ECE 269 - Discussion Session

Fall 2020

1 Introduction

Suppose we have L data points $\mathbf{x}_1, \mathbf{x}_2. \dots \mathbf{x}_L \in \mathbb{R}^n$ and we form a matrix $\mathbf{X} \in \mathbb{R}^{n\times L}$ of the form

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_L]$$

Let us also assume that the rows of $\mathbf{X}$ are 0 mean i.e.

$$\sum_{i=1}^{L} \mathbf{X}_{ki} = 0 \quad 1 \le k \le n$$

Our aim is to find a new orthogonal coordinate system such that data has the greatest variance along the first direction, and the second greatest along the second and so on. So we must find a matrix $\mathbf{W}$ with orthonormal rows which transforms $\mathbf{X}$ to $\mathbf{Y}$ as

$$\mathbf{Y} = \mathbf{W}\mathbf{X}$$

where $\mathbf{W} \in \mathbb{R}^{m\times n}$ and $\mathbf{Y} \in \mathbb{R}^{m\times L}$ such that the sample variance of row 1 of $\mathbf{Y}$ is the greatest and that of row 2 of $\mathbf{Y}$ is the second largest and so on.

Let $\tilde{\mathbf{y}}_i \in \mathbb{R}^L$ and $\mathbf{w}_i \in \mathbb{R}^n$ denote the row vectors of $\mathbf{Y}$ and $\mathbf{W}$ respectively.

$$\mathbf{Y} = \begin{bmatrix} \tilde{\mathbf{y}}_1^T \\ \vdots \\ \tilde{\mathbf{y}}_m^T \end{bmatrix} \qquad \mathbf{W} = \begin{bmatrix} \mathbf{w}_1^T \\ \vdots \\ \mathbf{w}_m^T \end{bmatrix}$$

We will start with $\mathbf{w}_1$ and $\tilde{\mathbf{y}}_1$. We will find $\mathbf{w}_1$ through the following maximization problem

$$\mathbf{w}_1 = \underset{\|\mathbf{w}_1\|=1}{\arg\max} \quad \sum_{k=1}^{L} \tilde{y}_{1,k}^2 = \underset{\|\mathbf{w}_1\|=1}{\arg\max} \quad \|\tilde{\mathbf{y}}_1\|^2 = \underset{\|\mathbf{w}\|=1}{\arg\max} \quad \tilde{\mathbf{y}}_1^T \tilde{\mathbf{y}}_1 = \underset{\|\mathbf{w}_1\|=1}{\arg\max} \quad \mathbf{w}_1^T \mathbf{X}\mathbf{X}^{\mathbf{T}} \mathbf{w}_1$$

We know that the objective function in the preceding problem is maximized when

$$\mathbf{w}_1 = \text{ the eigenvector corresponding to the largest eigenvalue of } \mathbf{X}\mathbf{X}^{\mathbf{T}}.$$

Note that $\mathbf{X}\mathbf{X}^{\mathbf{T}}$ is the sample covariance matrix scaled by $L-1$. We will proceed similarly to sequentially find all the required vectors. To find $\mathbf{w}_k$, we need a vector orthogonal to the span of the first $k-1$ vectors($\mathbf{w}_1, \dots, \mathbf{w}_{k-1}$) which maximizes the sample variance. Let $\mathcal{U}_{k-1}$ be the subspace orthogonal to the first $k-1$ vectors. We can find $\mathbf{w}_k$ as

$$\mathbf{w}_k = \underset{\substack{\mathbf{w}_k \in \mathcal{U}_{k-1} \\ \|\mathbf{w}_k\|=1}}{\arg\max} \quad \sum_{l=1}^{L} \tilde{y}_{k,l}^2 = \underset{\substack{\mathbf{w}_k \in \mathcal{U}_{k-1} \\ \|\mathbf{w}_k\|=1}}{\arg\max} \quad \|\tilde{\mathbf{y}}_k\|^2 = \underset{\substack{\mathbf{w}_k \in \mathcal{U}_{k-1} \\ \|\mathbf{w}_k\|=1}}{\arg\max} \quad \tilde{\mathbf{y}}_1^T \tilde{\mathbf{y}}_1 = \underset{\substack{\mathbf{w}_k \in \mathcal{U}_{k-1} \\ \|\mathbf{w}_k\|=1}}{\arg\max} \quad \mathbf{w}_k^T \mathbf{X}\mathbf{X}^{\mathbf{T}} \mathbf{w}_k$$

We can similarly show that

$$\mathbf{w}_k = \text{ the eigenvector corresponding to the kth largest eigenvalue of } \mathbf{X}\mathbf{X}^{\mathbf{T}}$$

We can recover an estimate of $\mathbf{X}$ using $\mathbf{Y}$ as follows:

$$\hat{\mathbf{X}} = \mathbf{W}^T\mathbf{Y}$$

2 Face Recognition Using Principal Component Analysis

We are given L training images of size $N_1 \times N_2$. First we need to vectorize them to obtain a set of L vectors $\mathbf{\Gamma}_1, \ldots, \mathbf{\Gamma}_L \in \mathbb{R}^{N_1 N_2}$. Before constructing the covariance matrix, we need to make sure that the sample mean of each attribute is 0. To do so, first we calculate the average

$$\mathbf{\Psi} = \frac{1}{L}\sum_{i=1}^{L}\mathbf{\Gamma}_i$$

Then we can construct the (unnormalized) covariance matrix as

$$\mathbf{C} = \mathbf{A}\mathbf{A}^{\mathbf{T}}$$

where $\mathbf{A} = [\mathbf{\Gamma_1} - \mathbf{\Psi}, \ldots, \mathbf{\Gamma_L} - \mathbf{\Psi}] \in \mathbb{R}^{N_1 N_2 \times L}$. Since $\mathbf{C}$ is a symmetric matrix, it has the following eigen decomposition

$$\mathbf{C} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$$

Note that you cannot directly compute eigen decomposition of $\mathbf{A}\mathbf{A}^{\mathbf{T}}$ due to its large size. You need to find the eigenvectors (and eigenvalues) of $\mathbf{A}\mathbf{A}^{\mathbf{T}}$ using the eigen decomposition of $\mathbf{A}^{\mathbf{T}}\mathbf{A}$ as described in the reference paper.

The matrix $\mathbf{C}$ will have at most L non-zero eigenvalues. Denote the orthonormal eigen vectors corresponding to these eigenvalues as $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_L$. These are called eigenfaces.

Here is how we reconstruct a given vectorized image $\mathbf{\Gamma}$ using eigenfaces. First you need to determine M eigenfaces corresponding to the M largest eigenvalues of $\mathbf{C}$. The choice of M is up to you, however you need to justify your choice.

$$\hat{\mathbf{\Gamma}} = \sum_{i=1}^{M}\alpha_i\mathbf{u}_i + \mathbf{\Psi}$$

where the coefficients are $\alpha_i = \mathbf{u}_i^T(\mathbf{\Gamma} - \mathbf{\Psi})$