Date 12/5/25
Page

Lab 10

Implement dimensionality reduction using Principle Component Analysis (PCA)

Reduces dimension

Feature	Example 1	Example 2	Example 3	Example 4
x1	4	8	13	7
x2	11	4	5	14

Standardized data:

$$\bar{x_1} = \frac{4+8+13+7}{4} = 8$$

$$\bar{x_2} = \frac{11+4+5+14}{4} = 8.5$$

$$X_c^T = \begin{bmatrix} 4-8 & 8-8 & 13-8 & 7-8 \\ 11-8.5 & 4-8.5 & 5-8.5 & 14-8.5 \end{bmatrix}$$

$$X_c = \begin{bmatrix} -4 & 0 & 5 & -1 \\ -8.5 & -4.5 & -3.5 & 5.5 \end{bmatrix}$$

Covariance matrix:

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

Date 2/5/25
Page

$$\Rightarrow \frac{1}{3}\begin{bmatrix} -4 & 0 & 5 & -1 \\ +2.5 & -4.5 & -3.5 & 5.5 \end{bmatrix}\begin{bmatrix} -4 & +25 \\ 0 & -45 \\ 5 & -35 \\ -1 & 5.5 \end{bmatrix}$$

$$C = \frac{1}{3}\begin{bmatrix} 42 & -33 \\ -33 & 69 \end{bmatrix}$$

$$C = \begin{bmatrix} 14 & -11 \\ -11 & 23 \end{bmatrix}$$

Let $(C - \lambda I) = 0$

$$\begin{bmatrix} 14-\lambda & -11 \\ -11 & 23-\lambda \end{bmatrix} = 0$$

$$\lambda^2 - 37\lambda + 201 = 0$$

$$\lambda = 30.3849$$

Consider larger λ

$$\begin{bmatrix} 14 - 30.3849 & -11 \\ -11 & 23 - 30.3849 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} -16.3849x & -11y \\ -11x & -7.3849y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$-16.3849x - 11y = 0$$

$$x = -0.671y$$

Date 12/05/25
Page

take $y = 1$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -0.6741 \\ 1 \end{bmatrix}$$

Normalize

$$x = \frac{x}{\sqrt{x^2+y^2}} \; ; \; y = \frac{y}{\sqrt{x^2+y^2}}$$

$$C_1 = \begin{bmatrix} -0.5574 \\ 0.8303 \end{bmatrix}$$

Calculate PC :

$$Z = X C_1 = \begin{bmatrix} 4 & 26 \\ 0 & 48 \\ 5 & -35 \\ -1 & 55 \end{bmatrix} \begin{bmatrix} -0.5574 \\ 0.8303 \end{bmatrix}$$

$$Z = \begin{bmatrix} +4.305 \\ -3.732 \\ -5.693 \\ +5.134 \end{bmatrix}$$

For heart-csv dataset

Accuracy before PCA	Accuracy after PCA
Logistic regression: 0.9016	Logistic: 0.8689
SVM: 0.8525	SVM: 0.8689
Random forest: 0.8361	Random forest: 0.8689