ECE/CS/ME 539 – Fall 2024 — Activity Solution 8

Objective

Perform Principal Component Analysis (PCA) by solving the optimization problem to derive the first principal component for a given 2D dataset.

Dataset

We are given a 2D dataset with four data points:

Point	x_1	x_2
1	1	1
2	2	2
3	3	3

Step 1: Center the Data

1. **Calculate the Mean of Each Feature:**

$$\mu_1 = \frac{1}{3}(1+2+3) = 2, \quad \mu_2 = \frac{1}{3}(1+2+3) = 2$$

2. **Center the Data:**

   Subtract the mean from each data point:

Centered Data:

Point	$x_1 - \mu_1$	$x_2 - \mu_2$
1	-1	-1
2	0	0
3	1	1

Step 2: Set Up the PCA Optimization Problem

To find the first principal component, set up the following optimization problem:

$$\text{maximize}_{\phi_{11},\phi_{21}} \quad \frac{1}{3}\sum_{i=1}^{3}\left(\phi_{11}(x_{i1}-\mu_1)+\phi_{21}(x_{i2}-\mu_2)\right)^2$$

Subject to the constraint:

$$\phi_{11}^2 + \phi_{21}^2 = 1$$

Step 3: Solve the Optimization Problem Using Lagrange Multipliers

1. **Set Up the Lagrangian Function:**

$$L(\phi_{11}, \phi_{21}, \lambda) = \frac{1}{3}\sum_{i=1}^{3}(\phi_{11}(x_{i1} - \mu_1) + \phi_{21}(x_{i2} - \mu_2))^2 - \lambda(\phi_{11}^2 + \phi_{21}^2 - 1)$$

2. **Compute Partial Derivatives:** By substituting three data points and take the partial derivatives with respect to ϕ_{11}, ϕ_{21}, and λ, and set them to zero:

$$\frac{\partial L}{\partial \phi_{11}} = \frac{2}{3}(2\phi_{11} + 2\phi_{21}) - 2\lambda\phi_{11} = 0$$

$$\frac{\partial L}{\partial \phi_{21}} = \frac{2}{3}(2\phi_{21} + 2\phi_{11}) - 2\lambda\phi_{21} = 0$$

$$\frac{\partial L}{\partial \lambda} = -(\phi_{11}^2 + \phi_{21}^2 - 1) = 0$$

3. **Solve for ϕ_{11}, ϕ_{21}, and λ:**

   The system of equations is:

$$(4 - 6\lambda)\phi_{11} + 4\phi_{21} = 0$$

$$4\phi_{11} + (4 - 6\lambda)\phi_{21} = 0$$

   For a non-trivial solution, aka ϕ_{11} and ϕ_{21} are not both zero, the determinant of the coefficient matrix must be zero. So we can have:

$$(4 - 6\lambda)^2 - 16 = 0$$

   Solving for λ, we find $\lambda = 0$ or $\lambda = \frac{4}{3}$.
   - For $\lambda = 0$: $\phi_{11} + \phi_{21} = 0 \implies \phi_{11} = \frac{1}{\sqrt{2}}, \phi_{21} = -\frac{1}{\sqrt{2}}$
   - For $\lambda = \frac{4}{3}$: $\phi_{11} - \phi_{21} = 0 \implies \phi_{11} = \phi_{21} = \frac{1}{\sqrt{2}}$

Step 4: Compute the Principal Component

From the solution of the optimization problem, we have two possible sets of values for ϕ_{11} and ϕ_{21}:

1. $\phi_{11} = \frac{1}{\sqrt{2}}$, $\phi_{21} = -\frac{1}{\sqrt{2}}$
2. $\phi_{11} = \frac{1}{\sqrt{2}}$, $\phi_{21} = \frac{1}{\sqrt{2}}$

We choose the second set as it aligns with the direction of the data spread (positive correlation between x_1 and x_2).

First Principal Component

The first principal component is given by the vector:

$$\vec{v} = \begin{pmatrix} \phi_{11} \\ \phi_{21} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$$

Projection onto the Principal Component

For each centered data point (x_1, x_2), the projection onto the first principal component is calculated as:

$$\text{projection} = \phi_{11}x_1 + \phi_{21}x_2$$

Given the centered data points:

$$\begin{aligned} &\text{Point 1:} \quad (-1, -1) \\ &\text{Point 2:} \quad (0, 0) \\ &\text{Point 3:} \quad (1, 1) \end{aligned}$$

We calculate the projections:

$$\begin{aligned} \text{Projection}_1 &= \frac{1}{\sqrt{2}} \cdot (-1) + \frac{1}{\sqrt{2}} \cdot (-1) = -\sqrt{2} \\ \text{Projection}_2 &= \frac{1}{\sqrt{2}} \cdot (0) + \frac{1}{\sqrt{2}} \cdot (0) = 0 \\ \text{Projection}_3 &= \frac{1}{\sqrt{2}} \cdot (1) + \frac{1}{\sqrt{2}} \cdot (1) = \sqrt{2} \end{aligned}$$

So the 1D representation of the centered data along the first principal component is:

$$\begin{pmatrix} -\sqrt{2} \\ 0 \\ \sqrt{2} \end{pmatrix}$$

Step 5: Interpret the Results

1. Values of ϕ_{11} and ϕ_{21}

From our previous calculations:

$$\begin{aligned} \phi_{11} &= \frac{1}{\sqrt{2}} \\ \phi_{21} &= \frac{1}{\sqrt{2}} \end{aligned}$$

These values define the direction of the first principal component in the original 2D space.

