Linear Ranking Analysis

Weihong Deng, Jiani Hu, Jun Guo

Beijing University of Posts and Telecommunication.

E-mail: {whdeng, jnhu, guojun}@bupt.edu.cn.

Introduction

Two limitations of Linear Discriminant Analysis (LDA)

$$J(w) = \frac{w^T S_B w}{w^T S_W w}$$

$$S_b = \frac{1}{n}\sum_{i=1}^{C}\sum_{j=1}^{n_i} n_i (x_{ij} - \mu_i)(x_{ij} - \mu_i)^T$$

- Class separation problem : when the dimension is less than C−1, LDA is suboptimal and could merge the close classes.
- LDA does not model the relative strength of the high-level semantic attributes, which is effective to enhance object recognition and zero-shot learning

Contributions:

- We extend the LDA technique to linear ranking analysis (LRA), by considering the ranking order of classes centroids on the projected subspace.
- Under the constrain on the ranking order of the classes, two criteria are proposed:
  1) minimization of the classification error with the assumption that each class is homogenous Guassian distributed;
  2) maximization of the sum (average) of the k minimum distances of all neighboring-class (centroid) pairs.
- Demonstration of state-of-the-art performance on ranking learning, zeros-shot learning, and subspace selection.

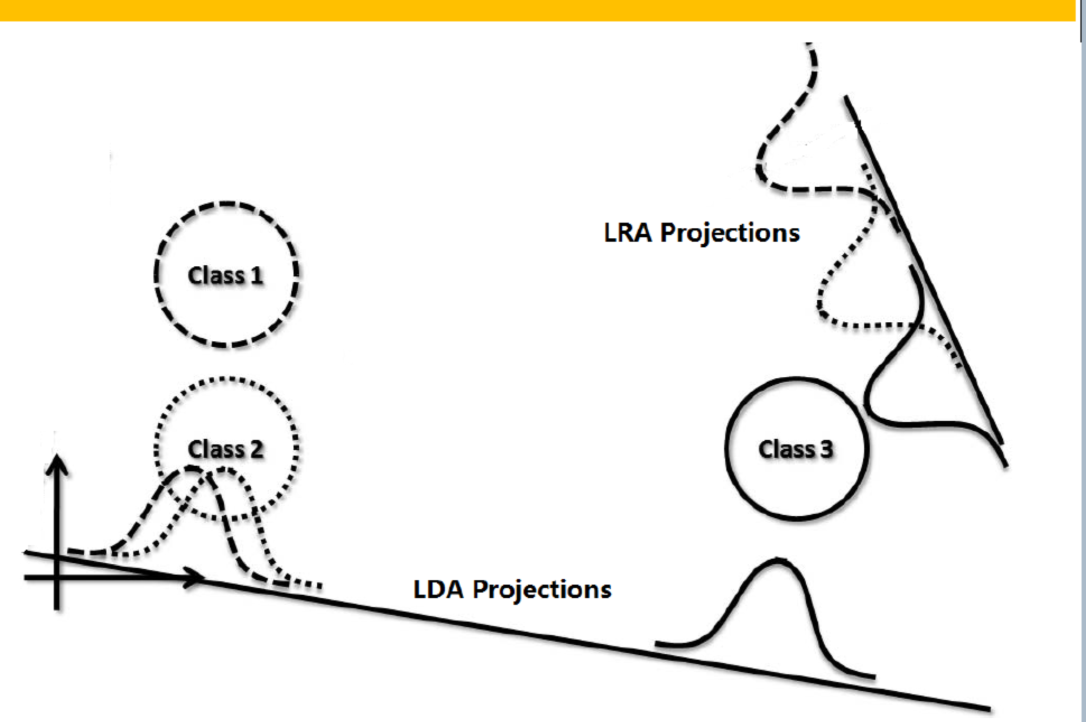


Figure 1. There are three classes (named 1, 2, and 3) of samples, which are drawn from a Gaussian distribution in each class. LDA finds a projection direction that maximize the between-class scatter, while merges class 1 and class 2. The proposed LRA aims to minimize the classification error (or maximize neighboring-class margin) while preserving the ranking order "Class 1≻Class 2≻Class 3"

Method

Learn a ranking function under the constraint of the order of projected class centroid.

