CS 663, Fall 2023

Assignment 4

Jahanvi Rajput 23D0378
Badri Vishal Kasuba 22M2119
Abhishek Kumar Singh 22M2104

Question 4:

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

   Solution:

   Let X be the centered data matrix, with rows representing data samples and columns representing features. If there are n data samples and d features, X is an $n \times d$ matrix.

   * **Centering the Data**:
     First, we calculate the mean vector, denoted as μ, by averaging each feature over all the data samples:

     $$\mu = \frac{1}{n} \sum_{i=1}^{n} x_i$$

     Here, x_i represents each row of the centered data matrix X.
     Then, we center the data by subtracting the mean vector μ from each data sample x_i:

     $$x_i = x_i - \mu$$

   * **Covariance Matrix**:
     The covariance matrix Σ is calculated as the matrix of covariances between the features. For two features i and j, the covariance is given by:

     $$\text{cov}(x_i, x_j) = \frac{1}{n} \sum_{i=1}^{n} x_i \cdot x_j$$

     The covariance matrix Σ is a $d \times d$ matrix, where each element Σ_{ij} represents the covariance between features i and j.
     In matrix notation, Σ can be expressed as:

$$\Sigma = \frac{1}{n} X^T X$$

where X^T represents the transpose of the centered data matrix X.

(i.) **Proving Symmetry**:

To prove that Σ is symmetric, we need to show that Σ is equal to its transpose Σ^T:

$$\Sigma = \frac{1}{n} X^T X$$

Now, let's calculate the transpose of Σ:

$$\Sigma^T = \left(\frac{1}{n} X^T X\right)^T = \frac{1}{n} (X^T X)^T$$

Using the properties of transposition, we have:

$$\Sigma^T = \frac{1}{n} X^T (X^T)^T$$

as X is symmetric implies $(X^T = X)$, which gives:

$$\Sigma^T = \frac{1}{n} X^T X$$

It's evident that Σ and Σ^T are equal:

$$\Sigma = \Sigma^T$$

Therefore, the covariance matrix Σ in PCA is symmetric, as Σ is equal to its transpose, Σ^T.

(ii.) **Proving positive semi-definite**:

To determine positive semi-definiteness, we need to examine the eigenvalues (λ_i) of Σ. The eigenvalues are found by solving the characteristic equation:

$$\det(\Sigma - \lambda I) = 0$$

The eigenvalues λ_i represent the variances of the principal components.
We need to show that all eigenvalues λ_i are non-negative:

$$\lambda_i \geq 0 \quad \text{for } i = 1, 2, \ldots, d$$

If all eigenvalues are non-negative, the covariance matrix Σ is positive semi-definite.
Consider the characteristic equation:

$$\det(\Sigma - \lambda I) = 0$$

Since Σ is a real symmetric matrix, its eigenvalues λ_i are real. If λ_i were non-positive (i.e., $\lambda_i \leq 0$) for any i, this would imply that there is a non-zero vector $\mathbf{v}$ such that:

$$(\Sigma - \lambda I)\mathbf{v} = \mathbf{0}$$

Multiplying both sides by $\mathbf{v}^T$:

$$\mathbf{v}^T(\Sigma - \lambda I)\mathbf{v} = \mathbf{v}^T\mathbf{0} = \mathbf{0}$$

We know that Σ is positive semi-definite because it is a covariance matrix. This implies that for any vector $\mathbf{v}$, $\mathbf{v}^T\Sigma\mathbf{v} \geq 0$. Now, using this property, we have:

$$\mathbf{v}^T(\Sigma - \lambda I)\mathbf{v} = \mathbf{v}^T\Sigma\mathbf{v} - \lambda\mathbf{v}^T\mathbf{v} \geq 0$$

But we assumed that $\lambda_i \leq 0$, which means that the term $-\lambda\mathbf{v}^T\mathbf{v}$ should be non-positive. However, this leads to a contradiction because it implies that $\mathbf{v}^T\Sigma\mathbf{v}$ is non-positive, which is not possible since Σ is a covariance matrix. Therefore, our assumption that $\lambda_i \leq 0$ for any i must be incorrect. Hence, we conclude that all eigenvalues λ_i of the covariance matrix Σ are positive, and the covariance matrix is positive definite.

