Classification task

Given: \( \{ [x_1, y_1], [x_2, y_2], \ldots, [x_n, y_n] \} \)
- \( x_i \) - feature vector
- \( y_i \in \{-1, 1\} \) - label

The task is: To design \( y = H(x) \) that predicts a label \( y \) for a new vector \( x \)
- Many approaches proposed (Boosting, Random Forests, Neural Nets, SVMs, ..)
- SVM seems to dominate the others

Locally Linear SVMs

- The decision boundary should be smooth
  - Approximately linear in a sufficiently small region
- Curvature is bounded
  - Functions \( w(x) \) and \( b(x) \) are Lipschitz
- \( w(x) \) and \( b(x) \) can be approximated using local coding

\[
H(x) = \sum_{i=1}^{n} \sum_{v \in \mathbb{V}} \gamma_v(x) w_v(x) \phi(v) + \sum_{v \in \mathbb{V}} \gamma_v(b) \phi(v)
\]

\[
= \gamma(x)^T W x + \gamma(x)^T b
\]

- Weights \( W \) and biases \( b \) are obtained as:

\[
\arg \min_{W,B} \frac{1}{2} ||W||^2 + \frac{1}{|S|} \sum_{k=1}^{S} \max(0, 1 - y_k H_w(b(x_k)))
\]

where

\[
\gamma_v(x) = \gamma_v(x) \phi(v)
\]

Local coding for manifold learning

- Points approximated as a weighted sum of anchor points
  - \( x \approx \sum_{v \in \mathbb{V}} \gamma_v(x) \phi(v) \)
- Coordinates obtained using:
  - Distance based methods (Gemert et al, 2008, Zhou et al, 2009)

- For a normalised coding any Lipschitz function \( f \) can be approximated:
  - \( \gamma_v(x) = \gamma_v(x) \phi(v) \)

(Kyu et al, 2009)

Relation to Other Models

- Generalisation of Linear SVM on \( x \):
  - \( H(x) = w^T x + b \)
  - \( w \) and \( b \) approximated using local coding
- Generalisation of linear SVM on \( y \):
  - \( H(x) = \sum_{v \in \mathbb{V}} \gamma_v(x) \phi(v) \)
  - \( \gamma_v(x) \) approximated using local coding
- Generalisation of model selecting Latent(MI) SVM:
  - \( H(x) = w^T x + b \)
  - \( w \) and \( b \) approximated using local coding

Locally Linear Support Vector Machines

Lubor Ladicky Philip H.S. Torr

Extension to Finite Kernels

- The finite kernel classifier takes the form:

\[
H(x) = \gamma(x)^T W(x) + \gamma(x)^T \tilde{b}
\]

- Weights \( W \) and \( \tilde{b} \) obtained as:

\[
\arg \min_{W,B} \frac{1}{2} ||W||^2 + \frac{1}{|S|} \sum_{k=1}^{S} \max(0, 1 - y_k H_w(b(x_k)))
\]

Experiments

- MNIST, LETTER & USPS datasets
  - Anchor points obtained using K-means clustering
  - Coordinates evaluated on kNN (k = 8) (slow part)
  - Spatial pyramid of BOW features
  - Coordinates evaluated based on image histograms
- CALTECH-101 (15 samples per class)
  - Coordinates evaluated on kNN (k = 5)
  - Approximated intersection kernel used
  - Raw data used

Linear SVMs
- Fast training and evaluation
- Applicable to large scale data sets
- Low discriminative power

Kernel SVMs
- Slow training and evaluation
- Not feasible for large scale data sets
- Better performance for hard problems

Motivation:
- Exploiting manifold learning techniques to design an SVM with
  - Good trade-off between the performance and speed
  - Scalability to large scale data sets (solvable with stochastic gradient descent)

Local coding for manifold learning

- Points approximated as a weighted sum of anchor points
  - \( x \approx \sum_{v \in \mathbb{V}} \gamma_v(x) \phi(v) \)
- Coordinates obtained using:
  - Distance based methods (Gemert et al, 2008, Zhou et al, 2009)

- For a normalised coding any Lipschitz function \( f \) can be approximated:
  - \( \gamma_v(x) = \gamma_v(x) \phi(v) \)

(Kyu et al, 2009)