2DLDA and Its Applications on the MNIST handwritten digits Classification Problems

Xixi Lu and Terry Situ
San Jose State University

About This Study

In this study, we are going to investigate how the algorithms of (2D) matrix-based linear discriminant analysis (LDA) perform on the classification problems of the MNIST handwritten digits dataset, and to compare its performance to the traditional (1D) vector-based dimension reduction method: Principal Compnent Analysis (PCA).

Linear Discriminant Analysis (LDA) is most commonly used as dimensionality reduction technique in the pre-processing step for pattern-classification and machine learning applications. The goal is to project a dataset onto a lower-dimensional space with good class-separability in order to avoid overfitting ("curse of dimensionality") and also reduce computational costs. The main ideas behind 2DLDA are that they are based on 2D matrices as opposed to the traditional LDA, which are based on 1D vector. [1]

With all the benefits come with 2DLDA, you may ask: "how does it performs with different classification methods (e.g. *k* nearest neighbors algorithm and LDA classification)?" or "is it better than 1D-PCA?"

We investigate these questions and analyze extensively them based on a large database of handwritten digits. The database contains 60,000 training images and 10,000 testing images. In addition, each image contains 28 x 28 pixels. [3]

The study is conducted in two parts:

1). We use 2DLDA to reduced the dimension of the **MNIST** handwritten digit dataset and applied different classification methods on the reduced dataset.

2). We compare the performance between 2DLDA and PCA (in term of test errors) by using different classifiers.

We compared the accuracy of those classifiers in term of test error rates.

How Does 2dlda Work?

2DLDA transforms $r \times c$ images to smaller $r' \times c'$ images.
Let $X \in R^{r \times c}$ be a given image. The transformation is defined by two matrices with orthonormal columns, $L \in R^{r \times r'}$ and $R \in R^{c \times c'}$:

$$Y = L^T X R \in R^{r' \times c'}$$

Like FDA, 2DLDA finds the best transformations *L*, *R* by preserving the most discriminatory information in the projection space:

$$\max_{L,R} \frac{between-class\ scatter}{within-class\ scatter}$$

And the between-class scatter and within-class scatter are defined as follow:

Within-class scatter:

$$\tilde{s}_w^2 = \sum_i \sum_{X \in \Pi_i} \| L^T (X - M_i) R \|_F^2 \qquad M_i = \frac{1}{n_i} \sum_{X \in \Pi_i} X$$

Between-class scatter:

$$\tilde{s}_b^2 = \sum_i n_i \| L^T (M_i - M) R \|_F^2 \qquad M = \frac{1}{n} \sum_i n_i M_i$$

Algorithm For 2DLDA

We need to obtain the optima *L* and *R* by solving the following equation:

$$\max_{L,R} \frac{\sum_i n_i \| L^T (M_i - M) R \|_F^2}{\sum_i \sum_{X \in \Pi_i} \| L^T (X - M_i) R \|_F^2}$$

Where $L \in R^{r \times r'}$ and $R \in R^{c \times c'}$ are tall matrices with orthonormal columns.

The iterative procedure:

1. Initialize $R = \begin{pmatrix} I_{c' \times c'} \\ 0_{(c-c') \times c'} \end{pmatrix} \in R^{c \times c'}$

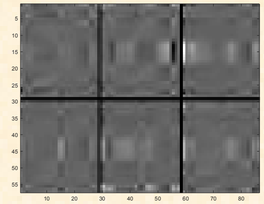
2. Iterative until convergence:
   - $L \leftarrow$ top r' eigenvectors of $(S_w^R)^{-1} S_b^R$
   - $R \leftarrow$ top c' eigenvectors of $(S_w^L)^{-1} S_b^L$

$$S_w^L = \sum_i \sum_{X \in \Pi_i} (X - M_i)^T L L^T (X - M_i) \in R^{c \times c} \qquad S_b^L = \sum_i n_i (M_i - M)^T L L^T (M_i - M) \in R^{c \times c}$$

$$S_w^R = \sum_i \sum_{X \in \Pi_i} (X - M_i) R R^T (X - M_i)^T \in R^{r \times r} \qquad S_b^R = \sum_i n_i (M_i - M) R R^T (M_i - M)^T \in R^{r \times r}$$

3. Return final versions of *L* and *R*.

Examples of Projected Displays by 2DLDA



The image on the left displays the first six images were projected here. The projected dimension is 15 x 15 for each image. Can our human recognize these images? (They are 5, 0, 4, 1, 9 & 2).

Investigations & Results: Case I 2DLDA + Local kMeans

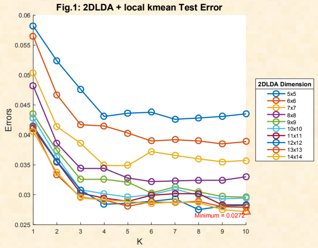


Figure 1: 2DLDA + Local kMeans. [4]
The lowest test error (0.0272) reached when the dimension is 14 x 14 and k = 10.

