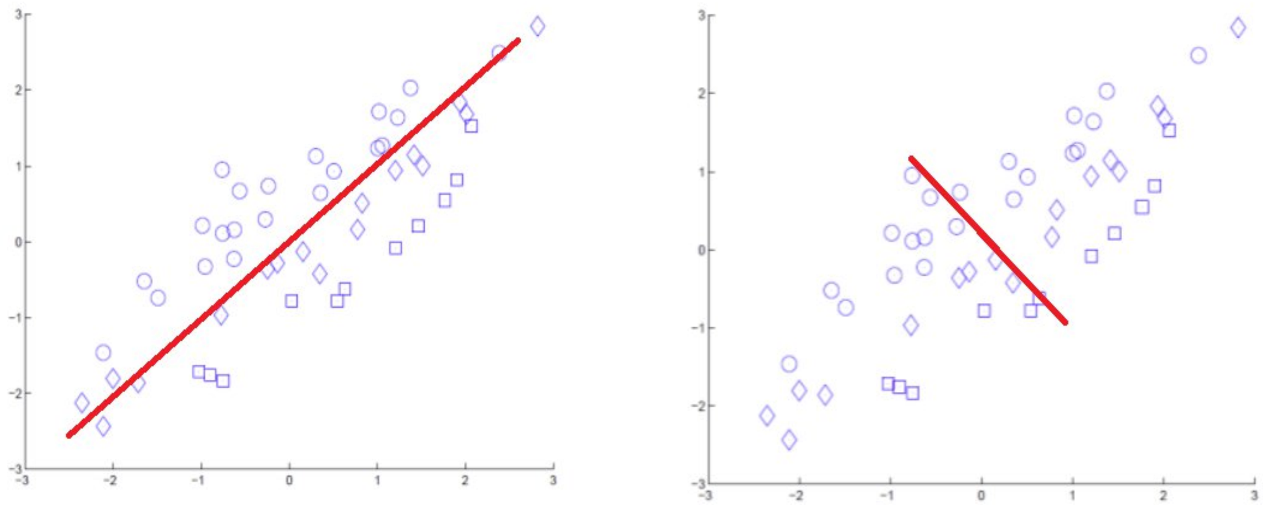
Problem 3: Principal Component Analysis (PCA) (20 points)

(1) Given labels of the data, the goal of Fisher's Linear Discriminant is to find the projection direction that maximizes the ratio of between-class variance and the within-class variance. While PCA aims to reduce the dimension of the data by finding projection directions that maximizes the variance after projection. Note that PCA does not consider the label information. In the following figures, consider round points as positive class, and both diamond and square points as negative class. Please draw (a) the direction of the first principal component in the left figure by ignoring the label of the data points, and (b) the Fisher's linear discriminant direction in the right figure. Please draw a line to show the direction for each of them.

Answer: The direction of the first principal component is shown in the left figure and the fisher's linear discriminant direction in the right figure-



(2) Consider 3 data points in the 2D space: (2,2), (0,0), (-2,-2). Please answer the following questions.

a) Calculate the first principal component by calculating the eigenvalue (non-zero) and eigenvector of the covariance matrix. You need to provide the actual vector of the first principal component (with length=1). You can use the unbiased estimation of the covariance.

b) If we project the three data points into the 1D subspace by the principal component obtained in (a), what are the new coordinates of the three data points in the 1D subspace? What is the variance of the data after projection?

c) What is the cumulative explained variance of the first principal component? Is there any variance that is not captured by it?

Answer:

Problem-3: Principal Component Analysis

(2) Consider data points in the 2D space: (2,2), (0,0), (-2,-2). Please answer the following.

A: a) The given points are: (2,2), (0,0) and (-2,-2)

Mean of X $= \frac{2+0-2}{3} = 0 = \overline{X}$

Mean of Y $= \frac{2+0-2}{3} = 0 = \overline{Y}$

Calculating $var(X) = \frac{1}{N-1} \sum_{n=1}^{N} (X_n - \overline{X})^2$

$Var(X) = \frac{1}{3-1} \left[(2-0)^2 + (0-0)^2 + (-2-0)^2 \right] = \frac{1}{2} \times 8 = \underline{4}$

$Var(Y) = \frac{1}{3-1} \left[(2-0)^2 + (0-0)^2 + (2-0)^2 \right] = \frac{1}{2} \times 8 = \underline{4}$

Calculating $cov(X,Y) = \frac{1}{N-1} \sum_{n=1}^{N} (X_n - \overline{X})(Y_n - \overline{Y})$

$$\text{cov}(X,Y) = \frac{1}{3-1}\left[(2)(2) + (0)(0) + (-2)(-2)\right] = \frac{8}{2} = \underline{4}$$

