Classification of sets

- Variable number of instances (documents, sections, zones of interest, conformations) in a input set ('bag')
- Straightforward generalization of neural networks: max pooling of outputs/inputs
- Proposal:
  - max pooling at an intermediate level (hidden features)
  - application to Restricted Boltzmann Machines (RBM)

Examples of such inputs:
- work done at the University of Toronto
- Several zones of interest in an image (blobs)
- Several parts in a text

Extension of Classification RBM for variable-size sets (of vectors)

Class. RBM

Class. Set RBM

Class. Set RBM [XOR]

Class. Set RBM [OR]

Energy function - defines a joint probability \( p(\ldots) = \exp(-E(\ldots)/Z) \)

\[
E(x, h, y) = -b^T x - c^T h - d^T y - h^T W x - h^T U y
\]

\( \forall j, \sum h^j \in (0, 1) \)

Bottom Up Inference

\[
p(h = 0 | x) = \text{sigmoid}(c + W x + U y)
\]

\[
p(h = 1 | x) = \frac{\exp(\text{act}(x^h)) + 1}{1 + \sum \exp(\text{act}(x^h)) + U y}
\]

\[
p(h = 0 | y) = 1 / (1 + \sum \exp(\text{act}(x^h)) + U y)
\]

\[
p(h = 1 | h) = \text{sigmoid}(h + b + H y)
\]

\[
p(y = 1 | h) = \frac{\text{sigmoid}(h + b + H y)}{1 + \sum \text{sigmoid}(h + b + H y)}
\]

\[
p(y = 0 | h) = \frac{1 - \text{sigmoid}(h + b + H y)}{1 + \sum \text{sigmoid}(h + b + H y)}
\]

Class posterior probabilities for discriminative training

\[
p(y | X) = \frac{\exp(-FE(X, y))}{\sum_{y'} \exp(-FE(X, y'))}
\]

\[
FE(X, y) = - \sum \text{softmax}(\text{act}(x^h)) + U y
\]

Training

MAX pooling approximation

Experiments: Multi-Instance Learning (MIL)

- Binary classification (detection)
- 166 / 230 features, 6 / 10 instances per bag
- 5 x 5-fold cross validation on small datasets (≤ 200 bags)

Table of accuracies (%)

Experiments: Multi-page mail paper classification

- Multi-class classification (6 / 11 classes)
- Typically between 4 and 20 scanned pages per mail
- 5-fold cross validation with 10/15000 mails
- Sub-resolution image (6 x 8 grid)
- Automatic Word Recognition: Sparse and noisy binary bag-of-words, with the 10,000 most frequent words

Conclusions

- Class.Set RBM [XOR] is very competitive, and much faster than SVM with kernels of bags (miGraph or localMax).
- The MAX pooling approximation is efficient, and effective for Class.Set RBM [XOR] (but do not necessarily improves performance for RBM (OR)).
- Future Work on generative/hybrid learning:
  - Experiment non-binary input units on MIL datasets.
  - Experiment simplification to be tractable for text classification (high-dimensional inputs).