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

Solution:

Let A be a real symmetric matrix. This means that A is a square matrix, and it satisfies the property $A = A^T$.

(a) **Eigenvectors are Orthogonal**:

Consider two distinct eigenvectors $\mathbf{v}_i$ and $\mathbf{v}_j$ corresponding to eigenvalues λ_i and λ_j, respectively, where $i \neq j$.

The eigenvector equation for a symmetric matrix is:

$$A\mathbf{v}_i = \lambda_i \mathbf{v}_i$$
$$A\mathbf{v}_j = \lambda_j \mathbf{v}_j$$

Now, take the dot product of these equations:

$$\mathbf{v}_i^T A\mathbf{v}_j = \lambda_i \mathbf{v}_i^T \mathbf{v}_j$$
$$\mathbf{v}_j^T A\mathbf{v}_i = \lambda_j \mathbf{v}_j^T \mathbf{v}_i$$

Since $A = A^T$ (symmetric matrix), the left-hand sides are the same:

$$\mathbf{v}_i^T A\mathbf{v}_j = \mathbf{v}_j^T A\mathbf{v}_i$$

This implies:

$$\lambda_i \mathbf{v}_i^T \mathbf{v}_j = \lambda_j \mathbf{v}_j^T \mathbf{v}_i$$

Subtracting the two equations:

$$(\lambda_i - \lambda_j)\mathbf{v}_i^T \mathbf{v}_j = 0$$

Since λ_i and λ_j are distinct eigenvalues ($\lambda_i \neq \lambda_j$), this implies that:

$$\mathbf{v}_i^T \mathbf{v}_j = 0$$

Thus, the eigenvectors $\mathbf{v}_i$ and $\mathbf{v}_j$ are orthogonal for $i \neq j$.

(b) **Eigenvectors are Normalized**:

To show that the eigenvectors are normalized, we need to prove that the magnitude of each eigenvector is 1. This is done by showing that $\|\mathbf{v}_i\| = 1$ for all i.

Consider an eigenvector $\mathbf{v}_i$ corresponding to the eigenvalue λ_i. We know that:

$$A\mathbf{v}_i = \lambda_i \mathbf{v}_i$$

Now, take the dot product of both sides with itself:

$$\mathbf{v}_i^T A\mathbf{v}_i = \mathbf{v}_i^T \lambda_i \mathbf{v}_i$$

Since $A = A^T$ (symmetric matrix), the left-hand side becomes:

$$\mathbf{v}_i^T A\mathbf{v}_i = \lambda_i \mathbf{v}_i^T \mathbf{v}_i$$

$$\mathbf{v}_i^T A\mathbf{v}_i = \lambda_i \|\mathbf{v}_i\|^2$$

Divide both sides by λ_i (assuming $\lambda_i \neq 0$):

$$\|\mathbf{v}_i\|^2 = \frac{\mathbf{v}_i^T A\mathbf{v}_i}{\lambda_i}$$

Now, we know that $A\mathbf{v}_i = \lambda_i \mathbf{v}_i$, so:

$$\mathbf{v}_i^T A\mathbf{v}_i = \mathbf{v}_i^T \lambda_i \mathbf{v}_i$$

$$\mathbf{v}_i^T A\mathbf{v}_i = \lambda_i \mathbf{v}_i^T \mathbf{v}_i$$

Substituting this into the equation:

$$\|\mathbf{v}_i\|^2 = \frac{\lambda_i \mathbf{v}_i^T \mathbf{v}_i}{\lambda_i}$$

$$\|\mathbf{v}_i\|^2 = \mathbf{v}_i^T \mathbf{v}_i$$

$$\|\mathbf{v}_i\|^2 = \|\mathbf{v}_i\|^2$$

Therefore, $\|\mathbf{v}_i\|^2$ is equal to itself, and it follows that:

$\|\mathbf{v}_i\|^2 = 1$

Hence, $\|\mathbf{v}_i\| = 1$, and the eigenvectors are normalized.

Combining the results of Step 1 (orthogonality) and Step 2 (normalization), we can conclude that the eigenvectors of a symmetric matrix are orthonormal.

