PCA

Example 1: Given the data in Table, reduce the dimension from 2 to 1 using the Principal Component Analysis (PCA).

Feature	Sample 1	Sample-2	Sample-3	Sample-4
x_1	4	8	13	7
x_2	11	4	5	14

$N = 4$

Step 1: Calculate Mean:

$$\bar{x}_1 = \frac{1}{4}(4+8+13+7) = 8$$

$$\bar{x}_2 = \frac{1}{4}(11+4+5+14) = 8.5$$

Step 2: Calculation of the covariance matrix.

$$S = \begin{bmatrix} Cov(x_1,x_1) & Cov(x_1,x_2) \\ Cov(x_2,x_1) & Cov(x_2,x_2) \end{bmatrix}$$

$$Cov(x_1,x_1) = \frac{1}{N-1}\sum_{K=1}^{N}(x_{1K}-\bar{x}_1)(x_{1K}-\bar{x}_1)$$

$$= \frac{1}{3}\left((4-8)^2 + (8-8)^2 + (13-8)^2 + (7-8)^2\right)$$

$$= 14$$

$$Cov(x_1,x_2) = \frac{1}{N-1}\sum_{K=1}^{N}(x_{1K}-\bar{x}_1)(x_{2K}-\bar{x}_2)$$

$$= \frac{1}{3}\left((4-8)(11-8.5) + (8-8)(4-8.5)\right.$$
$$\left. + (13-8)(5-8.5) + (7-8)(14-8.5)\right)$$

$$= -11$$

$$Cov(x_2,x_1) = Cov(x_1,x_2) = -11$$

$$Cov(x_2,x_2) = \frac{1}{N-1}\sum_{K=1}^{N}(x_{2K}-\bar{x}_2)(x_{2K}-\bar{x}_2)$$

$$= \frac{1}{3}\left((11-8.5)^2 + (4-8.5)^2 + (5-8.5)^2 + (14-8.5)^2\right)$$

$$= 23$$

$$\Rightarrow \quad S = \begin{bmatrix} 14 & -11 \\ -11 & 23 \end{bmatrix}$$

Step 3: Eigenvalues of the covariance matrix

Determinant$(S - \lambda I) = 0$, I is Identity Matrix

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

$$\Rightarrow (14-\lambda)(23-\lambda) - (-11)\times(-11) = 0$$

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

So, $\lambda = \frac{1}{2}(37 \pm \sqrt{565})$

$= 30.3849,\ 6.6151$

$= \lambda_1,\ \lambda_2$ (Say)

Step 4: Computation of the eigenvectors, $U = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$

$$\Rightarrow (S - \lambda I)U = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 14-\lambda & -11 \\ -11 & 23-\lambda \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} (14-\lambda)u_1 - 11u_2 \\ -11u_1 + (23-\lambda)u_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\Rightarrow \begin{aligned} (14-\lambda)u_1 - 11u_2 &= 0 \\ -11u_1 + (23-\lambda)u_2 &= 0 \end{aligned}$$

$$\Rightarrow \frac{u_1}{11} = \frac{u_2}{14-\lambda} = t$$

$$\Rightarrow u_1 = 11t,\quad u_2 = (14-\lambda)t$$

If we assume $t = 1$, then, $u_1 = 11$, $u_2 = 14-\lambda$.

$$\Rightarrow U_0 = \begin{bmatrix} 11 \\ 14-\lambda \end{bmatrix}$$

⊛ The second eigen value can be left out as $\lambda_2 (6.6151) < \lambda_1$

So, we will calculate the eigen vector corresponding to the eigen value $\lambda_1 = 30.3849$.

$$U_1 = \begin{bmatrix} 11 \\ 14-\lambda_1 \end{bmatrix}$$

* To find a unit eigen vector (e_1) we compute the length of U_1, which is given by,

$$\|U_1\| = \sqrt{11^2 + (14-\lambda_1)^2}$$

$$= \sqrt{11^2 + (14 - 30.3849)^2} = 19.7348$$

$$e_1 = \begin{bmatrix} 11/\|U_1\| \\ (14-\lambda_1)/\|U_1\| \end{bmatrix} = \begin{bmatrix} 11/19.7348 \\ (14-30.3849)/19.7348 \end{bmatrix}$$

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

* Thus, principal component for the given data set is.

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} 0.5574 \\ -0.8303 \end{bmatrix}$$

* The feature vector (4, 11) gets transformed

to = Transpose of Eigen vector × [Feature Vector - Mean Vector]

$$= \begin{bmatrix} 0.5574 \\ -0.8303 \end{bmatrix}^T \times \left(\begin{bmatrix} 4 \\ 11 \end{bmatrix} - \begin{bmatrix} 8 \\ 8.5 \end{bmatrix} \right)$$

$$= \begin{bmatrix} 0.5574 & -0.8303 \end{bmatrix} \times \begin{bmatrix} -4 \\ 2.5 \end{bmatrix}$$

$$= 4.30535$$