1st PIXL Retreat

Support Vector Decomposition Machine

Yongxin (Taylor) Xi

Working with Prof. Peter Ramadge
& NIAM Group
Data from Prof. Jim Haxby

Core Idea of SVDM

- For a learning problem, suppose there are n data samples $x_i \in R^m$ with labels $y_i \in \{0,1\}$

Possible Methods:

- Classification on original data
- Dimension reduction then classification (noisy, redundant, latent variables)
- Dimension reduction and classification concurrently (SVDM)

Background: SVD & SVM

- SVD

$$X_{nxm} = U_{nxn} \Sigma_{nxn} V^T{}_{nxm}$$

$$U^T U = I \qquad V^T V = I$$

Approximate X by two lower dimensional matrices:

$$X_{nxm} \approx Z_{nxl} W_{lxm}$$

Background: SVD & SVM

- SVM (non-separable Case):

$$Min \quad \frac{1}{2} \| q \|^2 + C \sum_{i=1}^{n} h_i$$

$$Subject\ to: \ y_i (q^T x_i + b) \geq 1 - h_i, \quad i = 1, 2, ..., n$$

$$h_i \geq 0, \forall i$$

- Equivalent to:

$$Min \quad \frac{1}{2} \| q \|^2 + C \sum_{i=1}^{n} \max(0, 1 - y_i (q^T x_i + b))$$

SVDM Problem formulation (Pereira 06)

$$\min_{Z,W,Q} \| X - ZW \|_F^2 + \lambda \sum_{i=1}^{n} \sum_{j=1}^{K} \max(0, \mu - Y_{ij}[ZQ]_{ij})$$

$$Subject\ to: \quad Z_{i,1} = 1$$

$$Z_{i,2:end} \leq 1 \ \ , i = 1, 2, ..., n$$

$$\| Q_{:,j} \|^2 \leq 1, \ i = 1, 2, ..., l; \ \ \text{j=1,2,....,k}$$

$$\lambda > 0, u > 0, l \in Z^{++}$$

$$X_{nxm} \in R^{nxm}, Y_{nxK} \in \{1, -1\}^{nxK}$$

$$Z \in R^{nxl}, W \in R^{lxm}, Q \in R^{lxK}$$

What's changed

- $\|Q\|_F^2$ omitted from the objective
- $\|Q_{:,j}\|^2 \leq 1$ instead of $\mathrm{Min}\,\|Q_{:,j}\|^2$ regulates the margin in each SVM to be greater than 2.
- Intercept b removed
- Threshold changed from 1 to parameter u

Alternating Optimizing Scheme

- Fixing two of the variable matrices while optimizing the other

$$\min_{Z,W,Q} \| X - ZW \|_F^2 + \lambda \sum_{i=1}^{n} \sum_{j=1}^{K} \max(0, \mu - Y_{ij}[ZQ]_{ij})$$

- Given Z, Q, solve for W—Linear regression
- Given Z, W, solve for Q—LP problem
- Given W, Q, solve for Z—QP problem
- Converge to local optimum

Testing methods

- **Cross Validation (leave one out)**

  a. 7 category test (ManF, WomanF, MonkeyF, DogF, House, Shoe, Chair)

  b. 2 category test (face versus object)

- **Independent Testing**

  Inter - subjects

  I. Testing on new subjects (train on first 70% subjects, test on 30%)

  - a. 7 category test
  - b. 2 category test (face versus object)

  II. Testing on new runs (train on first 75% collectively from each subject)

  - a. 7 category test
  - b. 2 category test (face versus object)

  Intra - subjects *(train on first 75% of one subject only)*

  - a. 7 category test
  - b. 2 category test (face versus object)

Testing results

<table>
<tr><th colspan="2">Test of SVDM on new data</th><th>7 Categories (K=7)</th><th>2 Categories (K=2)</th></tr>
<tr><td colspan="2">Random guessing</td><td>14.28%</td><td>50%</td></tr>
<tr><td rowspan="2">Independent among subjects</td><td>Generalization on new subject</td><td>30%</td><td>69%</td></tr>
<tr><td>Generalization on new sample</td><td>48%</td><td>93.9%</td></tr>
<tr><td colspan="2">Independent within subjects</td><td>53%</td><td>92%</td></tr>
<tr><td colspan="2">Cross-Validation</td><td>50%</td><td>81%</td></tr>
</table>

An interesting experiment

Scheme:

Randomly select 70% data of two categories from all the subjects to train the classifier, and test on the remaining 30%.

accuracy	F face	M face	Monkey	Dog	House	Chair	Shoe
F face	0	0.46429	0.60714	0.64286	0.85714	0.89286	0.67857
M face	0.46429	0	0.64286	0.60714	0.89286	0.96429	0.82143
Monkey	0.60714	0.64286	0	0.67857	0.75	0.75	0.78571
Dog	0.64286	0.60714	0.67857	0	0.96429	0.96429	0.92857
House	0.85714	0.89286	0.75	0.96429	0	0.75	0.75
Chair	0.89286	0.96429	0.75	0.96429	0.75	0	0.64286
Shoe	0.67857	0.82143	0.78571	0.92857	0.75	0.64286	0

Problem Reformulation

- Original problem *(Francisco Pereira 06)*

$$\underset{Z,W,Q}{Min} \| X - ZW \|_F^2 + \lambda \sum_{i=1}^{n} \sum_{j=1}^{K} \max(0, \mu - Y_{ij}[ZQ]_{ij})$$

New formulation

$$\underset{Z,W,Q}{Min} \| X - ZW \|_F^2 + \lambda (\frac{1}{2} \| Q \|_F^2 + C \sum_{i=1}^{n} \sum_{j=1}^{K} \max(0, 1 - Y_{ij}([ZQ]_{ij} + b_j)))$$

$$Subject\ to: \quad |W(i,:)|_2 = 1 \ , \ i = 1, 2, ..., l$$

1. Standard SVM (to maximize the margin of classes)
2. In this way u can be set to 1, rather than tedious searching for the best parameter.
3. With bj, j=1,2,…K added

accuracy	F face	M face	Monkey	Dog	House	Chair	Shoe
F face	0	0.46429	0.60714	0.64286	0.85714	0.89286	0.67857
M face	0.46429	0	0.64286	0.60714	0.89286	0.96429	0.82143
Monkey	0.60714	0.64286	0	0.67857	0.75	0.75	0.78571
Dog	0.64286	0.60714	0.67857	0	0.96429	0.96429	0.92857
House	0.85714	0.89286	0.75	0.96429	0	0.75	0.75
Chair	0.89286	0.96429	0.75	0.96429	0.75	0	0.64286
Shoe	0.67857	0.82143	0.78571	0.92857	0.75	0.64286	0

Table 2. Result by Pereira's formulation

accuracy	F face	M face	Monkey	Dog	House	Chair	Shoe
F face	0	0.57143	0.67857	0.57143	0.82143	0.92857	0.67857
M face	0.57143	0	0.75	0.53571	0.96429	0.96429	0.85714
Monkey	0.67857	0.75	0	0.78571	0.85714	0.78571	0.82143
Dog	0.57143	0.53571	0.78571	0	0.96429	0.96429	0.96429
House	0.82143	0.96429	0.85714	0.96429	0	0.75	0.85714
Chair	0.92857	0.96429	0.78571	0.96429	0.75	0	0.64286
Shoe	0.67857	0.85714	0.82143	0.96429	0.85714	0.64286	0

Table 3. Result by the modified formulation

24/42 classifiers increase in performance; 12/42 classifiers remain the same; 6/42 classifiers decrease

Thank you!

Q & A