Fast Classification of Handwritten Digits Using 2D-DCT Based Sparse PCA

Darya Ismailova and Wu-Sheng Lu

Department of Electrical and Computer Engineering
University of Victoria, Victoria, BC, Canada

Outline

Introduction

- The problem of handwritten digit recognition (HWDR) has broad applications, where both accuracy and speed of digit recognition are critical indicators of system performance.
- Problem itself: given a training data set $\{\mathcal{D}_j, j = 0, 1, \ldots, 9\}$ develop an approach to train a multi-class classifier to recogrnize the digit outsde the training data.
- The primary challenge of the HWDR problem: variation in the hadwriting styles.

PCA for Multi-Category Classification

- Given a data set $\{\boldsymbol{x}_i, i = 1, 2, \ldots, n\}, \boldsymbol{x}_i \in R^{mx1}$, its average vector and covariance matrix are defined as

$$\overline{\boldsymbol{x}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i, \quad \boldsymbol{C} = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^T$$

- Since $\boldsymbol{C} \succeq 0$ its singular value decomposition (SVD) is identical to its eigen-decomposition

$$\boldsymbol{C} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{U}^T$$

  where $\boldsymbol{U} = [\boldsymbol{u}_1\, \boldsymbol{u}_2 \ldots \boldsymbol{u}_m]$ is orthogonal, $\boldsymbol{S} = diag\{\sigma_1, \sigma_2 \ldots \sigma_m\}$ with $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_m$.

PCA for Multi-Category Classification

- An L_2-optimal rank-K approximation of covariance matrix $\boldsymbol{C}$ can be obtained as

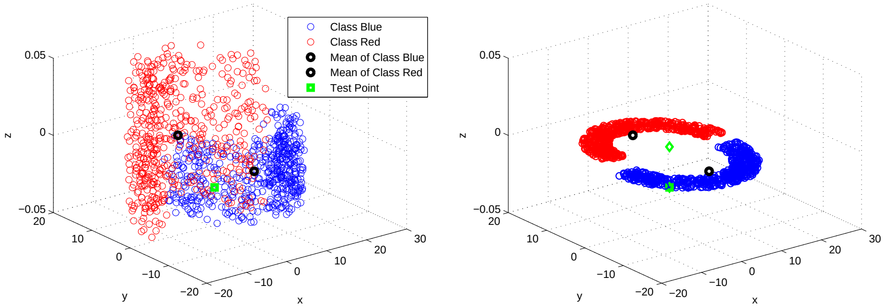
$$\boldsymbol{C} \approx \boldsymbol{U}_K \boldsymbol{S}_K \boldsymbol{U}_K^T$$

- The usefulness of approximation may be understood from two perspectives:

1. Dimention reduction from R^m to R^K
2. Supervised multi-category classification

PCA for Multi-Category Classification

Example of the supervised multi-category classification

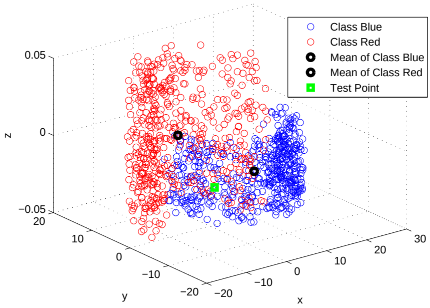


(a) Plot in the Original Space R^3

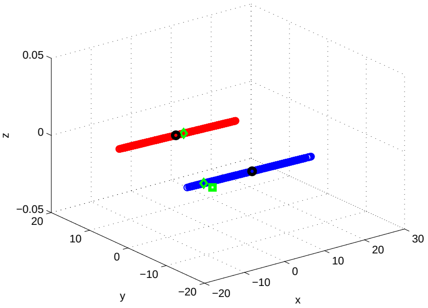
(b) Projection of Training Dataset to the 2-dimentional Subspace

Figure: Example "double semi-circle" data set for supervised classification of 1000 random samples.