3. **Consider a dataset of some N vectors in d dimensions given by $\{\boldsymbol{x_i}\}_{i=1}^N$. 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 $\frac{1}{N}\sum_{i=1}^N \|\tilde{\boldsymbol{x}}_{\boldsymbol{i}} - \boldsymbol{x_i}\|_2^2$ will be small. What will be the value of this error in terms of the eigenvalues of the covariance matrix?**

Solution:

We have a dataset of N vectors in d dimensions denoted as $\{x_i\}_{i=1}^N$ where i ranges from 1 to N, and the mean vector is denoted as $\bar{x}$. We are given that only k eigenvalues of the covariance matrix of this dataset are large, and the remaining eigenvalues are very small.

Now, let's find the value of the error $\|\tilde{x}_i - x_i\|_2^2$ in terms of the eigenvalues of the covariance matrix. Let $\lambda_1, \lambda_2, \ldots, \lambda_d$ be the eigenvalues of the covariance matrix in descending order, and $V_1, V_2, \ldots, V_d$ be the corresponding eigenvectors. We are given that only the first k eigenvalues $(\lambda_1, \lambda_2, \ldots, \lambda_k)$ are large, and the rest $(\lambda_{k+1}, \lambda_{k+2}, \ldots, \lambda_d)$ are very small.

The error $\|\tilde{x}_i - x_i\|_2^2$ can be expressed as:

$$\|\tilde{x}_i - x_i\|_2^2 = \|\bar{x} + \sum_{l=1}^{k}(V_l \cdot \alpha_{il}) - x_i\|_2^2$$

Using properties of norms, we have:

$$\|\bar{x} + \sum_{l=1}^{k}(V_l \cdot \alpha_{il}) - x_i\|_2^2 = \|\bar{x} - x_i\|_2^2 + \left\|\sum_{l=1}^{k}(V_l \cdot \alpha_{il})\right\|_2^2$$

The term $\|\bar{x} - x_i\|_2^2$ is independent of the eigenvectors and eigenvalues and represents the error introduced by the mean shift. It is typically not very large.

The term $\left\|\sum_{l=1}^{k}(V_l \cdot \alpha_{il})\right\|_2^2$ represents the approximation error, which depends on the eigenvectors and coefficients α_{il}. Now, the key observation is that the eigenvectors V_1 to V_k capture most of the variance in the data, as they correspond to the large eigenvalues. Therefore, $\left\|\sum_{l=1}^{k}(V_l \cdot \alpha_{il})\right\|_2^2$ is relatively small because we are focusing on the principal components that matter most.

Why the Error is Small:

(a) The eigenvectors of the covariance matrix represent the directions of maximum variance in the data. Large eigenvalues correspond to principal components with significant variance. Small eigenvalues correspond to directions of very little variance or noise.

(b) **Truncation of Small Eigenvalues**: By only considering the first k principal components, we are effectively ignoring the directions with small eigenvalues. This means that we are ignoring the noisy components in the data. In other words, we are focusing on the directions that carry most of the information.

Error in Terms of Eigenvalues of the Covariance Matrix

The approximation error can be expressed as:

$$\left\|\sum_{l=1}^{k}(V_l \cdot \alpha_{il})\right\|_2^2 = \sum_{l=1}^{k}\|V_l \cdot \alpha_{il}\|_2^2$$

Now, consider that $V_1, V_2, \ldots, V_k$ are the eigenvectors corresponding to the large eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$. These eigenvectors capture most of the variance in the data.

So, $\|V_l \cdot \alpha_{il}\|_2^2$ for $l = 1$ to k can be related to the corresponding eigenvalues as follows:

$$\|V_l \cdot \alpha_{il}\|_2^2 = \left(\alpha_{il}^2\right) \|V_l\|_2^2 = \alpha_{il}^2 \|V_l\|_2^2 = \alpha_{il}^2$$

Therefore, the approximation error can be expressed in terms of the eigenvalues as:

$$\left\| \sum_{l=1}^{k} (V_l \cdot \alpha_{il}) \right\|_2^2 = \sum_{l=1}^{k} \alpha_{il}^2$$

This approximation error is a sum of the squares of the coefficients α_{il} for the first k principal components and is not directly dependent on the eigenvalues themselves. The error depends on how well these coefficients capture the data's variance in the principal component space.

