Multiview Fisher Discriminant Analysis

Tom Diethe, David R. Hardoon, John Shawe-Taylor

{t.diethe,d.hardoon,j.shawe-taylor}@cs.ucl.ac.uk

Department of Computer Science
University College London

13/12/2008

Outline

Classification Function

- FDA is Bayes optimal for two normal distributions with equal covariance (KFDA is Bayes optimal in the feature space)
- Outputs of the classifier can be seen as probabilities
- If we simply sum the outputs of KFDA trained on two different view of the same data, the output will always be that of the most confident classifier

$$f(x_i) = \operatorname{sgn}\left(\sum_{j=1}^{k} \langle \mathbf{w}_j, \phi_j(x_i)\rangle + b\right)$$

Convex Formulation

- We follow the approach outlined by [Mika et al., 2001] and provide a disciplined convex form of Multiview Fisher Discriminant Analysis.
- Since we are trying to minimise the variance of the data along the projection whilst maximising the distance between the average outputs for each class over all of the views.
- This can be done it two ways, which we have denoted as the block method and concatenation method.

Block Method

$$\begin{aligned} \min_{\boldsymbol{\alpha}, b, \boldsymbol{\xi}} \quad & H(\boldsymbol{\xi}) + \lambda P(\boldsymbol{\alpha}, \boldsymbol{\beta}) & \\ \text{s.t.} \quad & \mathbf{K}_j \boldsymbol{\alpha}_j + \mathbf{1} b = \mathbf{y} + \boldsymbol{\xi} & j = 1, \ldots, k \\ & \mathbf{1}_i' \boldsymbol{\xi} = 0 & i = 1, 2 \end{aligned}$$

where,

$H(\cdot)$ is the Loss function
$P(\cdot)$ is the Regularisation function

Concatenation Method

$$\min_{\boldsymbol{\alpha}_j, b, \boldsymbol{\xi}} \quad H(\boldsymbol{\xi}) + \lambda P(\boldsymbol{\alpha}_j), \qquad\qquad j = 1, \ldots, k$$

$$\text{s.t.} \quad \sum_{j=1}^{k} \mathbf{K}_j \boldsymbol{\alpha}_j + \mathbf{1} b = \mathbf{y} + \boldsymbol{\xi}$$

- Concatenation method more efficient (scales linearly)
- Block method closer to spirit of KCCA

Regularisation

- The natural choices for the regularisation function $P(\boldsymbol{\alpha}, \boldsymbol{\beta})$ would either be the L_2-norm of the dual weight vectors

$$P(\boldsymbol{\alpha}) = \sum_{j=1}^{k} \|\boldsymbol{\alpha}_j\|_2^2$$

- or the L_2-norm of the primal weight vector

$$P(\boldsymbol{\alpha}) = \sum_{j=1}^{k} \boldsymbol{\alpha}_j' \mathbf{K}_i \boldsymbol{\alpha}_j$$

- More interesting is the L_1-norm of the dual weight vector, as this choice leads to sparse solutions

$$P(\boldsymbol{\alpha}) = \sum_{j=1}^{k} \|\boldsymbol{\alpha}_j\|_1$$

Loss Functions

- Standard loss function assumes Gaussian Densities:
  - $H(\xi) = \|\xi\|_2^2 \quad \Longrightarrow \quad \frac{1}{\sqrt{2\pi}} \exp(-\frac{\xi^2}{2})$
- Can remove Gaussian noise model, resulting in different loss functions, e.g. Laplacian noise model:
  - $H(\xi) = \|\xi\|_1 \quad \Longrightarrow \quad \frac{1}{2} \exp(-|\xi|)$
- Using L_1 regulariser with Laplacian Loss results in a linear programme
- Other loss functions: ϵ-insensitive loss, Huber robust loss etc.

EEG & Music Dataset

- EEG experiment analysed in [Diethe et al., 2008]. The principal hypothesis was that neural patterns should reflect relative changes in the key of music that a listener is attending to.

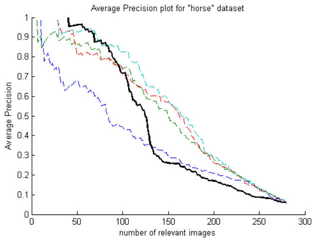
Table: Test errors for within-subject classification for Tonal vs Atonal. The SVM classification results are on the EEG data alone. The KCCA + SVM, and SMFDA classification used a kernel on the music as a second view. ** denotes significance at the $p < 0.001$ level

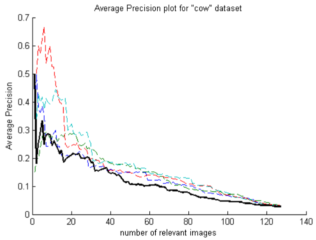
Classifier	# Train	# Test	Linear
SVM (Linear)	1152	383	0.2298**
SVM (RBF)	1152	383	0.1175**
KCCA + SVM (linear)	1152	383	0.0157**
SMFDA	1152	393	0.0000**

VOC 2007 Dataset

- Dataset from the PASCAL VOC2007 challenge [1]
- Features from [Viitaniemi and Laaksonen, 2008], with an extra feature extraction method (SIFT)

Figure: Average precision recall curves for 3 VOC 2007 datasets for Multiview Fisher Discriminant Analysis using the Block method plotted against SOM results





Selected References

Diethe, T., Durrant, S., Shawe-Taylor, J., and Neubauer, H. (2008).
Semantic dimensionality reduction for the classification of EEG according to musical tonality.
Technical report, Presented at the NIPS 2008 workshop Learning from Multiple Sources.

Mika, S., Smola, A. J., and Schölkopf, B. (2001).
An improved training algorithm for kernel fisher discriminants.
In *In AISTATS*, pages 98–104.

Viitaniemi, V. and Laaksonen, J. (2008).
Techniques for image classification, object detection and object segmentation applied to VOC challenge 2007.
Technical report, Department of Information and Computer Science, Helsinki University of Technology (TKK).