# Multi-label Classification: A Deep Sparse Decomposition Perspective

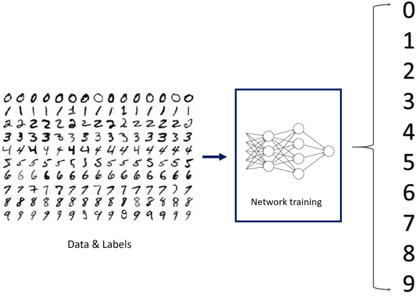
**Mehrdad Yaghoobi**
**In collaboration with: Mike E. Davies, Cecile Chenot and Yuqi Wang**
*Institute for Digital Communications, University of Edinburgh*

UDRC Themed Meeting on Deep Learning and Defence,
Belfast, UK, 14 November, 2019

EPSRC
Engineering and Physical Sciences Research Council

# Multi-class vs Multi-label Classification

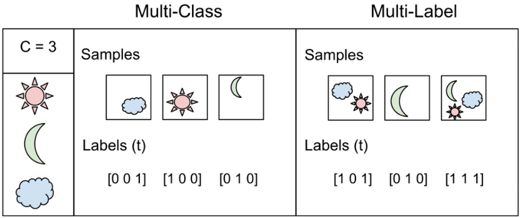
# Multi-class Classification




https://bit.ly/2X0xMbz

- **Multi-class** classification; the most popular ML task, i.e. finding the corresponding class to a given sample, among different classes of objects.
- Has been usually learned **supervised** or **semi-supervised**.
- It can be cast as nearest neighbour search in a feature space.

# Multi-label Classification




https://bit.ly/2WYKIif

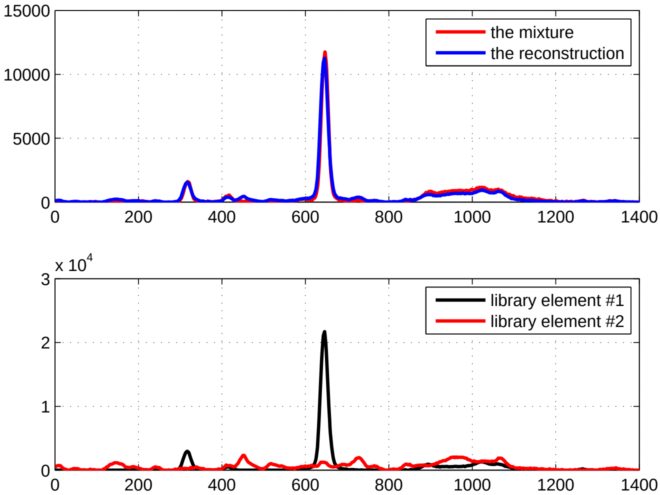
- **Multi-label** classification: secondary classes are **not mutually exclusive**, i.e. each given sample may have multiple associated primary classes.
- The number of secondary classes can become huge. Solving the problem is computationally intractable, i.e. NP-hard .
- Under some **dissimilarity criterion** between classes, the multi-label classification becomes a polynomial time .

# Exemplar Applications

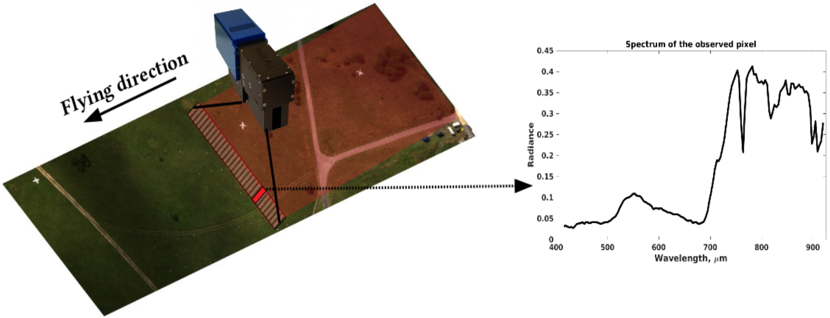
# Raman Spectroscopy


- Raman Spectrometers measure the **spectral shifts** of the transmitted laser beam.
- Identifying chemical components of samples from spectral data.
- Low computational power and high accuracy in hand-help spectrometers.

# Raman Spectral Decomposition

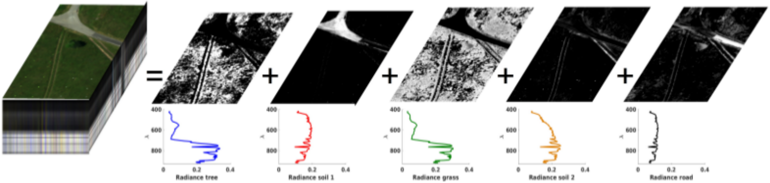


# Hyperspectral Imagery




- (Push broom) Hyperspectral Imagers measure the **spectral features** of the reflected lights.
- Big data problem; how to classify the materials for real-time clustering and object detection.