Example of the supervised multi-category classification



(a) Plot in the Original Space R^3



(b) Projection of Training Dataset to the 1-dimentional Subspace

Figure: Example Data Set for Supervised Classification of 1000 Random Samples in Double Semi-Circle

2-D DCT

- Given a digital image of size N by N represented by its light intensity $\{x(i,j), i,j = 1,2,\ldots,N\}$, the 2-D DCT of the image is a 2-D array of the same size, denoted by $\{D(k,l), k,l = 1,2,\ldots,N\}$ where

$$D(k,l) = \frac{2\alpha(k)\alpha(l)}{N}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} x(i,j)\cdot\cos\left(\frac{(2i+1)k\pi}{2N}\right)\cos\left(\frac{(2j+1)l\pi}{2N}\right)$$

with

$$\alpha(k) = \begin{cases} \frac{1}{\sqrt{2}}, \text{ for } k = 0 \\ 1 \text{ for } k \neq 0 \end{cases}$$

2-D DCT

Important property of DCT is energy compaction.

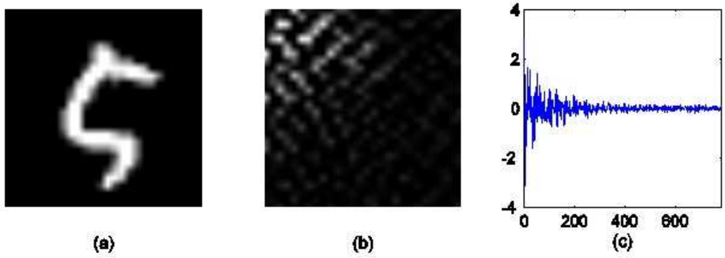


Figure: (a) An example digit from MNIST database, (b) the 2-D DCT of the image of size 28 by 28 in (a), and (iii) the 784 DCT coefficients as a 1-D sequence.

2-D DCT-Based Sparse PCA

- The main point is to use 2-D DCT at the pre-processing stage is to reduce the dimension m of the input data space.
- Given training data $\{\boldsymbol{x}_i, i = 1, 2, \ldots, n\}$ with $\boldsymbol{x}_i \in R^{mx1}$, first re-shape vector $\boldsymbol{x}_i$ to its original image size and apply 2-D DCT.
- Convert the 2-D DCT coefficients to a 1-D sequence by zig-zag scanning the coefficients.
- Retain the first r DCT coefficients to construct $\boldsymbol{d}_i \in R^r$ thus constructing a reduced data set $\{\boldsymbol{d}_i, i = 1, 2, \ldots, n\}$.

2-D DCT-Based Sparse PCA

- The equation for PCA holds for reduced data set as:

$$\boldsymbol{C} \approx \boldsymbol{U}_K \boldsymbol{S}_K \boldsymbol{U}_K^T$$

$$\overline{\boldsymbol{d}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{d}_i, \quad \boldsymbol{C} = \frac{1}{n-1}\sum_{i=1}^{n} (\boldsymbol{d}_i - \overline{\boldsymbol{d}})(\boldsymbol{d}_i - \overline{\boldsymbol{d}})^T$$

$$\boldsymbol{S}_K = diag\{\sigma_1, \sigma_2, \ldots, \sigma_K\}, \quad \boldsymbol{U}_K = [\boldsymbol{u}_1\, \boldsymbol{u}_2 \ldots \boldsymbol{u}_K]$$

- The L data classes $\{\boldsymbol{x}_i^{(j)}, i = 1, 2, \ldots, n_j\}$ for $j = 0, 1, \ldots, L-1$ are well represented by L reduced "data" sets $\{\overline{\boldsymbol{d}_j} \in R^r, \boldsymbol{U}_K^{(j)}\}$.