2. Variance Captured by the Principal Component

To calculate the variance captured, we proceed as follows:

a) Total variance in the data

$$\begin{aligned}\mathrm{Var}(X) &= \mathrm{Var}(x_1) + \mathrm{Var}(x_2) \\ &= \frac{2}{3} + \frac{2}{3} = \frac{4}{3}\end{aligned}$$

b) Variance along the principal component

This is the average of the squared projections.

Projections: $-\sqrt{2}$, 0, $\sqrt{2}$

Squared projections: 2, 0, 2

Average:

$$\frac{2+0+2}{3} = \frac{4}{3}$$

c) Ratio of variance captured

$$\begin{aligned}\text{Ratio} &= \frac{\text{Variance along PC}}{\text{Total Variance}} \\ &= \frac{4/3}{4/3} = 1\end{aligned}$$

The first principal component captures 100% of the variance in the data.

3. 1D Representation of the Data

The 1D representation of the centered data projected onto the first principal component is:

$$\begin{pmatrix} -\sqrt{2} \\ 0 \\ \sqrt{2} \end{pmatrix}$$

This 1D representation preserves the relative distances between the points along the direction of maximum variance.

Interpretation

1. The values $\phi_{11} = \phi_{21} = \frac{1}{\sqrt{2}}$ indicate that the first principal component is oriented at a 45-degree angle in the original 2D space, showing equal contribution from both original variables.

2. The first principal component captures 100% of the variance in the data. This is a special case due to the perfect correlation in our small dataset. In real-world scenarios with more data points and dimensions, the first principal component usually captures less than 100% of the variance.

3. The 1D representation $\left[-\sqrt{2}, 0, \sqrt{2}\right]$ maintains the relative spacing of the original data points along the direction of maximum variance.

This result demonstrates that for this particular dataset, we can represent all the information using just one dimension instead of two, without losing any information. This exemplifies the core concept of dimensionality reduction in Principal Component Analysis.

Step 6: Connection to Eigenvalue Decomposition

The connection between Principal Component Analysis (PCA) and eigenvalue decomposition lies in the fact that the principal components are eigenvectors of the covariance matrix of the data, and the corresponding eigenvalues represent the amount of variance explained by each principal component.

Covariance Matrix

For our centered 2D dataset, let's first compute the covariance matrix:

$$\mathbf{X} = \begin{bmatrix} -1 & -1 \\ 0 & 0 \\ 1 & 1 \end{bmatrix}$$

The covariance matrix $\mathbf{C}$ is given by:

$$\mathbf{C} = \frac{1}{n-1}\mathbf{X}^T\mathbf{X} = \frac{1}{2}\begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

Eigenvalue Decomposition

The eigenvalue equation for the covariance matrix is:

$$\mathbf{C}\mathbf{v} = \lambda\mathbf{v}$$

where $\mathbf{v}$ is an eigenvector and λ is the corresponding eigenvalue.
Solving this equation:

$$\begin{aligned} \begin{vmatrix} 1-\lambda & 1 \\ 1 & 1-\lambda \end{vmatrix} &= 0 \\ (1-\lambda)^2 - 1 &= 0 \\ \lambda^2 - 2\lambda &= 0 \\ \lambda(\lambda - 2) &= 0 \end{aligned}$$

The eigenvalues are $\lambda_1 = 2$ and $\lambda_2 = 0$.

Eigenvectors

For $\lambda_1 = 2$:

$$\begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

This gives us $v_1 = v_2$, and after normalization:

$$\mathbf{v}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

For $\lambda_2 = 0$:

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

This gives us $v_1 = -v_2$, and after normalization:

$$\mathbf{v}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Connection to PCA

1. The eigenvectors $\mathbf{v}_1$ and $\mathbf{v}_2$ are the principal components we found earlier.
2. The eigenvalues represent the variance explained by each principal component:

- $\lambda_1 = 2$ corresponds to the first PC, explaining 100% of the variance.
- $\lambda_2 = 0$ corresponds to the second PC, explaining 0% of the variance.

3. The optimization problem in PCA is equivalent to finding the eigenvector corresponding to the largest eigenvalue of the covariance matrix.
4. The proportion of variance explained by each PC is given by $\frac{\lambda_i}{\sum_j \lambda_j}$, which matches our earlier calculations.

This eigenvalue decomposition confirms and provides a theoretical foundation for the results we obtained through the optimization approach in the previous steps.

Solution Summary

The principal component values and the variance captured are derived using optimization and are connected to the eigenvalue decomposition technique for PCA.