$$r(x) = w^T x$$

Minimum Error Criterion

- Assume that the class conditional distribution is homoscedastic Gaussian, minimizes the Bayes optimal classification error as follows

$$J(w) = \frac{2}{C}\sum_{i=1}^{C-1}\Phi\left(\frac{-w^T\mu_{i,i+1}}{2}\right)$$

- The Jacobian and Hessian is well-defined

$$\frac{\partial J}{\partial w} = -\frac{1}{C}\sum_{i=1}^{C-1}\frac{1}{\sqrt{2\pi}}e^{(w^T\mu_{i,i+1})^2/4}\mu_{i,i+1}$$

$$\frac{\partial^2 J}{\partial^2 w} = \frac{1}{4C\sqrt{2\pi}}\sum_{i=1}^{C-1}e^{(w^T\mu_{i,i+1})^2/4}(w^T\mu_{i,i+1})\mu_{i,i+1}\mu_{i,i+1}^T$$

- Mathematically elegant but limit on non-Guassian-distributed classes

Large Margin Criterion

- Maximize the average distance of the K nearest neighboring classes in the projected space

$$\max_w \Theta_k(y)$$

- K=1, maximize the distance between the nearest neighboring projected classes.
- K=C-1, maximize the extension of the projected classes.
- K={1,3,…,C-1}, adaptive to the various class distributions (selected by cross-validation).

- Objective function is convex and can be maximized efficiently by linear programming.

Theorem *Given the collection of linear function, $\{g_i(w) = w^T(\mu_{i+1} - \mu_i)\}_{i=1}^{C-1}$, the problem of maximizing $\Theta_k(g(w))$, the sum of the k smallest function is convex, which can be formulated as a linear program as follows*

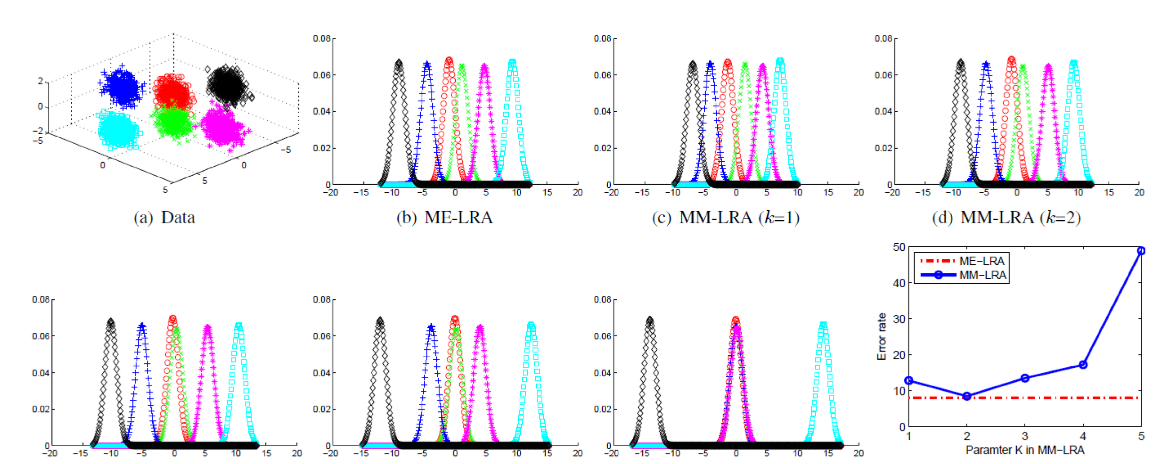
$$\max\left(kt - \textstyle\sum_{i=1}^{C-1}\xi_i\right)$$

subject to

$$\xi_i \geq t - g_i(w),\ i = 1,\ldots,C-1,$$
$$\xi_i \geq 0,\ i = 1,\ldots,C-1,$$
$$w^T w \leq 1$$

Result

A. Gaussian Toy Data



(a) Data (b) ME-LRA (c) MM-LRA (k=1) (d) MM-LRA (k=2)
(e) MM-LRA (k=3) (f) MM-LRA (k=4) (g) MM-LRA (k=5) (h) Accuracy

Figure 2. The LRA learning of ranking function of six normal distributions with the predefined ordering (black ≺ blue ≺ red ≺ green ≺ magenta ≺ light-blue). (a) three-dimensional scatter plot of six classes. (b) Bayes optimal 1D result . (c–g) show the class distributions in the max-k-min optimal 1D subspace with $k = \{1, 2, 3, 4, 5\}$. (h) Dependence of error rate (using the nearest mean classification) on the parameter k of the MM-LRA using the homogeneous Gaussian data set.

