Question 1: Assignment 4: CS 663, Fall 2024

Amitesh Shekhar
IIT Bombay
22b0014@iitb.ac.in

Anupam Rawat
IIT Bombay
22b3982@iitb.ac.in

Toshan Achintya Golla
IIT Bombay
22b2234@iitb.ac.in

September 24, 2024

1. The aim of this exercise is to help you understand the mathematics behind PCA more deeply. Do as directed: [3+3+4+5=15 points]

   (a) Prove that the covariance matrix in PCA is symmetric and positive semi-definite.

   (b) Prove that the eigenvectors of a symmetric matrix are orthonormal.

   (c) Consider a dataset of some N vectors in d dimensions given by $\{\boldsymbol{x_i}\}_{i=1}^N$ with mean vector $\bar{\boldsymbol{x}}$. Note that each $\boldsymbol{x_i} \in \mathbb{R}^d$ and also $\bar{\boldsymbol{x}} \in \mathbb{R}^d$. Suppose that only k eigenvalues of the corresponding covariance matrix are large and the remaining are very small in value. Let $\tilde{\boldsymbol{x}}_{\boldsymbol{i}}$ be an approximation to $\boldsymbol{x_i}$ of the form $\tilde{\boldsymbol{x}}_{\boldsymbol{i}} = \bar{\boldsymbol{x}} + \sum_{l=1}^k \boldsymbol{V_l}\alpha_{il}$ where $\boldsymbol{V_l}$ stands for the lth eigenvector (it has d elements) and α_{il} (it is a scalar) stands for the lth eigencoefficient of $\boldsymbol{x_i}$. Argue why the error $\frac{1}{N}\sum_{i=1}^N ||\tilde{\boldsymbol{x}}_{\boldsymbol{i}} - \boldsymbol{x_i}||_2^2$ will be small. What will be the value of this error in terms of the eigenvalues of the covariance matrix?

   (d) Consider two uncorrelated zero-mean random variables (X_1, X_2). Let X_1 belong to a Gaussian distribution with variance 100 and X_2 belong to a Gaussian distribution with variance 1. What are the principal components of (X_1, X_2)? If the variance of X_1 and X_2 were equal, what are the principal components?

Soln:

(a) Covariane matrix C is given by

$$C = \frac{XX^T}{(N-1)}$$

And since $C^T = C$, we can conclude that the covariance matrix in PCA is symmetric.
A matrix M $\in$ L(V)is positive semi-definite iff,

   i. M is symmetric
   ii. $v^T Mv \geq 0 \ \forall \ v \in V$

Now,

$$v^T Cv = \frac{v^T XX^T v}{(N-1)} = \frac{||X^T v||^2}{(N-1)} \geq 0$$

We had already proved above that C is symmetric, thus we can say that the covariance matrix in PCA is symmetric and positive semi-definite.

(b) For any real matrix A and any vectors x and y, we have

$$\langle Ax, y\rangle = \langle x, A^T y\rangle.$$

Now assume that A is symmetric, and (x, λ) and (y, μ) are arbitrary (eigenvectors, eigenvalues) pairs of A, such that $\lambda \neq \mu$ and $x \neq y$. Then

$$\lambda\langle x, y\rangle = \langle \lambda x, y\rangle = \langle Ax, y\rangle = \langle x, A^T y\rangle = \langle x, Ay\rangle = \langle x, \mu y\rangle = \mu\langle x, y\rangle.$$

Therefore,

$$(\lambda - \mu)\langle x, y\rangle = 0.$$

Since $\lambda \neq \mu$, thus $\langle x, y\rangle = 0$, i.e., $x \perp y$. Also, since eigenvectors by default are normalised, we can say that the eigenvectors of a symmetric matrix is perpendicular and normalised, i.e. orthonormal.