2-D DCT-Based Sparse PCA

- To classify a test point $\boldsymbol{x} \in R^m$:

  (i) Apply 2-D DCT to the image constructed from $\boldsymbol{x}$ and keep the first r DCT coefficients to construct vector $\boldsymbol{d} \in R^r$;

  (ii) Project point $\boldsymbol{d} - \overline{\boldsymbol{d}_j}$ into the jth data class: $\boldsymbol{z}_j = \boldsymbol{U}_K^{(j)T}(\boldsymbol{d} - \overline{\boldsymbol{d}_j})$;

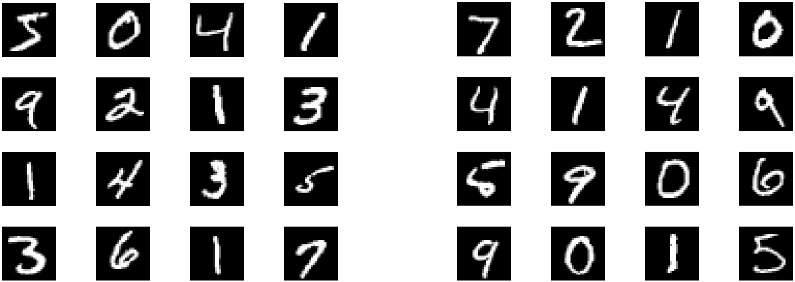
  (iii) Approximate point $\boldsymbol{d}$ in the jth class as $\widehat{\boldsymbol{d}_j} = \boldsymbol{U}_K^{(j)}\boldsymbol{z}_j + \overline{\boldsymbol{d}_j}$;

  (iv) Compute $e_j = \|\boldsymbol{d} - \widehat{\boldsymbol{d}_j}\|$ for $j = 0, 1, \ldots, L-1$;

  (v) Classify point $\boldsymbol{x}$ to class j^* if e_{j^*} reaches the minimum among $\{e_j, j = 0, 1, \ldots, L-1\}$.

Application to Handwritten Digit Recognition

The MNIST Database



(a) Training set (b) Testing set

Figure: Typical images from the MNIST database

Application to Handwritten Digit Recognition

Generation of Input Data for HWDR

- The MNIST database: 60,000 labeled handwritten digits in the training set, 10,000 handwritten digits in the test set. Each data sample is a vector of length 784 representing a 28 by 28 gray-scale image of the digit.

- Input data for the algorithm: ten sets of training data $\mathcal{D}_j = \{(\boldsymbol{x}_i^{(j)}, y_j), i = 1, 2, \ldots, n_j\}$ for $j = 0, 1, \ldots, 9$, where each $\mathcal{D}_j$ contains a total of n_j digits representing the same numeral j. In our experiments, n_j was set to 1200 for all sets and, for a fixed j, $\{\boldsymbol{x}_i^{(j)}, i = 1, 2, \ldots, 1200\}$ were selected at random from those in the training data that collects all the digits representing numeral j.

Application to Handwritten Digit Recognition

Performance Evaluation of 2-D DCT Based Sparse PCA

- Appropriate ranges for r and K were found to be $180 \leq r \leq 400$ and $22 \leq K \leq 31$, respectively.
- A smaller r yields a faster classifier, but using an r too small degrades recognition accuracy.

Application to Handwritten Digit Recognition

Performance Evaluation of 2-D DCT Based Sparse PCA

	Sparse PCA	**Conventional PCA**
Dim. of the classifier input	196	784
K	26	25
Accuracy	96.21%	96.26%
Normalized time	0.643	1

Conclusions

- A 2-D DCT-based sparse PCA classifier for handwritten digit recognition has been proposed.
- The ability of 2-D DCT to compress image-related signals allows a significant dimensionality reduction of the input space.
- The sparse PCA classifier is shown to perform HWDR considerably faster than the conventional PCA classifier without sacrificing recognition accuracy.

Q & A

Appendix

2-D DCT-Based Sparse PCA. The Algorithm

Input: Training data $\mathcal{D}_j = \{(\boldsymbol{x}_i^{(j)}, y_j), i = 1, 2, \ldots, n_j\}$ for $j = 0, 1, \ldots, L-1$; target dimension of reduced input space r; number K of principal components to be retained; and testing data $\mathcal{T}$.

