PCA Example

Sample	Hours x_1	Score x_2
A	2	4
B	4	6
C	6	8

$$A = \begin{bmatrix} 2 & 4 \\ 4 & 6 \\ 6 & 8 \end{bmatrix}$$

Compute the mean of each feature

Compute the mean of each column

$$A = \begin{bmatrix} 2 & 4 \\ 4 & 6 \\ 6 & 8 \end{bmatrix}$$

$$\mu_1 = \frac{2+4+6}{3} = 4 \qquad \mu_2 = \frac{4+6+8}{3} = 6$$

$$\boxed{\mu = \begin{bmatrix} 4 & 6 \end{bmatrix}}$$

Center the Data

$$A = \begin{bmatrix} 2 & 4 \\ 4 & 6 \\ 6 & 8 \end{bmatrix}$$

$$\mu = \begin{bmatrix} 4 & 6 \end{bmatrix}$$

Subtract the mean for each sample

$$A_{centered} = \begin{bmatrix} 2-4 & 4-6 \\ 4-4 & 6-6 \\ 6-4 & 8-6 \end{bmatrix}$$

$$\boxed{A_{centered} = \begin{bmatrix} -2 & -2 \\ 0 & 0 \\ 2 & 2 \end{bmatrix}}$$

Compute the Covariance Matrix

$$\boxed{\Sigma = \frac{1}{n-1} A_{centered}^{T} A_{centered}}$$

$$A_{centered} = \begin{bmatrix} -2 & -2 \\ 0 & 0 \\ 2 & 2 \end{bmatrix} \qquad A_{centered}^{T} = \begin{bmatrix} -2 & 0 & 2 \\ -2 & 0 & 2 \end{bmatrix}$$

Compute the transpose of $A_{centered}$

Compute the Covariance Matrix

$$\boxed{\Sigma = \frac{1}{n-1} A_{centered}^{T} A_{centered}}$$

$$A_{centered} = \begin{bmatrix} -2 & -2 \\ 0 & 0 \\ 2 & 2 \end{bmatrix} \qquad A_{centered}^{T} = \begin{bmatrix} -2 & 0 & 2 \\ -2 & 0 & 2 \end{bmatrix}$$

$$\Sigma = \frac{1}{n-1} A_{centered}^{T} A_{centered}$$

Use the formula by plugging in the matrices

Compute the Covariance Matrix

$$\boxed{\Sigma = \frac{1}{n-1} A_{centered}^{T} A_{centered}}$$

$$A_{centered} = \begin{bmatrix} -2 & -2 \\ 0 & 0 \\ 2 & 2 \end{bmatrix} \qquad A_{centered}^{T} = \begin{bmatrix} -2 & 0 & 2 \\ -2 & 0 & 2 \end{bmatrix}$$

$$\Sigma = \frac{1}{n-1} \begin{bmatrix} -2 & 0 & 2 \\ -2 & 0 & 2 \end{bmatrix} \begin{bmatrix} -2 & -2 \\ 0 & 0 \\ 2 & 2 \end{bmatrix}$$

Compute the Covariance Matrix

$$\Sigma = \frac{1}{n-1}\begin{bmatrix} -2 & 0 & 2 \\ -2 & 0 & 2 \end{bmatrix}\begin{bmatrix} -2 & -2 \\ 0 & 0 \\ 2 & 2 \end{bmatrix}$$

Multiply the two matrices

$$\Sigma = \frac{1}{3-1}\begin{bmatrix} 8 & 8 \\ 8 & 8 \end{bmatrix}$$

Because we have three samples in our dataset, n will be 3

$$\Sigma = \frac{1}{2}\begin{bmatrix} 8 & 8 \\ 8 & 8 \end{bmatrix}$$

Now, divide the matrix by 2

$$\boxed{\Sigma = \begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix}}$$

$$\sum = \begin{pmatrix} cov(X,X) & cov(X,Y) \\ cov(X,Y) & cov(Y,Y) \end{pmatrix}$$

$$\Sigma = \begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix}$$

- Variance of x_1 is 4
- Variance of x_2 is 4
- Covariance is 4

$$\boxed{Av = \lambda v}$$

Where:

