Principle Component Analysis Example Question

Question:

Given the following data, use PCA to reduce dimension from 2 to 1.

Feature	**Example 1**	**Example2**	**Example3**	**Example4**
x	**4**	**8**	**13**	**7**
Y	**11**	**4**	**5**	**14**

Solution:

In the dataset above we have 2 dimensions x and y,our task is to reduce them to 1 using PCA.

Step 1: Dataset

Feature	**Example 1**	**Example2**	**Example3**	**Example4**
x	**4**	**8**	**13**	**7**
Y	**11**	**4**	**5**	**14**

No of features, n = 2 (x, y)

No. of samples, N = 4 (Example1,Example2,Example3,Example4)

Step2: Computation of mean of variables.

$\bar{x} = (4 + 8 + 13 + 7) \div 4 = 8$

mean of x is 8

$\bar{y} = (11 + 4 + 5 + 14) \div 4 = 8.5$

mean of y is 8.5

Step3: Computation of Covariance matrix

1. Get the Ordered pairs:
   (x, x) , (x ,y) , (y , x) , (y, y)

Ordered pairs are all combinations of variables used to compute entries in the covariance matrix.

Formula for getting no. of ordered pairs is (n^2) , 'n' representing the number of variables. Since here we have 2 variables x , y so the ordered pairs will be 4.

2. Covariance of these ordered pairs:
   Cov(x , x)

   Formula if the values are similar:

$$\mathbf{1/(N-1)} \times \sum_{i=1}^{n} (x_i - \bar{x})^2$$

i) **Cov(x , x)** =
   $1/(4-1)\ [(4-8)^2 + (8-8)^2 + (13-8)^2 + (7-8)^2]$
   $=14$

Formula if the values are not similar:

$$\mathbf{1/(N-1)} \times \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

ii) **Cov(x ,y)** =
   $1/(4-1)\ [(4-8)(11-8.5) + (8-8)(4-8.5) + (13-8)(5-8.5) + (7-8)(14-8.5)]$
   $= -11$

iii) **Cov(y , x)** = Cov(x , y) = -11

iv) **Cov(y , y)** =

   $1/(4-1)\ [(11-8.5)^2 + (4-8.5)^2 + (5-8.5)^2 + (14-8.5)^2]$
   $=23$

3. Covariance matrix n*n

$$S = \begin{bmatrix} cov(x, x) & cov(x, y) \\ cov(y, x) & cov(y, y) \end{bmatrix}$$

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

Step4: Eigen value, Eigen vector, Normalized eigen vector

i) **Eigen value**

Formula: determinant(S – λI) = 0

det(S – λI) = 0

S = Covariance matrix

λ = Eigen value

I = Identity matrix

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

(λI)

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

S – λI (S – λI)

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

$$\det * \left(\begin{bmatrix} 14-\lambda & -11 \\ -11 & 23-\lambda \end{bmatrix} \right) = 0$$

To find the determinant, we take 14- λ into 23- λ ,**minus** -11 into -11 . Meaning ;

(14- λ) *(23- λ) – (-11) * (-11) = 0

The output is a quadratic equation;

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

We then find the root of this quadratic equation

Formula: $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$ a=1,b=-37,c=201

λ = 30.3849, 6.6151

λ> λ1
λ=30.3849 →First principle component since it's the greatest value.
λ1= 6.6151

ii) Eigen vector of λ1

Formula: (S – λI)u1 = 0 U1=Eigen vector

[14- λ -11] * [U1] = 0
[-11 23- λ] [U2]

(14-λ)U1 + (-11)U2] = [0]
[(-11)U1 + (23-λ)U2] [0]

2 linear equations are formed:

(14 -λ)U1 - 11U2 = 0
(-11)U1 + (23-λ)U2 = 0

Next we find out the value of U1 and U2 from any of this linear equations;
I will consider (14 -λ)U1 - 11U2 = 0 linear equation,

U1/11 = U2/(14-λ) = t
When t is equal to 1 (which acts as a scaling factor),
U1 = 11
U2 = 14-λ

Eigen vector u1 of λ = [11]
[14-λ]

= [11]
[14 - 30.3849]

=[11]
[-16.3849]

iii)Normalize the eigen vector u1
This means to scale it so that its length (magnitude) is 1.
To compute the magnitude ;

$$\sqrt{(11)2+(-16.3849)2}$$

Divide each component by the magnitude

11/$\sqrt{(11)2+(-16.3849)2}$

-16.3849/$\sqrt{(11)2+(-16.3849)2}$

e1 = [0.5574]
[-0.8303]

λ 1 e2 = [0.8303]
[0.5574]

Step5: Derive new dataset

First principle component PC1	Example1	Example2	Example3	Example4
	P11	**P12**	**P13**	**P14**