Step 1: Apply 2-D DCT to $D_j, j = 0, 1, \ldots, L-1$ to obtain reduced data set $R_i = \{(\boldsymbol{d}_i^{(j)}, y_j), i = 1, 2, \ldots, n_j\}$ of dimension r.

Step 2: For $j = 0, 1, \ldots, L-1$ compute

$$\overline{\boldsymbol{d}}_j = \frac{1}{n}\sum_{i=1}^{n_j} \boldsymbol{d}_i^{(j)}, \boldsymbol{C}_j = \frac{1}{n_j - 1}\sum_{i=1}^{n_j}(\boldsymbol{d}_i^{(j)} - \overline{\boldsymbol{d}}_j)(\boldsymbol{d}_i^{(j)} - \overline{\boldsymbol{d}}_j)^T$$

and the K eigenvectors $\boldsymbol{U}_K^{(j)} = [\boldsymbol{u}_1^{(j)}\ \boldsymbol{u}_2^{(j)} \ldots \boldsymbol{u}_K^{(j)}]$ associated with the K largest eigenvalues of $\boldsymbol{C}_j$.

2-D DCT-Based Sparse PCA. The Algorithm

Step 3: For each vector $\boldsymbol{x}$ from test data $\mathcal{T}$:

(i) Compute vector $\boldsymbol{d} \in R^r$ by applying 2-D DCT to the image constructed from $\boldsymbol{x}$ and retaining the r most significant DCT coefficients (refer to Sec. 3.A);

(ii) Perform projections for $j = 0, 1, \ldots, L-1$;

(iii) Compute approximating points $\widehat{\boldsymbol{d}_j} = \boldsymbol{U}_K^{(j)} \boldsymbol{z}_j + \overline{\boldsymbol{d}_j}$ for $j = 0, 1, \ldots, L-1$;

(iv) Compute $e_j = \|\boldsymbol{d} - \widehat{\boldsymbol{d}_j}\|$ for $j = 0, 1, \ldots, L-1$;

(v) Classify point $\boldsymbol{x}$ to class j^* if e_{j^*} reaches the minimum among $\{e_j, j = 0, 1, \ldots, L-1\}$.

References I

- S. Haykin, *Neural Networks - A Comprehensive Foundation*, Macmillan College Publishing Co., 1994.
- Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural Computation*, vol. 1, pp. 541–551, 1989.
- Y. LeCun, L. D. Jackel, L. Bottou, A. Brunot, C. Cortes, J. S. Denker, H. Drucker, I. Guyon, U. A. Muller, E. Sackinger, P. Simard, and V. Vapnik, "Comparison of learning algorithms for handwritten digit recognition", *Int. Conf. on Artificial Neural Networks*, pp. 53–60, 1995.
- G. E. Hinton, P. Dayan, and M. Revow, "Modeling the manifolds of images of handwritten digits," *IEEE Trans. Neural Networks*, vol. 8, no. 1, pp. 65–74, Jan. 1997.

References II

- Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, pp. 2278–2324, Nov. 1998.
- F. Lauer, C. Y. Suen, and G. Bloch, "A trainable feature extractor for handwritten digit recongnition," *Pattern Recognition*, vol. 40, pp. 1816–1824, 2007.
- Y. LeCun, C. Cortes, and C. J. C. Burges, "The MNIST database of handwritten digits," available online: *http://yann.lecun.com/exdb/mnist*
- C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
- I. T. Jolliffe, *Principal Component Analysis*, 2nd ed., Springer, 2002.

References III

- M. Kirby and L. Sirovich, "Application of Karhunen-Loeve procedure for the characterization of human faces," *IEEE Trans. PAMI*, vol. 12, pp. 103–108, 1990.
- M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," *Proc. CVPR*, pp. 586-591, 1991.
- K. Rao and P. Yip, *Discrete Cosine Transform: Algorithms, Advantages, Applications,* Academic Press, 1992.
- W. B. Pennebaker and J. L. Mitchell, *JPEG Still Image Data Compression Standard*, 3rd ed., Springer, 1993.