Principal Component Analysis

Def: PCA is a dimensionality reduction algorithm that finds orthogonal (uncorrelated) direction in the data where variance is maximized.

Purpose: Imagine a dataset with 42 features and 1 target. Building a ML model on this dataset will require extensive computational power. But if you apply PCA, you can get most of the information of all 42 features in k (k<42) features.

Advantages

1. Maximizes speed
2. Less feature but maximum information
3. Improves accuracy (often)

Disadvantages

1. Loss of explainability
2. Assumes linearity
3. Sensitive to scaling

Example : A dataset is given. Reduce the number of features from 3 to 2. (i) Apply PCA to all features (ii)

X_1	X_2	X_3	y
2	3	5	-
4	6	10	-
1	5	8	-
3	4	7	-

Solution: Step 1 - Center the data

$\mu_1 x_1 = 2.5$, $\mu_2 x_2 = 4.5$, $\mu_3 x_3 = 7.5$

$$X_{center} = \begin{bmatrix} -0.5 & -1 & -2 \\ 1.5 & 2 & 3 \\ -1.5 & 1 & 1 \\ 0.5 & 0 & 0 \end{bmatrix}$$

Step 2: Computing covariance matrix.

$$C = \frac{1}{n-1} X^T_{Centred} \; X_{Center}$$

$$= \frac{1}{4-1} \begin{bmatrix} -0.5 & 1.5 & -1.5 & 0.5 \\ -1 & 2 & 1 & 0 \\ -2 & 3 & 1 & 0 \end{bmatrix} \begin{bmatrix} -0.5 & -1 & -2 \\ 1.5 & 2 & 3 \\ -1.5 & 1 & 1 \\ 0.5 & 0 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 1.67 & 1.83 & 2.67 \\ 1.83 & 3.67 & 5.33 \\ 2.67 & 5.33 & 8.67 \end{bmatrix}$$

Step 3: Computing eigenvalues and eigenvectors

We need to solve the following characteristic equation.

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

$$\lambda I = \lambda \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix}$$

$$C-\lambda I = \begin{bmatrix} 1.67 & 1.83 & 2.67 \\ 1.83 & 3.67 & 5.33 \\ 2.67 & 5.33 & 8.67 \end{bmatrix} - \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix}$$

$$= \begin{bmatrix} 1.67-\lambda & 1.83 & 2.67 \\ 1.83 & 3.67-\lambda & 5.33 \\ 2.67 & 5.33 & 8.67-\lambda \end{bmatrix}$$

$$\det(C-\lambda I) = (1.67-\lambda)\{(3.67-\lambda)(8.67-\lambda) - (5.33)^2\}$$

$$- 1.83\ \{(1.83)(8.67-\lambda) - (5.33 \times 2.67)\}$$

$$+ 2.67\ \{(1.83 \times 5.33) - (2.67 \times (3.67-\lambda))\}$$

$$= k\lambda^2 + ----$$

$\lambda_1 = 13.23,\ \lambda_2 = 0.44,\ \lambda_3 = 0$

(Top 2 eigenvalues : $\lambda_1 = 13.23,\ \lambda_2 = 0.44$)

Eigenvector for λ_1 : $V_1 = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$

$$(C - \lambda_1 I) V_1 = 0$$

$$\Rightarrow \begin{bmatrix} 1.67-\lambda_1 & 1.83 & 2.67 \\ 1.83 & 3.67-\lambda_1 & 5.33 \\ 2.67 & 5.33 & 8.67-\lambda \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\Rightarrow a_1(1.67 - 13.23) + a_2(1.83) + a_3(2.67) = 0$$

$$a_1(1.83) + a_2(3.67 - 13.23) + a_3(5.33) = 0$$

$$a_1(2.67) + a_2(5.33) + a_3(8.67 - \lambda) = 0$$

solving;

$$V_1 = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} 0.24 \\ 0.48 \\ 0.84 \end{bmatrix}$$

similarly,

$$V_2 = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} -0.82 \\ 0.56 \\ -0.11 \end{bmatrix}$$

similarly, $v_3 = \begin{bmatrix} 0.52 \\ 0.67 \\ -0.53 \end{bmatrix}$

Final step: Project data onto v_1 and v_2

Recall that, we need to reduce dim by 1.

Formula: $Z = X_{centered} \cdot V_K$

$$= \begin{bmatrix} -0.5 & 1 & -2 \\ 1.5 & 2 & 3 \\ -1.5 & 1 & 1 \\ 0.5 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.24 & -0.82 \\ 0.48 & 0.56 \\ 0.84 & -0.11 \end{bmatrix}$$

$$= \begin{bmatrix} -2.31 & 0.78 \\ 4.62 & -1.56 \\ -0.77 & 2.34 \\ 0.00 & -0.39 \end{bmatrix}$$

Final Reduced data

PC 1	PC 2	y
-2.31	0.78	–
4.62	-1.56	–
-0.77	2.34	–
0.00	-0.39	–