochrane Database of Systematic Reviews, Issue 6, 2020

Looking Ahead: Empirical Risk Minimization

Looking Ahead: Empirical Risk Minimization

We studied a (binary) decision problem:

$$\underset{f:\mathbb{R}^d \to \{0,1\}}{\text{minimize}} \; \mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}), Y\big)\Big]$$

Looking Ahead: Empirical Risk Minimization

We studied a (binary) decision problem:

$$\underset{f:\mathbb{R}^d \to \{0,1\}}{\text{minimize}} \;\; \mathbb{E}_{\boldsymbol{X}Y}\left[\ell\big(f(\boldsymbol{X}), Y\big)\right]$$

Assumed known $f_{\boldsymbol{X}\,|\,H_i}(\boldsymbol{x}\,|\,H_i)$ and $\mathbb{P}(H_i)$

Looking Ahead: Empirical Risk Minimization

We studied a (binary) decision problem:

$$\underset{f:\mathbb{R}^d \to \{0,1\}}{\text{minimize}} \;\; \mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}), Y\big)\Big]$$

Assumed known $f_{\boldsymbol{X}\,|\,H_i}(\boldsymbol{x}\,|\,H_i)$ and $\mathbb{P}(H_i)$

What if they are unknown?

Looking Ahead: Empirical Risk Minimization

We studied a (binary) decision problem:

$$\underset{f:\mathbb{R}^d \to \{0,1\}}{\text{minimize}} \; \mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}), Y\big)\Big]$$

Assumed known $f_{\boldsymbol{X}\,|\,H_i}(\boldsymbol{x}\,|\,H_i)$ and $\mathbb{P}(H_i)$

What if they are unknown?

Empirical Risk Minimization (ERM):

Looking Ahead: Empirical Risk Minimization

We studied a (binary) decision problem:

$$\underset{f:\mathbb{R}^d \to \{0,1\}}{\text{minimize}} \;\; \mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}), Y\big)\Big]$$

Assumed known $f_{\boldsymbol{X}\,|\,H_i}(\boldsymbol{x}\,|\,H_i)$ and $\mathbb{P}(H_i)$

What if they are unknown?

Empirical Risk Minimization (ERM):

$$\underset{f:\mathbb{R}^d \to \{0,1\}}{\text{minimize}} \;\; \frac{1}{T}\sum_{t=1}^{T}\Big[\ell\big(f(\boldsymbol{x}_t), y_t\big)\Big]$$

Looking Ahead: Empirical Risk Minimization

We studied a (binary) decision problem:

$$\boxed{\underset{f:\mathbb{R}^d \to \{0,1\}}{\text{minimize}} \;\; \mathbb{E}_{\boldsymbol{X}Y}\Big[\ell\big(f(\boldsymbol{X}), Y\big)\Big]}$$

Assumed known $f_{\boldsymbol{X}\,|\,H_i}(\boldsymbol{x}\,|\,H_i)$ and $\mathbb{P}(H_i)$

What if they are unknown?

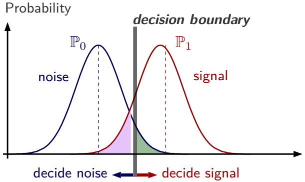
Empirical Risk Minimization (ERM):

$$\boxed{\underset{f:\mathbb{R}^d \to \{0,1\}}{\text{minimize}} \;\; \frac{1}{T}\sum_{t=1}^{T}\Big[\ell\big(f(\boldsymbol{x}_t), y_t\big)\Big]}$$

Assumption: we observe T samples $\{(\boldsymbol{x}_t,\, y_t)\}_{t=1}^{T}$

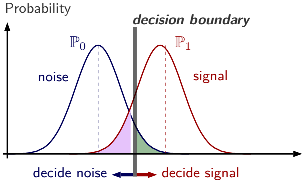
Conclusions

Conclusions



Optimal decision problems (binary case)

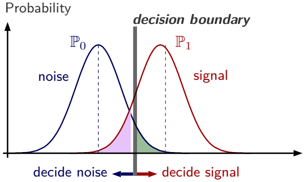
Conclusions



Optimal decision problems (binary case)

Likelihood ratio tests (LRT)

Conclusions

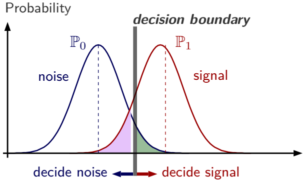


Optimal decision problems (binary case)

Likelihood ratio tests (LRT)

MAP and ML as particular cases

Conclusions

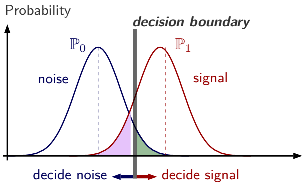


Optimal decision problems (binary case)

Likelihood ratio tests (LRT)

- MAP and ML as particular cases
- Optimality (Neyman-Pearson lemma)

Conclusions



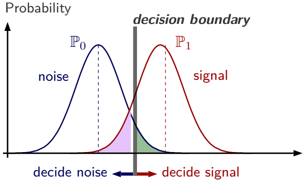
Optimal decision problems (binary case)

Likelihood ratio tests (LRT)

- MAP and ML as particular cases
- Optimality (Neyman-Pearson lemma)

Key assumption: known probability distributions

Conclusions



Optimal decision problems (binary case)

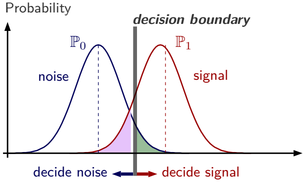
Likelihood ratio tests (LRT)

- MAP and ML as particular cases
- Optimality (Neyman-Pearson lemma)

Key assumption: known probability distributions

Types of errors

Conclusions



Optimal decision problems (binary case)

Likelihood ratio tests (LRT)

- MAP and ML as particular cases
- Optimality (Neyman-Pearson lemma)

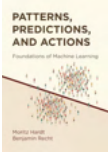
Key assumption: known probability distributions

Types of errors

ROC curves and properties

References

References

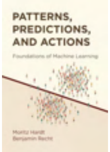


M. Hardt, B. Recht

Patterns, Predictions, and Actions

Princeton University Press, Oct, 2022

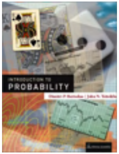
References



M. Hardt, B. Recht

Patterns, Predictions, and Actions

Princeton University Press, Oct, 2022



D. P. Bertsekas, J. N. Tsitsiklis

Introduction to Probability

Athena Scientific, 2nd edition, 2008

T. Fauwcett, “An introduction to ROC analysis,” Pattern Recognition Letters,

Vol. 27, pp. 861-874, 2006.

H. L. Van Trees, **Detection, Estimation, and Modulation: Part I**,

John Wiley & Sons, 2001.