14/10/2025

PCA
Principal Component Analysis

Assignment 1: Iris Dataset

Example #1

first 10 Rows of Iris Dataset.

Row	Sepal L	Sepal W	Petal L	Petal W
1	5.1	3.5	1.4	0.2
2	4.9	3.0	1.4	0.2
3	4.7	3.2	1.3	0.2
4	4.6	3.1	1.5	0.2
5	5.0	3.6	1.4	0.2
6	5.4	3.9	1.7	0.3
7	4.6	3.4	1.4	0.3
8	5.0	3.4	1.5	0.2
9	4.4	2.9	1.4	0.2
10	4.9	3.1	1.5	0.1

Step! 1 => create Data Matrix X

$$X = \begin{bmatrix} 5.1 & 3.5 & 1.4 & 0.2 \\ 4.9 & 3.0 & 1.4 & 0.2 \\ 4.7 & 3.2 & 1.3 & 0.2 \\ 4.6 & 3.1 & 1.5 & 0.2 \\ 5.0 & 3.6 & 1.4 & 0.2 \\ 5.4 & 3.9 & 1.7 & 0.3 \\ 4.6 & 3.4 & 1.4 & 0.3 \\ 5.0 & 3.4 & 1.5 & 0.2 \\ 4.4 & 2.9 & 1.4 & 0.2 \\ 4.9 & 3.1 & 1.5 & 0.1 \end{bmatrix}$$

Step 2 Standardize the features!:

* Mean of 0
* Standard Deviation of 1

Column Means:

Feature	Mean
Sepal L	4.86
Sepal W	3.31
Petal L	1.45
Petal W	0.22

We'll go with 4X4 Matrix (A Row of Dataset).

$$X = \begin{bmatrix} 5.1 & 3.5 & 1.4 & 0.2 \\ 4.9 & 3.0 & 1.4 & 0.2 \\ 4.7 & 3.2 & 1.3 & 0.2 \\ 4.6 & 3.1 & 1.5 & 0.2 \end{bmatrix}$$

$$\text{Mean } X = \begin{bmatrix} 4.825 & 3.2 & 1.4 & 0.2 \end{bmatrix}$$

Sepal Length Sepal Width Petal(L) Petal(W)

Step 3: Subtract the Mean (Center the Data)

Sepal L	Sepal W	Petal L	Petal W
$5.1 - 4.825 = 0.275$	$3.5 - 3.2 = 0.3$	$1.4 - 1.4 = 0.0$	$0.2 - 0.2 = 0.0$
$4.9 - 4.825 = 0.075$	$3.0 - 3.2 = -0.2$	$1.4 - 1.4 = 0.0$	$0.2 - 0.2 = 0.0$
$4.7 - 4.825 = -0.125$	$3.2 - 3.2 = 0$	$1.3 - 1.4 = -0.1$	$0.2 - 0.2 = 0.0$
$4.6 - 4.825 = -0.225$	$3.1 - 3.2 = -0.1$	$1.5 - 1.4 = 0.1$	$0.2 - 0.2 = 0.0$

Step 4: Compute Covariance

$$Cov(x,y) = \frac{1}{n-1} \sum (x_i - \bar{x})(y_i - \bar{y})$$

Here $n = 4$, so $n-1 \Rightarrow 4-1 = 3$

All pairs of Features (General Rule for PCA)

		x_1 Sepal L	x_2 Sepal W	x_3 petal L	x_4 petal W
x_1	Sepal L	$Var(x_1)$	$Cov(x_1 x_2)$	$Cov(x_1 x_3)$	$Cov(x_1 x_4)$
x_2	Sepal W	$Cov(x_2 x_1)$	$Var(x_2)$	$Cov(x_2 x_3)$	$Cov(x_2 x_4)$
x_3	petal L	$Cov(x_3 x_1)$	$Cov(x_3 x_2)$	$Var(x_3)$	$Cov(x_3 x_4)$
x_4	petal W	$Cov(x_4 x_1)$	$Cov(x_4 x_2)$	$Cov(x_4 x_3)$	$Var(x_4)$

Covariance Matrix

$$Var(x_1) = Cov(x_1, x_1) \Rightarrow [(0.275)^2 + (0.075)^2 + (0.125)^2 + (0.225)^2]/3$$
$$= 0.1474/3 = 0.0491$$

$$Var(x_2) = Cov(x_2, x_2) = [(0.3)^2 + (-0.2)^2 + 0^2 + (-0.1)^2]/3$$
$$Var(x_2) = 0.14/3 = 0.0467$$

$$Var(x_3) = Cov(x_3, x_3) = [0.0^2 + 0.0^2 + (-0.1)^2 + (0.1)^2]/3$$
$$= 0.02/3 = 0.0067$$

$$Var(x_4) = Cov(x_4, x_4) = [0.0^2 + 0.0^2 + 0.0^2 + 0.0^2] = 0$$

$$Cov(x_1 x_2) = [(0.275)(0.3) + (0.075) \times (-0.2) + (-0.125) \times (0) + (-0.225)(-0.1)]/3$$

$$= 0.09/3 = 0.03$$

$$Cov(X_1, X_3) = [0.0275 \times 0 + 0.075 \times 0 + (-0.125) \times (-0.1) + (-0.225)(0.1)]/3$$

$$= -0.01/3 = -0.0033$$

$$Cov(X_2, X_3) = [0.3 \times 0 + (-0.2) \times 0 + 0 + (-0.1)(0.1)]/3$$

$$= -0.01/3 = -0.0033$$

$Cov(X_4, X_4) = 0$ Since All X_4 are 0

Step 5: Build Covariance Matrix.

$$\Sigma = \begin{bmatrix} 0.0491 & 0.0300 & -0.0033 & 0 \\ 0.0300 & 0.0467 & -0.0033 & 0 \\ -0.0033 & -0.0033 & 0.0067 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Step 6: Calculating Eigen Values & Eigen Vectors

$$(\Sigma - \lambda I)v = 0$$

Solving

$$\det(\Sigma - \lambda I) = 0$$

$$\begin{bmatrix} 0.0491 & 0.0300 & -0.0033 & 0 \\ 0.0300 & 0.0467 & -0.0033 & 0 \\ -0.0033 & -0.0033 & 0.0067 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} - \begin{bmatrix} \lambda & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 \\ 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & \lambda \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 0.0491-\lambda & 0.0300 & -0.0033 & 0 \\ 0.0300 & 0.0467-\lambda & -0.0033 & 0 \\ -0.0033 & -0.0033 & 0.0067-\lambda & 0 \\ 0 & 0 & 0 & -\lambda \end{bmatrix} = 0$$

$$\det(\Sigma - \lambda I) = 0 \Rightarrow \lambda = \{0.078253, 0.017891, 0.006356, 0\}$$

Explained Variance Ratio:

PC	Eigen Value	% Variance
1	0.078523	76.34%
2	0.017891	17.46%
3	0.006356	6.20%
4	0	0%

Eigen Vectors (v)

$$(\Sigma - \lambda I)v = 0 \qquad v_1 = 0.078253$$

$$(\Sigma - 0.078253) = \begin{bmatrix} -0.029086 & 0.030000 & -0.009333 \\ 0.030000 & -0.031586 & -0.009333 \\ -0.009333 & -0.009333 & -0.071586 \end{bmatrix} * \begin{bmatrix} x \\ y \\ z \end{bmatrix} = 0$$

$$v_1 = [-0.704, 0.710, -0.027]^T$$

$$v_1 = [0.707, 0.707, -0.027]$$

PC1 combines mainly Sepal Length & Width $\Rightarrow$ Roughly Equal Weight