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?**

   Solution:

   To find the principal components of the random variables (X_1, X_2), we can use the eigenvalue decomposition of the covariance matrix Σ. The covariance matrix Σ for two random variables (X_1, X_2) can be calculated as follows:

$$\Sigma = \begin{bmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) \\ \text{Cov}(X_2, X_1) & \text{Var}(X_2) \end{bmatrix}$$

   Given that X_1 and X_2 are uncorrelated, $\text{Cov}(X_1, X_2) = 0$, and the covariance matrix Σ becomes:

$$\Sigma = \begin{bmatrix} \text{Var}(X_1) & 0 \\ 0 & \text{Var}(X_2) \end{bmatrix}$$

Now, let's find the eigenvalues and eigenvectors of this covariance matrix Σ.

Eigenvalues: The eigenvalues (λ) are the variances of the principal components. For a 2x2 matrix like Σ, the eigenvalues can be calculated by solving the characteristic equation:

$$\det(\Sigma - \lambda I) = 0$$

where I is the identity matrix:

$$\begin{vmatrix} \text{Var}(X_1) - \lambda & 0 \\ 0 & \text{Var}(X_2) - \lambda \end{vmatrix} = 0$$

Eigenvectors: The eigenvectors (V) are the directions of the principal components. They can be obtained by solving the equation:

$$(\Sigma - \lambda I) \cdot V = 0$$

where $\Sigma - \lambda I$ is the matrix with eigenvalues subtracted from the diagonal.

Case 1: $\text{Var}(X_1) = 100$ and $\text{Var}(X_2) = 1$

Given the covariance matrix Σ:

$$\Sigma = \begin{bmatrix} 100 & 0 \\ 0 & 1 \end{bmatrix}$$

For λ_1 (eigenvalue corresponding to the variance of the first principal component):

$$(\Sigma - \lambda_1 I) \cdot V_1 = 0$$

Solving this equation will give us λ_1 and V_1:

$$(100 - \lambda_1) \cdot V_{1_1} = 0 \Rightarrow \lambda_1 = 100$$
$$(1 - \lambda_1) \cdot V_{1_2} = 0 \Rightarrow \lambda_1 = 1$$

For $\lambda_1 = 100$, the corresponding eigenvector V_1 will be $(1, 0)$.

For $\lambda_1 = 1$, the corresponding eigenvector V_1 will be $(0, 1)$.

For λ_2 (eigenvalue corresponding to the variance of the second principal component):

$$(\Sigma - \lambda_2 I) \cdot V_2 = 0$$

Solving this equation will give us λ_2 and V_2:

$$(100 - \lambda_2) \cdot V_{2_1} = 0 \Rightarrow \lambda_2 = 100$$
$$(1 - \lambda_2) \cdot V_{2_2} = 0 \Rightarrow \lambda_2 = 1$$

For $\lambda_2 = 100$, the corresponding eigenvector V_2 will be $(1, 0)$.

For $\lambda_2 = 1$, the corresponding eigenvector V_2 will be $(0, 1)$.

Principal Component 1:

Eigenvalue: $\lambda_1 = 100$ Eigenvector: $V_1 = (1, 0)$

Principal Component 2:

Eigenvalue: $\lambda_2 = 1$ Eigenvector: $V_2 = (0, 1)$

Case 2: Equal Variances, $\mathbf{Var}(X_1) = \mathbf{Var}(X_2)$

In this case, the covariance matrix Σ is the same as in Case 1. Therefore, the principal components are the same as in Case 1.

Principal Component 1 ($\lambda_1 = 100$):

Eigenvalue: $\lambda_1 = 100$ Eigenvector: $V_1 = (1, 0$

Principal Component 2 ($\lambda_2 = 1$):

Eigenvalue: $\lambda_2 = 1$ Eigenvector: $V_2 = (0, 1$

The principal components are the same in both cases, but the eigenvalues represent the variances of the corresponding components, which differ between Case 1 and Case 2.