P11 = e1 [4-8] Remember,4 is the first x value in the original dataset ,we
[11-8.5] are subtracting its mean which is 8 vise versa as to Y.
We will multiply their output with the eigen vector e1

=[0.5574 -0.8303] [-4]
[2.5]

= -4.3052

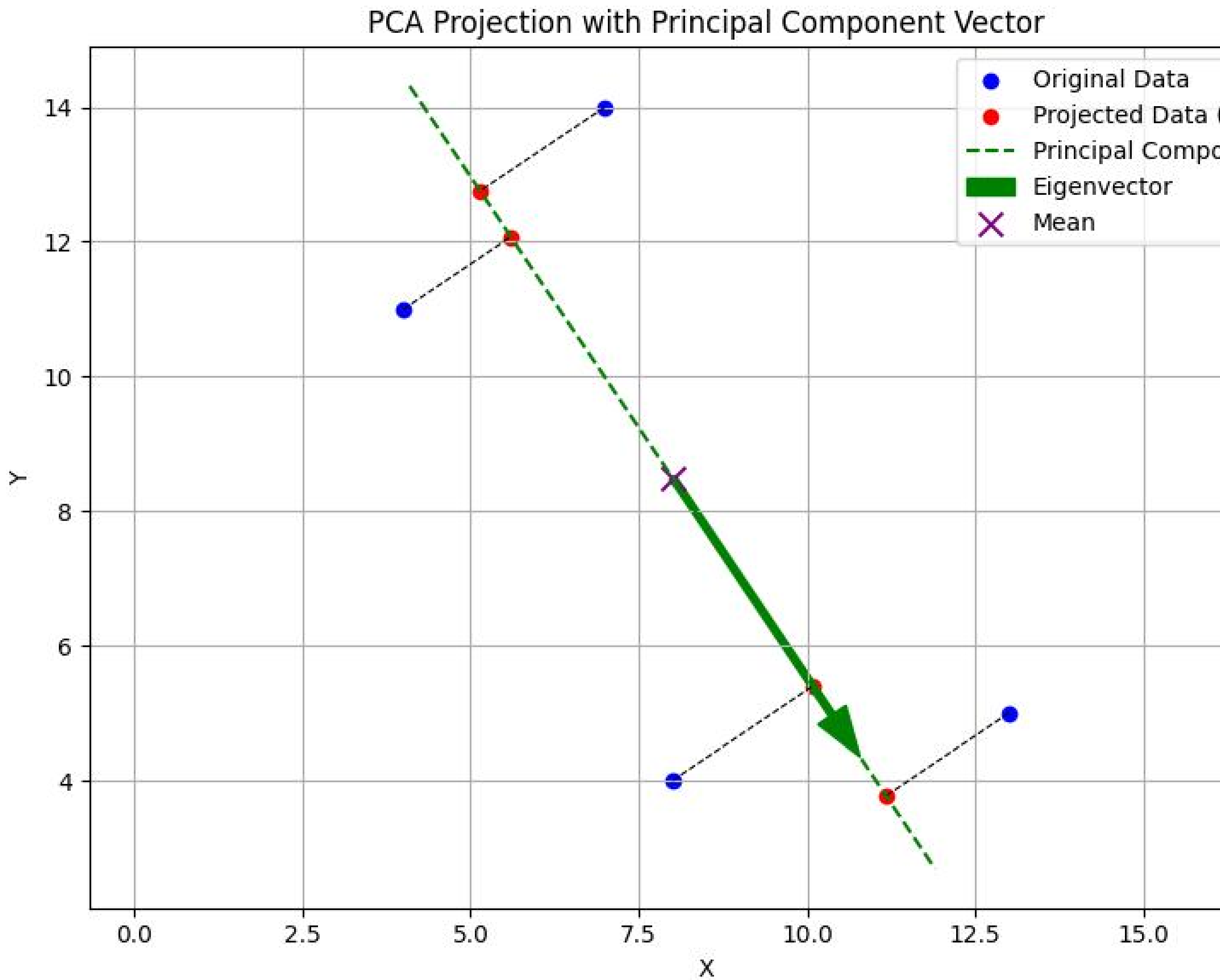
P12= [0.5574 -0.8303] [8-8]
[4-8.5]

= 3.7361

P13 = 5.6028
P14 = - 5.1238

	Example1	Example2	Example3	Example4
PC1	-4.3052	3.7361	5.6028	5.1238

The table above is a one dimension dataset.


PCA Projection with Principal Component Vector
Original Data
Projected Data (
Principal Compo
Eigenvector
Mean
Y
X
14
12
10
8
6
4
0.0
2.5
5.0
7.5
10.0
12.5
15.0