(c) We know that we can write $\boldsymbol{x_i} = \bar{\boldsymbol{x}} + \sum_{l=1}^{d} \boldsymbol{V_l}\alpha_{il} = \bar{\boldsymbol{x}} + \boldsymbol{V_l}\boldsymbol{\alpha_i}$.
Also, we are given that $\tilde{\boldsymbol{x}}_{\boldsymbol{i}} = \bar{\boldsymbol{x}} + \sum_{l=1}^{k} \boldsymbol{V_l}\alpha_{il} = \bar{\boldsymbol{x}} + \boldsymbol{V_l}\tilde{\boldsymbol{\alpha}}_{\boldsymbol{i}}$.
We can easily note that $\tilde{\alpha}_{il} = \alpha_{il}$ for $1 \le l \le k$ and $\tilde{\alpha}_{il} = 0$ for $k+1 \le l \le d$.
Then the error that we defined above becomes

$$\frac{1}{N}\sum_{i=1}^{N} \|\tilde{\boldsymbol{x}}_{\boldsymbol{i}} - \boldsymbol{x_i}\|_2^2 = \frac{1}{N}\sum_{i=1}^{N} \|\tilde{\boldsymbol{\alpha}}_{\boldsymbol{i}} - \boldsymbol{\alpha_i}\|_2^2 = \frac{1}{N}\sum_{i=1}^{N}\sum_{l=k+1}^{d} \alpha_{il}^2$$

BY applying suitable approximations, we can say that

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{l=k+1}^{d} {\alpha_{il}}^2 \approx \sum_{l=k+1}^{d} E[\alpha_l^2] = \sum_{l=k+1}^{d} \lambda_l$$

Since we know that $\lambda_{k+1}, ...\lambda_d$ are all small, we can concluse that the error turns out to be small.

(d) We are given that

$$X_1 \sim \mathcal{N}(0, 100) \quad \text{(Gaussian distribution with variance 100)}$$
$$X_2 \sim \mathcal{N}(0, 1) \quad \text{(Gaussian distribution with variance 1)}$$

Since X_1 and X_2 are uncorrelated, their covariance matrix Σ is given by:

$$\Sigma = \begin{pmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) \\ \mathrm{Cov}(X_1, X_2) & \mathrm{Var}(X_2) \end{pmatrix} = \begin{pmatrix} 100 & 0 \\ 0 & 1 \end{pmatrix}$$

To calculate the eigenvalues λ of the covariance matrix Σ, we solve $\det(\Sigma - \lambda I) = 0$

$$\det\begin{pmatrix} 100-\lambda & 0 \\ 0 & 1-\lambda \end{pmatrix} = (100-\lambda)(1-\lambda) = 0$$

Thus, the eigenvalues are $\lambda_1 = 100, \quad \lambda_2 = 1$. Now, for each eigenvalue, we find the corresponding eigenvector.
- For $\lambda_1 = 100$:

$$\begin{pmatrix} 100-100 & 0 \\ 0 & 1-100 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & -99 \end{pmatrix}$$

The corresponding eigenvector is $\mathbf{v_1} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$.
- For $\lambda_2 = 1$:

$$\begin{pmatrix} 100-1 & 0 \\ 0 & 1-1 \end{pmatrix} = \begin{pmatrix} 99 & 0 \\ 0 & 0 \end{pmatrix}$$

The corresponding eigenvector is $\mathbf{v_2} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.
The principal components are the normalized eigenvectors corresponding to the largest eigenvalues:
1. First principal component PC_1 corresponding to $\lambda_1 = 100$: $PC_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$
2. Second principal component PC_2 corresponding to $\lambda_2 = 1$: $PC_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

If the variances of X_1 and X_2 were equal, we denote them as σ^2. The covariance matrix becomes:

$$\Sigma = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix} = \sigma^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

The eigenvalues are:

$$\lambda_1 = \sigma^2, \quad \lambda_2 = \sigma^2$$

Clearly, for the above eigenvalues, the eigenvectors can be any orthonormal basis of $\mathbb{R}^2$, such as $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$. So, the eigenvectors remain the same as earlier. In conclusion:
- For uncorrelated X_1 and X_2 with variances 100 and 1, respectively: $PC_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $PC_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$
- If the variances were equal, then the principal components could remain as $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$.