Question 4

CS663 (Digital Image Processing) Assignment 4

Atishay Jain (210050026)
Cheshta Damor (210050040)
Kanad Shende (210050078)

210050026@iitb.ac.in
210050040@iitb.ac.in
210050078@iitb.ac.in

Contents

I Question 4 1

1 Part 1 2

1.1 Symmetricity 2

1.2 Positive semi-definite 2

2 Part 2 3

3 Part 3 4

4 Part 4 6

Question 4

PART

I

Problem 1

The aim of this exercise is to help you understand the mathematics behind PCA more deeply. Do as directed: [5+5+5+5=20 points]

1. Prove that the covariance matrix in PCA is symmetric and positive semi-definite.

2. Prove that the eigenvectors of a symmetric matrix are orthonormal.

3. Consider a dataset of some N vectors in d dimensions given by $\{\boldsymbol{x_i}\}_{i=1}^{d}$ with mean vector $\bar{\boldsymbol{x}}$. 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 $\tilde{\boldsymbol{x}}_{\boldsymbol{i}} = \bar{\boldsymbol{x}} + \sum_{l=1}^{k} \boldsymbol{V_l} \alpha_{il}$. Argue why the error $\|\tilde{\boldsymbol{x}}_{\boldsymbol{i}} - \boldsymbol{x_i}\|_2^2$ will be small. What will be the value of this error in terms of the eigenvalues of the covariance matrix?

4. 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?

Section 1

Part 1

Subsection 1.1

Symmetricity

The covariance matrix can be written as follows -

$$\Sigma = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)(x_i - \mu)^T$$

- n is the number of data points.
- x_i represents the data points.
- μ is the mean vector.

By performing transpose operation on Σ,

$$\begin{aligned}\Sigma^T &= \left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)(x_i - \mu)^T\right)^T \\ &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)(x_i - \mu)^T \\ &= \Sigma\end{aligned}$$

Hence as Σ is equal to Σ^T, we can conclude that the covariance matrix in PCA is symmetric

Subsection 1.2

Positive semi-definite

A matrix M is positive semi-definite if, for any non-zero vector x, the following inequality holds:

$$x^T M x \geq 0$$

For our covariance matrix Σ in PCA,

$$\begin{aligned}x^T \Sigma x &= x^T \left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)(x_i - \mu)^T\right) x \\ &= \frac{1}{n}\sum_{i=1}^{n} x^T (x_i - \mu)(x_i - \mu)^T x \\ &= \frac{1}{n}\sum_{i=1}^{n} \left\|(x_i - \mu)^T x\right\|^2 \geq 0\end{aligned}$$

This proves that the covariance matrix in PCA is positive semi-definite.

Section 2

Part 2

Assuming the matrix to have distinct eigenvalues.

Here, $\lambda_1 \neq \lambda_2$. Then: $$\begin{aligned} Av_1 &= \lambda_1 v_1 \\ Av_2 &= \lambda_2 v_2 \end{aligned}$$

Multiplying by v_2^T both sides of the first equation:

$$v_2^T A v_1 = \lambda_1 v_2^T v_1$$

Since A is symmetric, $A^T = A$, so:

$$\begin{aligned} v_2^T A v_1 &= v_2^T A^T v_1 \\ &= (Av_2)^T v_1 \\ &= (\lambda_2 v_2)^T v_1 \\ &= \lambda_2 v_2^T v_1 \end{aligned}$$

Equating the two expressions for $v_2^T A v_1$, we get:

$$\lambda_1 v_2^T v_1 = \lambda_2 v_2^T v_1$$

Since $\lambda_1 \neq \lambda_2$, this implies $v_2^T v_1 = 0$, which means v_1 and v_2 are orthogonal.

We can normalize v_1 and v_2 to unit vectors $\hat{v_1}$ and $\hat{v_2}$ by dividing them by their respective norms:

$$\hat{v_1} = \frac{v_1}{\|v_1\|}$$

$$\hat{v_2} = \frac{v_2}{\|v_2\|}$$

So, eigenvectors with distinct eigenvalues are orthonormal.

