

Stat 406 Spring 2008 Homework 2

1 Deriving the second principal component

1. Let

$$J(\mathbf{v}_2, \mathbf{z}_2) = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - z_{i1} \mathbf{v}_1 - z_{i2} \mathbf{v}_2)^T (\mathbf{x}_i - z_{i1} \mathbf{v}_1 - z_{i2} \mathbf{v}_2) \quad (1)$$

Show that $\frac{\partial J}{\partial \mathbf{z}_2} = 0$ yields $z_{i2} = \mathbf{v}_2^T \mathbf{x}_i$.

2. Show that the value of \mathbf{v}_2 that minimizes

$$\tilde{J}(\mathbf{v}_2) = -\mathbf{v}_2^T \mathbf{C} \mathbf{v}_2 + \lambda_2 (\mathbf{v}_2^T \mathbf{v}_2 - 1) + \lambda_{12} (\mathbf{v}_2^T \mathbf{v}_1 - 0) \quad (2)$$

is given by the eigenvector of \mathbf{C} with the second largest eigenvalue. Hint: recall that $\mathbf{C} \mathbf{v}_1 = \lambda_1 \mathbf{v}_1$ and

$$\frac{\partial \mathbf{x}^T \mathbf{A} \mathbf{x}}{\partial \mathbf{x}} = (\mathbf{A} + \mathbf{A}^T) \mathbf{x} \quad (3)$$

2 Deriving the residual error for PCA

1. Prove that

$$\|\mathbf{x}_i - \sum_{j=1}^K z_{ij} \mathbf{v}_j\|^2 = \mathbf{x}_i^T \mathbf{x}_i - \sum_{j=1}^K \mathbf{v}_j^T \mathbf{x}_i \mathbf{x}_i^T \mathbf{v}_j \quad (4)$$

Hint: first consider the case $K = 2$. Use the fact that $\mathbf{v}_j^T \mathbf{v}_j = 1$ and $\mathbf{v}_j^T \mathbf{v}_k = 0$ for $k \neq j$. Also, recall $z_{ij} = \mathbf{x}_i^T \mathbf{v}_j$.

2. Now show that

$$J_K \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^n \left(\mathbf{x}_i^T \mathbf{x}_i - \sum_{j=1}^K \mathbf{v}_j^T \mathbf{x}_i \mathbf{x}_i^T \mathbf{v}_j \right) = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i^T \mathbf{x}_i - \sum_{j=1}^K \lambda_j \quad (5)$$

Hint: recall $\mathbf{v}_j^T \mathbf{C} \mathbf{v}_j = \lambda_j \mathbf{v}_j^T \mathbf{v}_j = \lambda_j$.

3. If $K = d$ there is no truncation, so $J_d = 0$. Use this to show that the error from only using $K < d$ terms is given by

$$J_K = \sum_{j=K+1}^d \lambda_j \quad (6)$$

Hint: partition the sum $\sum_{j=1}^d \lambda_j$ into $\sum_{j=1}^K \lambda_j$ and $\sum_{j=K+1}^d \lambda_j$.



Figure 1: Left: test set (1 image per person for 40 people). Middle: the nearest neighbor in the training set using $K = 4$ PCA dimensions. A black rectangle in the top left indicates an error. Right: the nearest neighbor in the training set using $K = 25$ PCA dimensions. There are many fewer errors. Notice how the person in the bottom left of the test set is wearing glasses, but gets mapped to the same person without glasses in the training set. (Generating this Figure is Exercise 3.)

3 Derivation of Fisher's linear discriminant

Show that the maximum of

$$J(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}{\mathbf{w}^T \mathbf{S}_W \mathbf{w}} \quad (7)$$

is given by

$$\mathbf{S}_B \mathbf{w} = \lambda \mathbf{S}_W \mathbf{w} \quad (8)$$

$$(9)$$

where

$$\lambda = \frac{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}{\mathbf{w}^T \mathbf{S}_W \mathbf{w}} \quad (10)$$

Hint: recall that the derivative of a ratio of two scalars is given by

$$\frac{d}{dx} \frac{f(x)}{g(x)} = \frac{f'g - fg'}{g^2} \quad (11)$$

where $f' = \frac{d}{dx} f(x)$ and $g' = \frac{d}{dx} g(x)$. Also, recall that

$$\frac{d}{dx} \mathbf{x}^T \mathbf{A} \mathbf{x} = (\mathbf{A} + \mathbf{A}^T) \mathbf{x} \quad (12)$$

4 PCA and nearest neighbors for face recognition (Matlab)

Download the files `olivettifaces.mat` and `Code.zip` from the MLABA web page. Fill in the parts of `faceRecStub.m` marked with question marks, and thus reproduce Figure 1 and Figure 2. Turn in your code and plots.

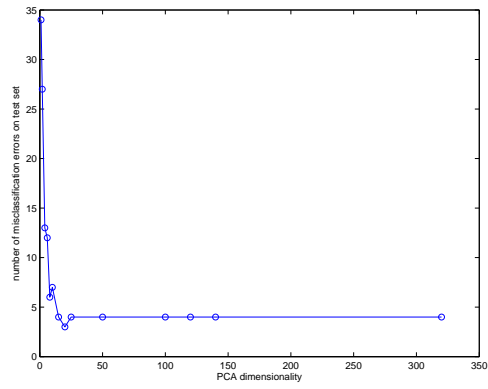


Figure 2: Misclassification rate vs number of PCA dimensions.