Question 1

Anshika Raman
Roll No: 210050014

Kushal Agarwal
Roll No: 210100087

Kavan Vavadiya
Roll No: 210100166

October 23, 2024

a) Let the data matrix be $X \in R^{n \times p}$, where n is the number of samples and p is the number of features. The covariance matrix $C \in R^{p \times p}$ is given by:

$$C = \frac{1}{n-1}(X - \bar{X})^\top (X - \bar{X})$$

where $\bar{X}$ is the mean of the data matrix X.
Symmetry:
To show that C is symmetric, we need to prove that $C = C^\top$.

$$C^\top = \left(\frac{1}{n-1}(X - \bar{X})^\top (X - \bar{X})\right)^\top = \frac{1}{n-1}(X - \bar{X})^\top (X - \bar{X})$$

Since the transpose of a product of matrices reverses the order of multiplication and $(A^\top)^\top = A$, it follows that:

$$C^\top = C$$

Thus, the covariance matrix is symmetric.
Positive Semi-Definiteness:
To prove that C is positive semi-definite, we need to show that for any vector $z \in R^p$,

$$z^\top C z \geq 0$$

Substituting $C = \frac{1}{n-1}(X - \bar{X})^\top (X - \bar{X})$, we have:

$$z^\top C z = \frac{1}{n-1} z^\top (X - \bar{X})^\top (X - \bar{X}) z$$

Let $y = (X - \bar{X})z$. Then,

$$z^\top C z = \frac{1}{n-1} y^\top y = \frac{1}{n-1}\|y\|^2$$

Since $\|y\|^2 \geq 0$ for all y, it follows that:

$$z^\top C z \geq 0$$

Thus, the covariance matrix C is positive semi-definite.

b) Let $A \in R^{n \times n}$ be a symmetric matrix, i.e., $A^\top = A$. Assume $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are eigenvectors of A corresponding to eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, respectively. This means:

$$A\mathbf{v}_i = \lambda_i \mathbf{v}_i \quad \text{for } i = 1, 2, \ldots, n$$

We want to show that eigenvectors corresponding to distinct eigenvalues are orthogonal, i.e., for $i \neq j$, we have:

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

Since A is symmetric, we have:

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

On the other hand, taking the transpose of $A\mathbf{v}_i = \lambda_i \mathbf{v}_i$, we get:

$$(A\mathbf{v}_i)^\top = (\lambda_i \mathbf{v}_i)^\top = \lambda_i \mathbf{v}_i^\top$$

Thus,

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

Since A is symmetric, we also know that:

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

Therefore, we have:

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

Now, since dot product is symmetric, this simplifies to:

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

For $i \neq j$, since $\lambda_i \neq \lambda_j$, it follows that:

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

Thus, the eigenvectors corresponding to distinct eigenvalues are orthogonal.
Orthonormality:
To make the eigenvectors orthonormal, we can normalize each eigenvector. Let:

$$\mathbf{u}_i = \frac{\mathbf{v}_i}{\|\mathbf{v}_i\|}$$

Then the set of eigenvectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ will form an orthonormal basis. Therefore, the eigenvectors of a symmetric matrix can be chosen to be orthonormal.

c) The dataset $\{\mathbf{x}_i\}_{i=1}^N$ is approximated by $\tilde{\mathbf{x}}_i = \bar{\mathbf{x}} + \sum_{l=1}^k V_l \alpha_{il}$. This implies that $\tilde{\mathbf{x}}_i$ is a projection of $\mathbf{x}_i$ onto a k-dimensional subspace spanned by the top k eigenvectors $\{V_l\}_{l=1}^k$ of the covariance matrix.

Since only k eigenvalues are large, the remaining $d-k$ eigenvalues are very small. These small eigenvalues correspond to directions in which the data has very low variance. When we approximate $\mathbf{x}_i$ using only the top k eigenvectors, the error is mostly due to neglecting the contributions from the remaining $d-k$ small-eigenvalue directions.

The error per data point is given by:

$$\|\tilde{\mathbf{x}}_i - \mathbf{x}_i\|_2^2 = \left\| \sum_{l=k+1}^{d} V_l \alpha_{il} \right\|_2^2.$$

Since the eigenvectors are orthogonal, this can be written as:

$$\|\tilde{\mathbf{x}}_i - \mathbf{x}_i\|_2^2 = \sum_{l=k+1}^{d} \alpha_{il}^2.$$

The total mean squared error across all N data points is:

$$\frac{1}{N} \sum_{i=1}^{N} \|\tilde{\mathbf{x}}_i - \mathbf{x}_i\|_2^2 = \frac{1}{N} \sum_{i=1}^{N} \sum_{l=k+1}^{d} \alpha_{il}^2.$$

Since α_{il}^2 is the projection of $\mathbf{x}_i$ in the direction of the l-th eigenvector, and the variance along this direction is given by the corresponding eigenvalue λ_l, we can express the error as:

$$\frac{1}{N} \sum_{i=1}^{N} \|\tilde{\mathbf{x}}_i - \mathbf{x}_i\|_2^2 = \sum_{l=k+1}^{d} \lambda_l.$$

Conclusion: Thus, the total mean squared error in approximating the data using only the top k eigenvectors is equal to the sum of the eigenvalues corresponding to the neglected directions:

$$\boxed{\sum_{l=k+1}^{d} \lambda_l}.$$

Since these eigenvalues are small, the approximation error will also be small. This justifies why the low-rank approximation works well when only a few eigenvalues are large.

d) Let us represent the random variables X_1 and X_2 as a random vector:

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}.$$

Since X_1 and X_2 are uncorrelated and both are zero-mean, the covariance matrix $\mathbf{C}$ of $\mathbf{X}$ is diagonal and can be written as:

$$\mathbf{C} = \begin{pmatrix} \mathrm{Var}(X_1) & 0 \\ 0 & \mathrm{Var}(X_2) \end{pmatrix} = \begin{pmatrix} 100 & 0 \\ 0 & 1 \end{pmatrix}.$$

The principal components are the eigenvectors of the covariance matrix $\mathbf{C}$, and their corresponding variances are the eigenvalues.
The eigenvalues of $\mathbf{C}$ are simply the diagonal elements: $\lambda_1 = 100$ and $\lambda_2 = 1$.
The corresponding eigenvectors are:

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Thus, the principal components of (X_1, X_2) are given by the eigenvectors $\mathbf{v}_1$ and $\mathbf{v}_2$, which correspond to the directions along the X_1 and X_2 axes.

Principal Components when Variances are Equal

Now, consider the case where the variances of X_1 and X_2 are equal, i.e., both random variables have the same variance. Let this common variance be σ^2. In this case, the covariance matrix becomes:

$$\mathbf{C} = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix} = \sigma^2 \mathbf{I},$$

where $\mathbf{I}$ is the identity matrix. The eigenvalues of this covariance matrix are both equal to σ^2, and the corresponding eigenvectors are still:

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

However, since both eigenvalues are equal, any linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$ is also an eigenvector. Thus, the principal components in this case are not unique and can lie in any direction in the plane.

Conclusion:

1. When the variances of X_1 and X_2 are unequal, the principal components are along the axes of X_1 and X_2.

2. When the variances are equal, the principal components can be any orthogonal directions in the plane, since the variance along any direction is the same.