Machine Learning

ASSIGNMENT 8

(75 points) **Question** This problem is on principal component analysis (PCA). The set of data points is $\{x_1, x_2, x_3\}$. These are

$$x_1 = \begin{bmatrix} 0 & 1 \end{bmatrix}^T, \quad x_2 = \begin{bmatrix} -2 & 0 \end{bmatrix}^T, \quad x_3 = \begin{bmatrix} -1 & -1 \end{bmatrix}^T$$

Determine the principal component of these data points. Also determine the corresponding error in the approximation.

Hint: The student is expected to complete the intermediate steps/ calculations.

Step 1: The number of data points is $n = 3$. Let the average value of the data points be

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \begin{bmatrix} -1 & 0 \end{bmatrix}^T$$

Define a column vector

$$\theta_i = (x_i - \overline{x}), \quad 1 \le i \le n$$

and a 2×3 matrix Θ as

$$\Theta = \begin{bmatrix} \theta_1 & \theta_2 & \theta_3 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix}$$

Step 2: The covariance matrix $\widetilde{\Sigma}$ of the data points is

$$\widetilde{\Sigma} = \frac{1}{(n-1)}\Theta\Theta^T = \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1 \end{bmatrix}$$

The 2×2 covariance matrix $\widetilde{\Sigma}$ is symmetric and positive-definite (in this problem).

Step 3: Therefore $\widetilde{\Sigma} = \Psi\Lambda\Psi^T$where Λ is a diagonal matrix with eigenvalues of the matrix $\widetilde{\Sigma}$ on its main diagonal. As the matrix $\widetilde{\Sigma}$ is symmetric and positive-definite, its eigenvalues are positive. Let the eigenvalues of the matrix $\widetilde{\Sigma}$ be $\lambda_j > 0, 1 \le j \le 2$. The columns of the matrix Ψ are the mutually orthogonal eigenvectors of the covariance matrix $\widetilde{\Sigma}$. Assume that these eigenvectors are orthonormal. Therefore $\Psi\Psi^T = I$, where I is an identity matrix of size 2. We determine the eigenvalues λ_1 and λ_2 and the corresponding eigenvectors ψ_1 and ψ_2 in this step.

(a) The eigenvalues of the matrix $\widetilde{\Sigma}$ are $\lambda_1 = 1/2$, and $\lambda_2 = 3/2$.

The corresponding eigenvectors of length one are

$$\psi_1 = \pm\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \end{bmatrix}^T, \quad \text{and} \quad \psi_2 = \pm\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \end{bmatrix}^T$$

respectively. Observe that $\psi_1^T\psi_2 = 0$. That is, these vectors are orthonormal.

(b) Thus

$$\Psi = \begin{bmatrix} \psi_1 & \psi_2 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$$

It can be verified that $\Psi\Psi^T = I$, where I is an identity matrix of size 2.

(c) Verify by direct calculation, that $\Psi\Lambda\Psi^T = \widetilde{\Sigma}$.

Step 4: As $\widetilde{\Sigma} = \Psi\Lambda\Psi^T$, we have

$$\begin{aligned} \Lambda &= \Psi^T\widetilde{\Sigma}\Psi = \frac{1}{(n-1)}\Psi^T\Theta\Theta^T\Psi \\ &= \frac{1}{(n-1)}WW^T \end{aligned}$$

where

$$\begin{aligned} W &= \Psi^T\Theta = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix} \\ &= \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & -1 & 1 \\ 2 & -1 & -1 \end{bmatrix} \end{aligned}$$

is a 2×3 matrix. Also

$$\Theta = \Psi W$$

Let

$$W = \begin{bmatrix} w_1 & w_2 & w_3 \end{bmatrix}$$

Thus

$$\theta_i = \Psi w_i, \quad 1 \le i \le 3$$

Step 5: We next approximate Θ as $\widehat{\Theta}$, where

$$\widehat{\Theta} = \begin{bmatrix} \widehat{\theta}_1 & \widehat{\theta}_2 & \widehat{\theta}_3 \end{bmatrix}$$

Recall that eigenvalues of the covariance matrix $\widetilde{\Sigma}$ are $\lambda_1 = 1/2$, and $\lambda_2 = 3/2$. As $\lambda_2 > \lambda_1$, we approximate $\Theta = \Psi W$ as

$$\widehat{\Theta} = \widehat{\Psi} W$$

where

$$\widehat{\Psi} = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$$

Thus

$$\begin{aligned} \widehat{\Theta} &= \widehat{\Psi}W = \frac{1}{2} \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & -1 & 1 \\ 2 & -1 & -1 \end{bmatrix} \\ &= \frac{1}{2} \begin{bmatrix} 2 & -1 & -1 \\ 2 & -1 & -1 \end{bmatrix} \end{aligned}$$

Therefore

$$\widehat{\Theta} = \frac{1}{2}\begin{bmatrix} 2 & -1 & -1 \\ 2 & -1 & -1 \end{bmatrix} = \begin{bmatrix} \widehat{\theta}_1 & \widehat{\theta}_2 & \widehat{\theta}_3 \end{bmatrix}$$

We summarize θ_i and $\widehat{\theta}_i$, for $1 \leq i \leq 3$ as

$$\begin{aligned} \theta_1 &= \begin{bmatrix} 1 & 1 \end{bmatrix}^T, \quad \text{and} \quad \widehat{\theta}_1 = \begin{bmatrix} 1 & 1 \end{bmatrix}^T \\ \theta_2 &= \begin{bmatrix} -1 & 0 \end{bmatrix}^T, \quad \text{and} \quad \widehat{\theta}_2 = \begin{bmatrix} -1/2 & -1/2 \end{bmatrix}^T \\ \theta_3 &= \begin{bmatrix} 0 & -1 \end{bmatrix}^T, \quad \text{and} \quad \widehat{\theta}_3 = \begin{bmatrix} -1/2 & -1/2 \end{bmatrix}^T \end{aligned}$$

Therefore, the mean-squared error E_{err} in the approximation of a data point is

$$E_{err} = \frac{1}{(3-1)} \sum_{i=1}^{3} \left\| \theta_i - \widehat{\theta}_i \right\|^2$$

where $\|\cdot\|$ is the Euclidean norm. It is

$$E_{err} = \frac{1}{2} = \lambda_1$$

Step 6: Therefore the principal component is $\lambda_2 = 3/2$, and its direction is given by the eigenvector ψ_2. The error in the approximation is $E_{err} = 0.5$.