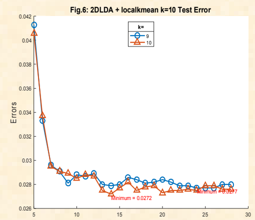


Figure 2: 2DLDA + Local kMeans (k=9 & 10).
The lowest test error (0.0272) reached when the projected dimension is 14 x 14 & k = 10.

Investigations & Results: Case II 2DLDA + kNN

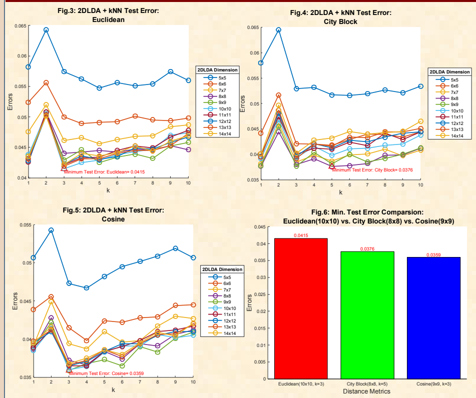


Figure 3 – 6: Different distant metrics with kNN. The lowest test error we obtained is 0.0359 with **Cosine Angle** distant function & the data size is 9 x 9.

Investigations & Results: Case III 2DLDA vs. PCA

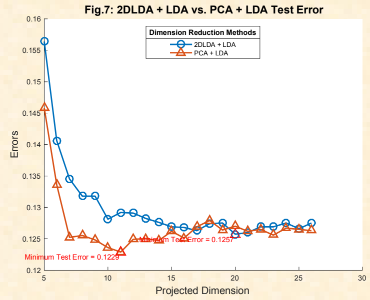


Figure 7: LDA: 2DLDA vs. PCA. The lowest test error(0.1229) obtained when using PCA with data size 11 x 11.

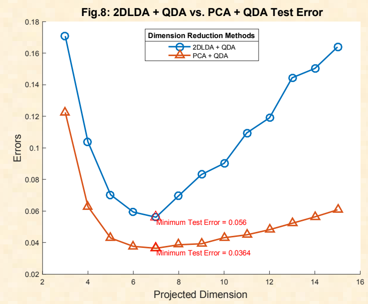


Figure 8: QDA: 2DLDA vs. PCA. The lowest test error(0.0364) obtained when using PCA with data size 7 x 7.

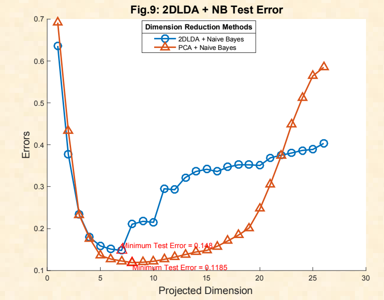


Figure 9: Naïve Bayes (Normal): 2DLDA vs. PCA. The lowest test error(0.1185) obtained when using PCA with data size 8 x 8.

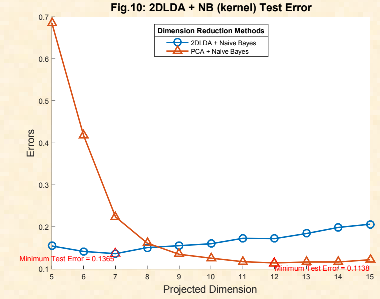


Figure 10: Naïve Bayes (kernel): 2DLA vs. PCA. The lowest test error(0.1138) obtained when using PCA with data size 12 x 12.

Conclusions & Limitations

- **For 2DLDA + kNN / Local Kmeans:**
  - the choice of distance functions will have a huge impact on the optimal combination of projected dimension and number of **k;**
  - This is possible for lower projected dimension to yield higher accuracy rate;
  - The higher the projected dimension will not always produce significant lower test error;
- **2DLDA vs. PCA:**
  - Both of LDA and QDA will produce lower min. test error rates PCA;
  - LDA & Naïve Bayes does not yield good classifying results;
- **Limitations:**
  - Due to the resource constrains, we DID NOT implement exhaustive search to obtain the global optimal combination(s) of projected dimension and classifier for either 2DLDA or PCA;
  - Even though we had obtained the global optimal combination here, that may ONLY apply to MNIST dataset.

References

[1]. Zheng, W., Lai, J., & Li, S. Z. (2008). 1D-LDA vs. 2D-LDA: When is vector-based linear discriminant analysis better than matrix-based? Pattern Recognition, 41(7), 2156-2172.
[2]. Raschka, S. (2014, August 3). Linear Discriminant Analysis – Bit by Bit. Retrieved March 30, 2016, from http://sebastianraschka.com/Articles/2014_python_lda.html
[3]. MNIST database. (n.d.). Retrieved April 03, 2016, from https://en.wikipedia.org/wiki/MNIST_database
[4]. Liang, Z (2008) 2DLDA PK LDA for feature extraction (Version 1.0) [Source code]. http://www.mathworks.com/matlabcentral/fileexchange/20174-2dlda-pk-lda-for-feature-extraction