$\boldsymbol{A}$ is a covariance matrix

$\boldsymbol{v}$ is the eigenvector

$\boldsymbol{\lambda}$ is the eigenvalue

$$\boxed{Av = \lambda v}$$

$$A v - \lambda v = 0$$

$$(A - \lambda)v = 0$$

$$\boxed{(A - \lambda I)v = 0} \quad ①$$

$$\boxed{det(A - \lambda I) = 0} \quad ②$$

I is an identity matrix $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

Compute the determinant

$$\boxed{det(A - \lambda I) = 0}$$

2

Plug in the identity matrix I and covariance matrix A

$$\left|\begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right| = 0$$

Perform scalar matrix multiplication

$$\left|\begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right| = 0$$

Subtract matrices

$$\begin{vmatrix} 4-\lambda & 4 \\ 4 & 4-\lambda \end{vmatrix} = 0$$

Find the determinant of the matrix

$$(4-\lambda)^2 - 16 = 0$$

Solve for the eigenvalues

$$(4-\lambda)^2 - 16 = 0$$

Simplify

$$(4-\lambda)(4-\lambda) - 16 = 0$$

$$16 - 4\lambda - 4\lambda + \lambda^2 - 16 = 0$$

$$\lambda^2 - 8\lambda = 0$$

$$(\lambda - 0)(\lambda - 8) = 0$$

$$\boxed{\lambda_1 = 8}$$

$$\boxed{\lambda_2 = 0}$$

$$A = \begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix}$$

$$v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

$$\lambda_1 = 8$$

Find the eigenvectors

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

1

Plug in the identity matrix I and covariance matrix A

$$\left(\begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0$$

Perform scalar matrix multiplication

$$\left(\begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right) \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0$$

Replace lambda with the first eigenvalue

$$\left(\begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix} - \begin{bmatrix} 8 & 0 \\ 0 & 8 \end{bmatrix}\right) \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0$$

Subtract matrices

$$\begin{bmatrix} -4 & 4 \\ 4 & -4 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0$$

Find the eigenvectors

$$A = \begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix}$$

$$v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

$$\boldsymbol{\lambda_1} = 8$$

$$\begin{bmatrix} -4 & 4 \\ 4 & -4 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0$$

Multiply the two matrices

$$\begin{bmatrix} -4v_1 + 4v_2 = 0 \\ 4v_1 - 4v_2 = 0 \end{bmatrix}$$

$$\begin{bmatrix} 4v_2 = 4v_1 \\ 4v_1 = 4v_2 \end{bmatrix}$$

$$\begin{bmatrix} v_2 = v_1 \\ v_1 = v_2 \end{bmatrix}$$

$$v = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Find the eigenvectors

$$A = \begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix}$$

$$v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

$$\boldsymbol{\lambda_2} = 0$$

$$\boxed{(A - \lambda I)v = 0}$$

1

Plug in the identity matrix I and covariance matrix A

$$\left(\begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0$$

Perform scalar matrix multiplication

$$\left(\begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right) \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0$$

Replace lambda with the first eigenvalue

$$\left(\begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}\right) \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0$$

Subtract matrices

$$\begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0$$

$$A = \begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix}$$

$$v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

$$\boldsymbol{\lambda_2} = 0$$

Find the eigenvectors

$$\begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0$$

Multiply the two matrices

$$\begin{bmatrix} 4v_1 + 4v_2 = 0 \\ 4v_1 + 4v_2 = 0 \end{bmatrix} \quad \begin{bmatrix} 4v_2 = -4v_1 \\ 4v_1 = -4v_2 \end{bmatrix} \quad \begin{bmatrix} v_2 = -v_1 \\ v_1 = -v_2 \end{bmatrix} \quad v = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Check if correct