Section 3

Part 3

Each data-point can be projected onto an eigenspace, giving a vector of d eigen-coefficients for that point. Thus, by the PCA algorithm, we can write each x_i as,

$$\begin{aligned} x_i &= x_{\text{mean}} + \sum_{l=1}^{d} V_l \alpha_{il} \\ &= x_{\text{mean}} + \sum_{l=1}^{k} V_l \alpha_{il} + \sum_{l=k+1}^{d} V_l \alpha_{il} \\ &= \tilde{x} + \sum_{l=k+1}^{d} V_l \alpha_{il} \end{aligned}$$

Error expression over all data points:

$$\begin{aligned} E &= \frac{1}{N} \sum_{i=1}^{N} \left(\|x_i - \tilde{x}\|^2 \right) \\ &= \frac{1}{N} \sum_{i=1}^{N} \left(\|\tilde{x} + \sum_{l=k+1}^{d} V_l \alpha_{il} - \tilde{x}\|^2 \right) \\ &= \frac{1}{N} \sum_{i=1}^{N} \left(\| \sum_{l=k+1}^{d} V_l \alpha_{il} \|^2 \right) \\ &= \frac{1}{N} \sum_{i=1}^{N} \left(\sum_{l=k+1}^{d} \sum_{m=k+1}^{d} V_l^T V_m \alpha_{il} \alpha_{im} \right) \end{aligned}$$

Since $V_l^T V_m = I$ (the Kronecker delta), the above expression simplifies to:

$$E = \frac{1}{N} \sum_{i=1}^{N} \left(\sum_{l=k+1}^{d} \alpha_{il}^2 \right)$$

Each eigenvalue λ_l can be written in terms of its corresponding unit eigenvector $\tilde{v}_l$ as follows:

$$\lambda_l = \sum_{i=1}^{N} \tilde{v}_l \left(x_i - \tilde{x} \right) \left(x_i - \tilde{x} \right)^T \tilde{v}_l^T$$

λ_l is the sum of projections of all samples on the l^{th} eigenvector.

$$\lambda_l = (N-1)\, E\left(\alpha_{il}^2\right)$$

Hence our final error term can be evaluated as follows:

$$E = \frac{1}{N}\sum_{i=1}^{N}\left(\sum_{l=k+1}^{d}\frac{\lambda_l}{N-1}\right)$$

This error term is small because the contribution of the eigenvalues from k+1 to d is very low, leading to low eigenvalues, thus the corresponding eigenvector captures less of the total variance in the data.

SECTION 4

Part 4

The principal components of (X_1, X_2) are the eigenvectors of its covariance matrix. The random vector (X_1, X_2) has zero mean and variances $\sigma_1^2 = 100$ and $\sigma_2^2 = 1$. In this case, the covariance matrix C for (X_1, X_2) is:

$$C = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} = \begin{bmatrix} 100 & 0 \\ 0 & 1 \end{bmatrix}$$

By solving the characteristic equation $\det(C - \lambda I) = 0$, we get:

$$C = det\left(\begin{bmatrix} 100 - \lambda & 0 \\ 0 & 1 - \lambda \end{bmatrix}\right) = (100 - \lambda)(1 - \lambda) = 0$$

Hence, we get eigenvalues as $\lambda_1 = 100$ and $\lambda_2 = 1$. Now, evaluating its eigenvectors,

$$Cv_1 = \lambda_1 v_1$$

$$\therefore (C - \lambda_1 I)\, v_1 = 0$$

$$\therefore \begin{bmatrix} 99 & 0 \\ 0 & 0 \end{bmatrix} v_1 = 0$$

This equation is satisfied by any unit vector v_1 in the plane. Similar results follow for v_2. So principal components can be any two orthonormal vectors in the plane.

If the variances were equal $(\sigma_1^2 = \sigma_2^2)$, then the covariance matrix would be $C = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}$, and the principal components would again result in any two orthonormal vectors in the plane.