# Hyperspectral Image Analysis

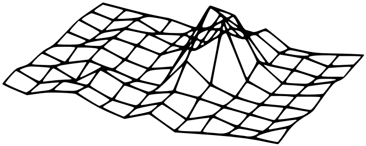


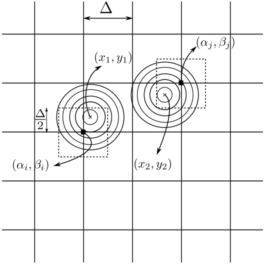

**Decomposing the spectral images or strips to the constituent spectra**

# Capacitive Touch Sensors


- Measuring the changes in the electrical field.
- The most popular multitouch interface for smart devices.
- Low computational power and real-time.

# Multitouch Sensing

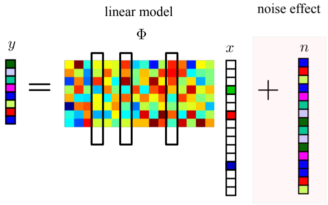





$$\begin{aligned} t(x,y) &\approx \sum_{1 \leq i \leq K} a_i e^{-\frac{(x-x_i)^2+(y-y_i)^2}{\sigma^2}} \\ &= e^{-\frac{x^2+y^2}{\sigma^2}} * \sum_{1 \leq i \leq K} a_i \delta_{x_i, x_j}(x,y) \\ &= g(x,y) * d(x,y) \end{aligned}$$

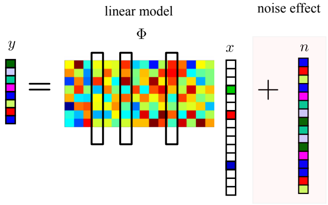
# Non-Negative Sparse Model for Multi-label Classification

# Non-negative Sparse Decomposition




- **y** is generated by weighted superposition of a few columns of **Φ**, *i.e.* atoms. **Φ** is called a dictionary. **x** holds the positive weights and **n** is an additive noise.
- Sparse decomposition is generally an NP-hard problem.

# Multi-label Classification with a Linear Mixing Model




- **y** can represent the convoluted sample, with labels $\mathcal{S}_y$.
- $\mathcal{S}_y$ is associated with the support of **x**, i.e. **non-negative element indices**.
- Columns of $\mathbf{\Phi}$, $\phi$, represent classes centre of masses.
- "Non-negative sparse decomposition" then becomes a multi-label classification problem.

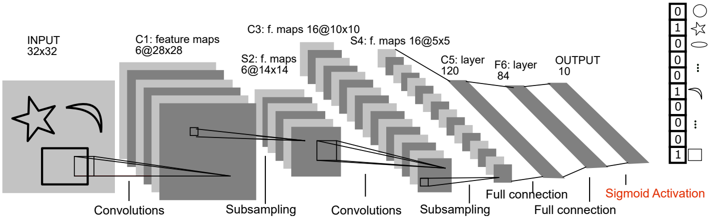
# Sparse Approximation Algorithms

- **Approximately** solving the problem, using greedy or iterative algorithms, or **problem relaxations**, *e.g.* $\ell_1$ relaxation.
- **Non-negative MP (NNMP)** is a simple greedy algorithm based on iteratively selecting maximally correlated to the residual signal.

**initialisation:** $s = \emptyset$, $k = 0$ and $\mathbf{r}_0 = \mathbf{y}$
**while** $k < K \quad \max(\boldsymbol{\Phi}^T \mathbf{r}_k) > 0$ **do**
  $\mathbf{s}_k = \mathbf{0}$
  $(\zeta, \iota) \leftarrow \max(\boldsymbol{\Phi}^T \mathbf{r}_k)$
  $\mathbf{s}_k[\iota] = \zeta$
  $\mathbf{r}_{k+1} \leftarrow \mathcal{P}\{\mathbf{r}_k - \zeta \phi_\iota\}$
  $k \leftarrow k + 1$
**end while**
$\mathbf{x} \leftarrow \sum_k \mathbf{s}_k$

# Deep Neural Networks

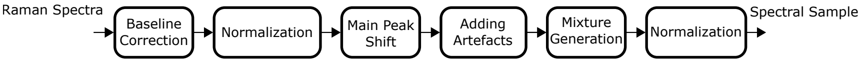
# Input to Representation Model

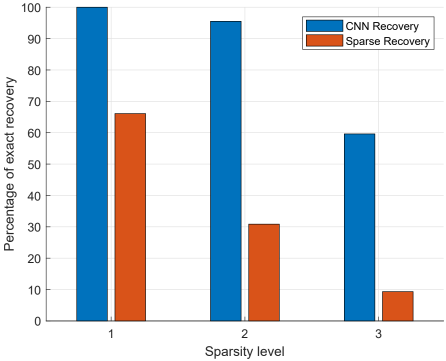



(Modified) LeCun5 CNN

- An end to end deep neural network for multi-label classification.
- The conventional **Softmax** output layer is replaced with **Sigmoid**.
- Supervised training should be done with **many** multi-label data. Data is generally augmented.