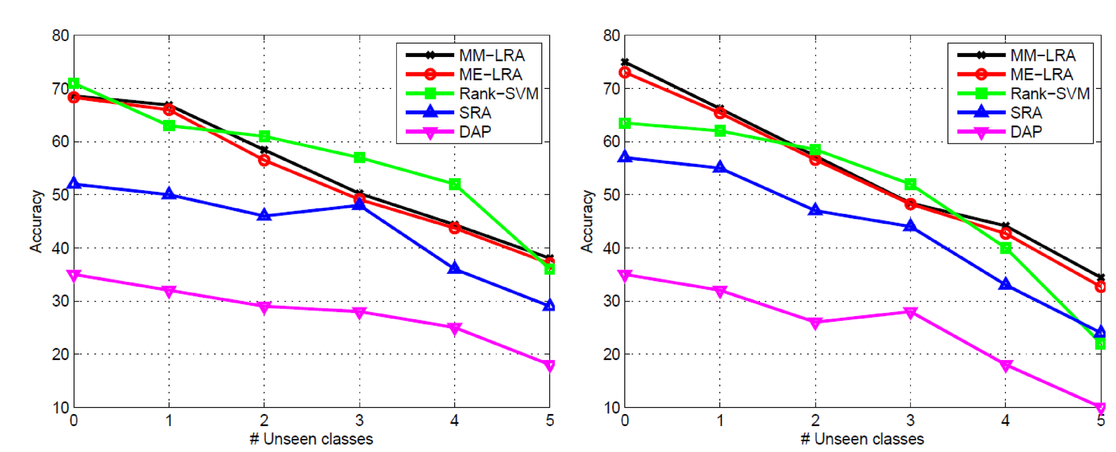
B. Attribute Ranking Learning

Table 1. The ranking function's accuracy on various attributes or Outdoor Scene Recognition and Public Figure Face Databases

Attributes	LDA	Binary SVM	Rank SVM	ME-LRA	MM-LRA
natural	92.25	90.74	**94.36**	94.03	94.04
open	68.48	84.54	**90.97**	89.88	89.92
perspective	78.87	78.22	**85.78**	85.17	85.45
large-objects	74.57	69.85	**86.36**	85.37	85.87
diagonal-plane	81.29	81.82	**87.52**	86.89	87.23
close-depth	72.96	86.89	**88.70**	83.38	86.56
masculine-looking	75.83	70.05	81.00	82.17	**82.47**
white	60.58	64.37	77.31	77.46	**78.46**
Young	78.40	74.48	81.05	81.46	**82.02**
Smiling	71.67	68.97	79.66	79.73	**80.65**
Chubby	60.94	61.65	76.14	75.85	**77.24**
visible-forehand	79.25	75.20	**87.91**	86.74	87.42
bushy-eyebrows	67.79	69.28	78.89	80.08	**80.57**
narrow-eyes	57.35	74.80	**80.72**	79.08	80.04
pointy-nose	58.11	68.75	74.84	77.35	**78.64**
big-lips	64.11	73.88	78.07	79.56	**80.12**
round-face	73.44	72.69	80.46	81.79	**81.96**
Average	71.52	74.48	82.93	82.71	**83.45**

C. Zero-Shot Learning

Experiments conducted in "Relative Attributes" (ICCV2011) with different ranking functions

