Question 2

Anshika Raman
Roll No: 210050014

Kushal Agarwal
Roll No: 210100087

Kavan Vavadiya
Roll No: 210100166

October 23, 2024

Minimized Direction e

We want to minimize the quantity:

$$\sum_{i=1}^{N} \|\mathbf{x}_i - \bar{\mathbf{x}} - (\mathbf{e} \cdot (\mathbf{x}_i - \bar{\mathbf{x}}))\mathbf{e}\|^2,$$

which represents the sum of squared distances between each vector $\mathbf{x}_i$ and its projection along the direction $\mathbf{e}$. This projection removes the variation along $\mathbf{e}$, leaving only the residual error orthogonal to $\mathbf{e}$.

Let $\mathbf{y}_i = \mathbf{x}_i - \bar{\mathbf{x}}$ represent the centered data. The goal is to minimize the residual error:

$$\|\mathbf{y}_i - (\mathbf{e}^\top \mathbf{y}_i)\mathbf{e}\|^2 = \mathbf{y}_i^\top \mathbf{y}_i - (\mathbf{e}^\top \mathbf{y}_i)^2$$

Summing over all vectors:

$$\sum_{i=1}^{N} \|\mathbf{y}_i - (\mathbf{e}^\top \mathbf{y}_i)\mathbf{e}\|^2 = \sum_{i=1}^{N} \mathbf{y}_i^\top \mathbf{y}_i - \sum_{i=1}^{N} (\mathbf{e}^\top \mathbf{y}_i)^2$$

Thus, the minimization of the residual error is equivalent to maximizing the second term:

$$\sum_{i=1}^{N} (\mathbf{e}^\top \mathbf{y}_i)^2 = \mathbf{e}^\top \left(\sum_{i=1}^{N} \mathbf{y}_i \mathbf{y}_i^\top \right) \mathbf{e} = \mathbf{e}^\top \mathbf{C} \mathbf{e}$$

Hence, we want to maximize $\mathbf{e}^\top \mathbf{C} \mathbf{e}$, where $\mathbf{C}$ is the covariance matrix of the dataset. The vector $\mathbf{e}$ that maximizes this expression is the eigenvector of $\mathbf{C}$ corresponding to its largest eigenvalue λ_1.

Direction f Perpendicular to e

Next, we want to find a direction $\mathbf{f}$ that is perpendicular to $\mathbf{e}$ and maximizes $\mathbf{f}^\top \mathbf{C} \mathbf{f}$. Since the eigenvectors of a symmetric matrix like $\mathbf{C}$ form an orthogonal set, the vector $\mathbf{f}$ must be an eigenvector of $\mathbf{C}$ that is orthogonal to $\mathbf{e}$.

Let the eigenvalues of $\mathbf{C}$ be $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$. We have already found that $\mathbf{e}$ corresponds to the largest eigenvalue λ_1.

To maximize $\mathbf{f}^\top \mathbf{C} \mathbf{f}$ subject to the constraint $\mathbf{f}^\top \mathbf{e} = 0$, we must choose $\mathbf{f}$ to be the eigenvector corresponding to the second-largest eigenvalue λ_2, since it provides the maximum possible value of $\mathbf{f}^\top \mathbf{C} \mathbf{f}$ while being orthogonal to $\mathbf{e}$.

Extension to the Third Direction g

Given the assumption $\text{rank}(\mathbf{C}) > 2$, we know that there are at least three non-zero eigenvalues. Therefore, there exists a third eigenvector $\mathbf{g}$, orthogonal to both $\mathbf{e}$ and $\mathbf{f}$.
We now extend the argument to a third direction $\mathbf{g}$, which is perpendicular to both $\mathbf{e}$ and $\mathbf{f}$, and maximizes $\mathbf{g}^\top \mathbf{C}\mathbf{g}$. Since $\mathbf{g}$ is constrained to be a unit vector (i.e., $\|\mathbf{g}\| = 1$), we maximize the quadratic form subject to this normalization constraint.
As before, by the orthogonality of the eigenvectors of $\mathbf{C}$, the vector $\mathbf{g}$ must be an eigenvector of $\mathbf{C}$ that is orthogonal to both $\mathbf{e}$ and $\mathbf{f}$. The vector $\mathbf{g}$ that maximizes $\mathbf{g}^\top \mathbf{C}\mathbf{g}$, subject to $\|\mathbf{g}\| = 1$, is the eigenvector corresponding to the third-largest eigenvalue λ_3.
Conclusion:

1. The direction $\mathbf{f}$ that is orthogonal to $\mathbf{e}$ and maximizes $\mathbf{f}^\top \mathbf{C}\mathbf{f}$ is the eigenvector corresponding to the second-largest eigenvalue λ_2 of $\mathbf{C}$.

2. The direction $\mathbf{g}$, which is orthogonal to both $\mathbf{e}$ and $\mathbf{f}$ and is a unit vector, maximizes $\mathbf{g}^\top \mathbf{C}\mathbf{g}$ and is the eigenvector corresponding to the third-largest eigenvalue λ_3, given that $\text{rank}(\mathbf{C}) > 2$.