The covariance matrix is given by,

$$S = \begin{bmatrix} \text{var}(x) & \text{cov}(x,y) \\ \text{cov}(x,y) & \text{var}(y) \end{bmatrix} = \begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix}$$

- Calculating Eigen-values & Eigen-vectors,

$$\det(S - \lambda I) = \det\begin{pmatrix} 4-\lambda & 4 \\ 4 & 4-\lambda \end{pmatrix}$$

$$\Rightarrow (4-\lambda)(4-\lambda) - 16 = 0$$

$$\Rightarrow \cancel{16} - 4\lambda - 4\lambda + \lambda^2 - \cancel{16} = 0$$

$$\Rightarrow \lambda^2 - 8\lambda = 0$$

$$\Rightarrow \lambda(\lambda - 8) = 0$$

$$\therefore \boxed{\lambda_1 = 8} \text{ and } \boxed{\lambda_2 = 0}$$

- Here, $\boxed{\lambda_1 = 8}$ is the first principal component,

$$\begin{bmatrix} -4 & 4 \\ 4 & -4 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\left[\begin{array}{cc|c} -4 & 4 & 0 \\ 4 & -4 & 0 \end{array}\right] \xrightarrow[R_1 \to R_1 * 1/4]{R_2 \to R_2 + R_1} \left[\begin{array}{cc|c} -1 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right] = 0$$

$$-1x_1 + x_2 = 0$$

let $\boxed{x_2 = 1}$ then $-x_1 + 1 = 0$

$$\boxed{x_1 = 1}$$

Therefore, for $\lambda_1 = 8$, $v = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is the eigen vector.

- The eigen vector for second eigen value $\lambda = 0$ is

$$\begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\left[\begin{array}{cc|c} 4 & 4 & 0 \\ 4 & 4 & 0 \end{array}\right] \xrightarrow[R_1 \to R_1 \times \frac{1}{4}]{R_2 \to R_2 - R_1} \left[\begin{array}{cc|c} 1 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right]$$

$x_1 + x_2 = 0$

Let $\boxed{x_2 = 1}$ then $x_1 + 1 = 0 \Rightarrow \boxed{x_1 = -1}$

$\therefore$ For $\lambda_2 = 0$, eigen vector is $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$

This is the second principal component.

- Let e_1 be the normalized eigen vector and the first principal component.

$$e_1 = \begin{bmatrix} 1/\sqrt{1^2+1^2} \\ 1/\sqrt{1^2+1^2} \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 0.707 \\ 0.707 \end{bmatrix}$$

2) b) Now, if we project the three data points into the 1D subspace, the principal component obtained in (a),

- Let P_{11}, P_{12} and P_{13} be the new first principal component analysis,

- Now, using e_1 from (a),

- $P_{11} = e_1^T \begin{bmatrix} X_1 - \overline{X} \\ Y_1 - \overline{Y} \end{bmatrix} = \begin{bmatrix} 0.707 & 0.707 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix}$

$= 1.414 + 1.414 = 2.828$

$\therefore \boxed{P_{11} = 2.828}$

- $P_{12} = e_1^T \begin{bmatrix} X_2 - \overline{X} \\ Y_2 - \overline{Y} \end{bmatrix} = \begin{bmatrix} 0.707 & 0.707 \end{bmatrix} \begin{bmatrix} 0-0 \\ 0-0 \end{bmatrix}$

$= 0 + 0 = 0$

$\therefore \boxed{P_{12} = 0}$

- $P_{13} = e_1^T \begin{bmatrix} X_3 - \overline{X} \\ Y_3 - \overline{Y} \end{bmatrix} = \begin{bmatrix} 0.707 & 0.707 \end{bmatrix} \begin{bmatrix} -2 \\ -2 \end{bmatrix}$

$= -1.414 - 1.414 = -2.828$

$\therefore \boxed{P_{13} = -2.828}$

Therefore, the new coordinates of the three data points in 1D subspace is,

PC1	$P_{11} = 2.828$	$P_{12} = 0$	$P_{13} = -2.828$
	DP_1	DP_2	DP_3

where DP_1, DP_2 & DP_3 are data points,

- The ~~do~~ variance of the data after projection is,

$$\mathrm{var}(X) = \frac{1}{N-1} \sum_{n=1}^{N} (x_n - \bar{x})^2$$

$$= \frac{1}{3-1} \left[(2.828-0)^2 + (0-0)^2 + (-2.828-0)^2 \right]$$

$$= \frac{1}{2} [16]$$

$$\boxed{\mathrm{var}(X) = 8}$$

2) c) Cumulative explained variance of the first Principal component is,

$$= \lambda_1 / (\lambda_1 + \lambda_2)$$

$$= \frac{8}{8+0} = 1/1$$

$\therefore$ The first principal component captures the complete variance.