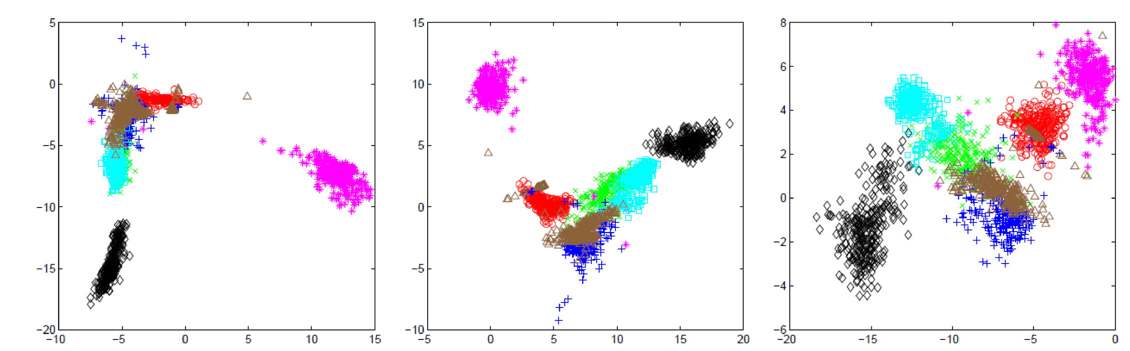


(a) OSR (b) PubFig

Figure 3. Zero-shot learning performance as the proportion of unseen classes increases. Total number of classes is constant at 8.

D. Subspace Selection for Classification

Exhaustive search the optimal cross-validation accuracy across all possible projected class orders.



(a) LDA (b) BLDA (c) LRA

Table 2. Average Classification Error Rate by Nearest Mean Classifier (upper table) and Nearest Neighbor Classifier (lower table) on Image Segmentation Data Set

Dim	1	2	3	4	5	6	Rank
LDA	49.39	31.21	15.71	11.04	9.74	8.53	5.60
GMSS	31.43	16.84	17.19	11.77	9.70	8.53	4.20
HMSS	32.47	21.21	15.54	10.95	9.91	8.53	4.40
aPAC	44.85	30.09	17.32	11.65	9.44	8.53	5.20
MMDA	49.05	19.70	14.68	12.90	8.87	8.53	4.20
BLDA	28.53	19.35	15.02	11.82	9.26	8.53	3.40
LRA	**27.14**	**15.71**	**12.21**	**9.52**	**8.74**	8.53	**1.00**

Dim	1	2	3	4	5	6	Rank
LDA	42.38	21.17	8.63	3.20	3.29	2.73	5.80
GMSS	31.22	14.16	7.84	3.85	3.90	2.73	5.20
HMSS	30.81	14.46	7.88	3.24	3.12	2.73	4.20
aPAC	41.90	21.82	7.10	3.20	3.16	2.73	5.00
MMDA	34.61	17.14	6.88	3.14	3.03	2.73	3.40
BLDA	**26.84**	13.77	5.93	3.38	3.16	2.73	3.20
LRA	27.97	**9.91**	**4.33**	**2.94**	**2.60**	2.73	**1.20**

Search the sub-optimal cross-validation accuracy under the constraint of the class order defined by a state-of-the-art dimension reduction method.

Table 4. Classification Error Rate by Nearest Neighbor Classifier on USPS Database

Dim	3	5	7	9	15	20	Rank
LDA	38.27	16.29	11.86	10.96	–	–	5.50
aPAC	31.44	16.79	11.41	11.06	–	–	5.25
HDA	35.18	25.16	19.88	17.34	–	–	11.00
WFLDA	43.10	23.42	14.15	10.96	–	–	8.75
FS-LDA	33.38	17.94	14.30	11.31	–	–	7.00
LFLDA	34.73	18.09	13.35	11.01	8.82	7.42	6.50
HFLDA	35.18	23.47	17.79	12.81	9.47	8.27	10.50
ODA	39.83	26.17	16.36	11.62	10.60	9.70	11.50
MODA	39.50	28.50	15.76	10.32	10.27	9.37	9.50
GMSS	**28.75**	15.74	11.16	9.87	5.98	5.83	2.25
ERE	39.01	20.23	15.60	11.36	6.73	5.33	9.25
ERE+BLDA	30.34	16.29	10.26	8.92	6.48	5.33	2.75
ERE+LRA	29.60	**14.85**	**10.16**	**8.87**	**6.43**	5.33	**1.25**

Conclusions

Consideration the ranking of the projected class centroids could achieve state-of-the-art performance on ranking learning, zeros-shot learning, and subspace selection.

Acknowledgments

- National Natural Science Foundation of China (NSFC)
- Program for New Century Excellent Talents in University.