$$\boxed{A\,v = \lambda v}$$

$$\begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix} v = \lambda v \qquad \leftarrow \qquad \boxed{A = \begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix}}$$

$$\begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix} v = 8v \qquad \leftarrow \qquad \boxed{\boldsymbol{\lambda_1} = 8}$$

$$\begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 8 \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \leftarrow \qquad \boxed{v = \begin{bmatrix} 1 \\ 1 \end{bmatrix}}$$

$$\begin{bmatrix} 4+4 \\ 4+4 \end{bmatrix} = \begin{bmatrix} 8 \\ 8 \end{bmatrix} \qquad \boxed{\begin{bmatrix} 8 \\ 8 \end{bmatrix} = \begin{bmatrix} 8 \\ 8 \end{bmatrix}}$$

Check if correct

$$\boxed{A\,v = \lambda v}$$

$$\begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix} v = \lambda v \leftarrow \boxed{A = \begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix}}$$

$$\begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix} v = 0v \leftarrow \boxed{\boldsymbol{\lambda_2} = 0}$$

$$\begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 0 \begin{bmatrix} 1 \\ -1 \end{bmatrix} \leftarrow \boxed{v = \begin{bmatrix} 1 \\ -1 \end{bmatrix}}$$

$$\begin{bmatrix} 4 - 4 \\ 4 - 4 \end{bmatrix} = 0 \qquad \boxed{\begin{bmatrix} 0 \\ 0 \end{bmatrix} = 0}$$

Project the Data Along the Principal Component

$$\boxed{\lambda_1 = 8}$$

$$\boxed{\lambda_2 = 0}$$

$$\boxed{v = \begin{bmatrix} 1 \\ 1 \end{bmatrix}}$$

$$\boxed{v = \begin{bmatrix} 1 \\ -1 \end{bmatrix}}$$

We can see that our first principal component has a larger eigenvalue that the second principal component, we will use this because it has the highest maximum variance

Project the Data Along the Principal Component

$$v_{normal} = \frac{v}{\|v\|}$$

$$\|v\| = \sqrt{(v_1)^2+(v_2)^2}$$

$$\|v\| = \sqrt{1^2+1^2} = \sqrt{2}$$

$$\lambda_1 = 8$$

$$v = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$v_{normal} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

We need to normalize our eigenvectors so that they only represent directions not magnitude.

Project the Data Along the Principal Component

$$\boxed{new\ projection = A_{centered} \times v_{normal}}$$

This is the formula to project our data along the axis of our principal component

$$A_{centered} = \begin{bmatrix} -2 & -2 \\ 0 & 0 \\ 2 & 2 \end{bmatrix}$$

$$\boxed{v_{normal} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}}$$

$$new\ projection = \begin{bmatrix} -2 & -2 \\ 0 & 0 \\ 2 & 2 \end{bmatrix} \times \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Project the Data Along the Principal Component

$$new\ projection = \begin{bmatrix} -2 & -2 \\ 0 & 0 \\ 2 & 2 \end{bmatrix} \times \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$firstpoint = \frac{-2-2}{\sqrt{2}} = \frac{-4}{\sqrt{2}} = -2\sqrt{2}$$

$$secondpoint = 0$$

$$thirdpoint = \frac{2+2}{\sqrt{2}} = \frac{4}{\sqrt{2}} = 2\sqrt{2}$$

$$\boxed{new\ projection = \begin{bmatrix} -2\sqrt{2} \\ 0 \\ 2\sqrt{2} \end{bmatrix}}$$

Project the Data Along the Principal Component

$$data\ in\ 2D = \begin{bmatrix} 2 & 4 \\ 4 & 6 \\ 6 & 8 \end{bmatrix}$$

$$\boxed{new\ projection = \begin{bmatrix} -2\sqrt{2} \\ 0 \\ 2\sqrt{2} \end{bmatrix}}$$

This is the the 1D representation of our original data