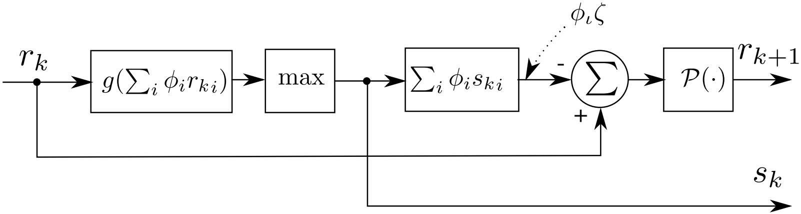
# Robust Raman Spectral Decomposition Result

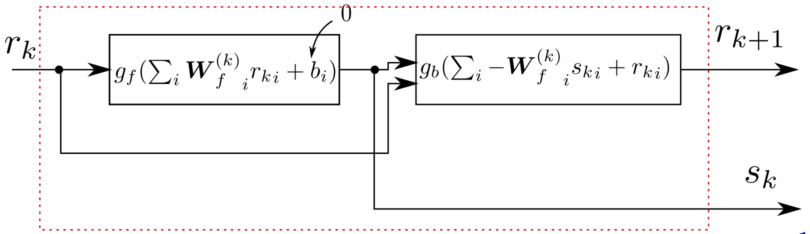


# NNMP: From Greedy to Deep Neural Networks

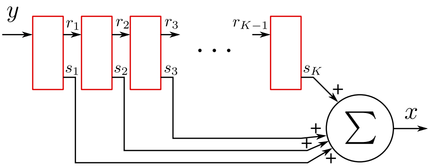
$(\zeta, \iota) \leftarrow \max(\mathbf{\Phi}^T \mathbf{r}_k)$ $\qquad$ $\mathbf{r}_{k+1} \leftarrow \mathcal{P}\{\mathbf{r}_k - \zeta \boldsymbol{\phi}_\iota\}$



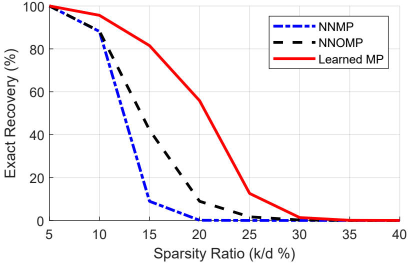
# Block Diagram and Update Rules

$$\begin{pmatrix} \mathbf{r}_{k+1} \\ \mathbf{x}_{k+1} \end{pmatrix} = g_b \begin{pmatrix} -\mathbf{W}_b^{(k)} g_f(\mathbf{W}_f^{(k)}{}^T \mathbf{r}_k) + \mathbf{r}_k \\ g_f(\mathbf{W}_f^{(k)}{}^T \mathbf{r}_k) + \mathbf{x}_k \end{pmatrix}$$

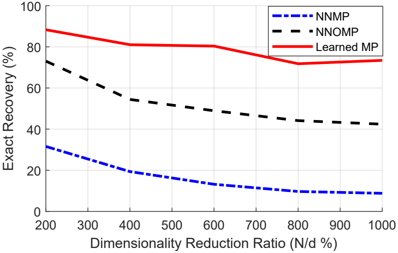



- $\mathbf{W}_f^{(k)}$ and $\mathbf{W}_b^{(k)}$ are weight matrices in $k^{th}$ layer.
- $g_f$ is **hard-max** and $g_b$ is **ReLU** nonlinear operators.
- Back-propagation algorithm has been used with the smoothed activation functions for training the weights.

# Multi-label Classification: Synthetic Data Experiments




$\mathbf{\Phi} \in \mathbb{R}^{20\times 40}$




$|\mathcal{S}| = 3,\ \mathbf{\Phi} \in \mathbb{R}^{20\times 40-200}$

# Conclusion and Future Work

- **Conclusion:**
  - Some defence classification tasks are multi-label classifications
  - Conventional sparse decompositions can be used. Lack of robustness to the model mismatch
  - Deep neural networks can be good data-adaptive alternatives
  - Non-negative Matching Pursuit algorithm can be reformulated as a DNN and trained to better fit to the task
- **Future work:**
  - More **real data** experiments
  - **Sample complexity** and **accuracy analysis** of DeepMP
  - **Robustness verification** of the DNNs in multi-label classifications

**We gratefully acknowledge the support from:**

UDRC

EPSRC
Engineering and Physical Sciences
Research Council

[dstl]

# Thanks for your attention.