Principal Component Analysis

Dr. SINGARA SINGH KASANA

ASSOCIATE PROFESSOR, CSED

TIET, PATIALA

Dataset

X	Y
4	11
8	4
13	5
7	14

Number of features =2
Number of samples =4
Mean of X = 8
Mean of Y = 8.5

Computation of covariance matrix:

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

$cov_{x,y}$ = covariance between variable a and y

x_i = data value of x

y_i = data value of y

$\bar{x}$ = mean of x

$\bar{y}$ = mean of y

N = number of data values

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

Similarly we can get
Cov (X,Y) = [1/ (4-1)]x [(4-8)(11-8.5) + (8-8)(4-8.5) + (13-8)(5-8.5) + (7-8)(14-8.5)] = -11

Cov(Y, Y) = [1/ (4-1)]x [$(11-8.5)^2$ + $(4-8.5)^2$ + $(5-8.5)^2$ + $(14-8.5)^2$] = 23

Covariance matrix

$$A = \begin{bmatrix} Cov\,(x,x) & Cov\,(x,y) \\ Cov\,(y,x) & Cov\,(y,y) \end{bmatrix}_{2\times2}$$

$$A = \begin{bmatrix} 14 & -11 \\ -11 & 23 \end{bmatrix}_{2\times2}$$

Eigen value

for this solve the eqn

$$det\,(A - \lambda I) = 0$$

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

By expanding this eqn we get

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

$$\boxed{\lambda = 30.3849,\ 6.6151}$$

$$(A - \lambda_1 I) \, U_1 = 0$$

Here U_1 is eigen vector for λ_1

$$\begin{bmatrix} 14 - \lambda_1 & -11 \\ -11 & 23 - \lambda_1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\left. \begin{array}{c} (14 - \lambda_1) u_1 - 11 u_2 = 0 \\ -11 u_1 + (23 - \lambda_1) u_2 = 0 \end{array} \right\}$$

from first eg^n

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

Let $t = 1$

$$u_1 = 11$$

$$u_2 = 14 - \lambda_1 = 14 - 30.3849 = -16.3849$$

To normalize this eigen vector.

$$length = \sqrt{(11)^2 + (16.3849)^2}$$

$$e_1 = \begin{bmatrix} 0.5574 \\ -0.8303 \end{bmatrix}$$

for second eigen vector

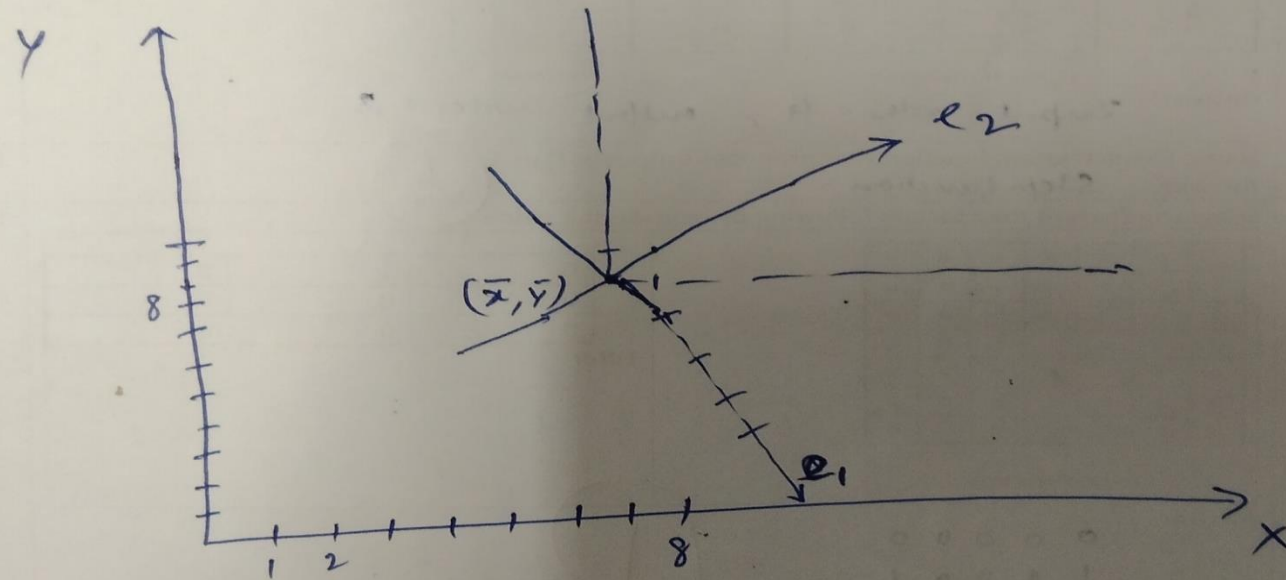$$e_2 = \begin{bmatrix} 0.8303 \\ 0.5574 \end{bmatrix}$$

· Drive new data set : —

$$P_{11} = e_1^T \begin{bmatrix} 4 - 8 \\ 18 - 8.5 \end{bmatrix}$$

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

$$= -2.2296 - 2.07575$$

$$= -4.3052$$

PC1

−4.3052
3.7361
5.6928
−5.1238

Singular Value Decomposition

Matrix A is given

3	2	2
2	3	-2

Thanks