Principal Component Analysis :- (PCA)

ex:

F	1	2	3	4
X_1	4	8	13	7
X_2	11	4	5	14

① Calculate mean of each feature

$\mu_{X_1} = 8$

$\mu_{X_2} = 8.5$

② Calculation of the covariance Matrix.

2 features H are

so 2×2 Matrix

$$S = \begin{bmatrix} cov(X_1X_1) & cov(X_1X_2) \\ cov(X_2X_1) & cov(X_2X_2) \end{bmatrix}$$

$$cov(X_1X_1) = \frac{1}{N-1}\sum_{K=1}^{N}(X_{1K} - \mu_{X_1})(X_{1K} - \mu_{X_1}) = 14$$

$$cov(X_1X_2) = \frac{1}{N-1}\sum_{1}^{N}(X_{1K} - \mu_{X_1})(X_{2K} - \mu_{X_2})$$

$$= \frac{1}{3}\sum_{1}^{4}(X_{1K} - 8)(X_{2K} - 8.5)$$

$$= \frac{1}{3}[(4-8)(11-8.5) + (8-8)(0) + (13-8)(5-8.5) + (7-8)(14-8.5)]$$

$$= -11$$

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

③ Eigenvalues of the covariance Matrix

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

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

$(14-\lambda)(23-\lambda) - 121 = 0$

$\lambda_1 = 30.3849$

$\lambda_2 = 6.6151$

(4) computation of EigenVectors.

$$\begin{bmatrix}0\\0\end{bmatrix} = (S-\lambda I)U \Rightarrow \begin{bmatrix}0\\0\end{bmatrix} = \begin{bmatrix}14-\lambda & -11\\-11 & 23-\lambda\end{bmatrix}\begin{bmatrix}U_1\\U_2\end{bmatrix}$$

$$(14-\lambda)U_1 - 11U_2 = 0$$

$$-11U_1 + (23-\lambda)U_2 = 0$$

$$(23-\lambda)U_2 = 11U_1$$

$$\frac{U_2}{11} = \frac{U_1}{23-\lambda} = t$$

$$U_1 = (23-\lambda)t$$
$$U_2 = 11t$$

$$\begin{bmatrix}23-\lambda\\11\end{bmatrix}$$

Note:
For one PC
select the Highest
eigen value.

$$U = \begin{bmatrix}23-30.3849\\11\end{bmatrix} = \begin{bmatrix}-7.3849\\11\end{bmatrix}$$

$$\|U\| = [(-7.3849)^2 + 11^2]^{1/2}$$

$$= 13.249$$

unit Eigen vector $= \frac{1}{13.249}\begin{bmatrix}-7.3849\\11\end{bmatrix} = \begin{bmatrix}-0.557\\0.830\end{bmatrix}$

$$e_1 = \begin{bmatrix}-0.557\\0.830\end{bmatrix}$$

By $\lambda = \lambda_2 = 6.615$
you get e_2 [2nd PC]

(5) computation of First PC.

$$e_1^T\begin{bmatrix}X_{1K} - \bar{X}_1\\X_{2K} - \bar{X}_2\end{bmatrix} = [-0.557 \quad 0.830]\begin{bmatrix}4-8\\11-8.5\end{bmatrix}$$

$x_1 = 4$
$x_2 = 11$

$$= [-0.557 \quad 0.830]\begin{bmatrix}-4\\2.5\end{bmatrix}$$

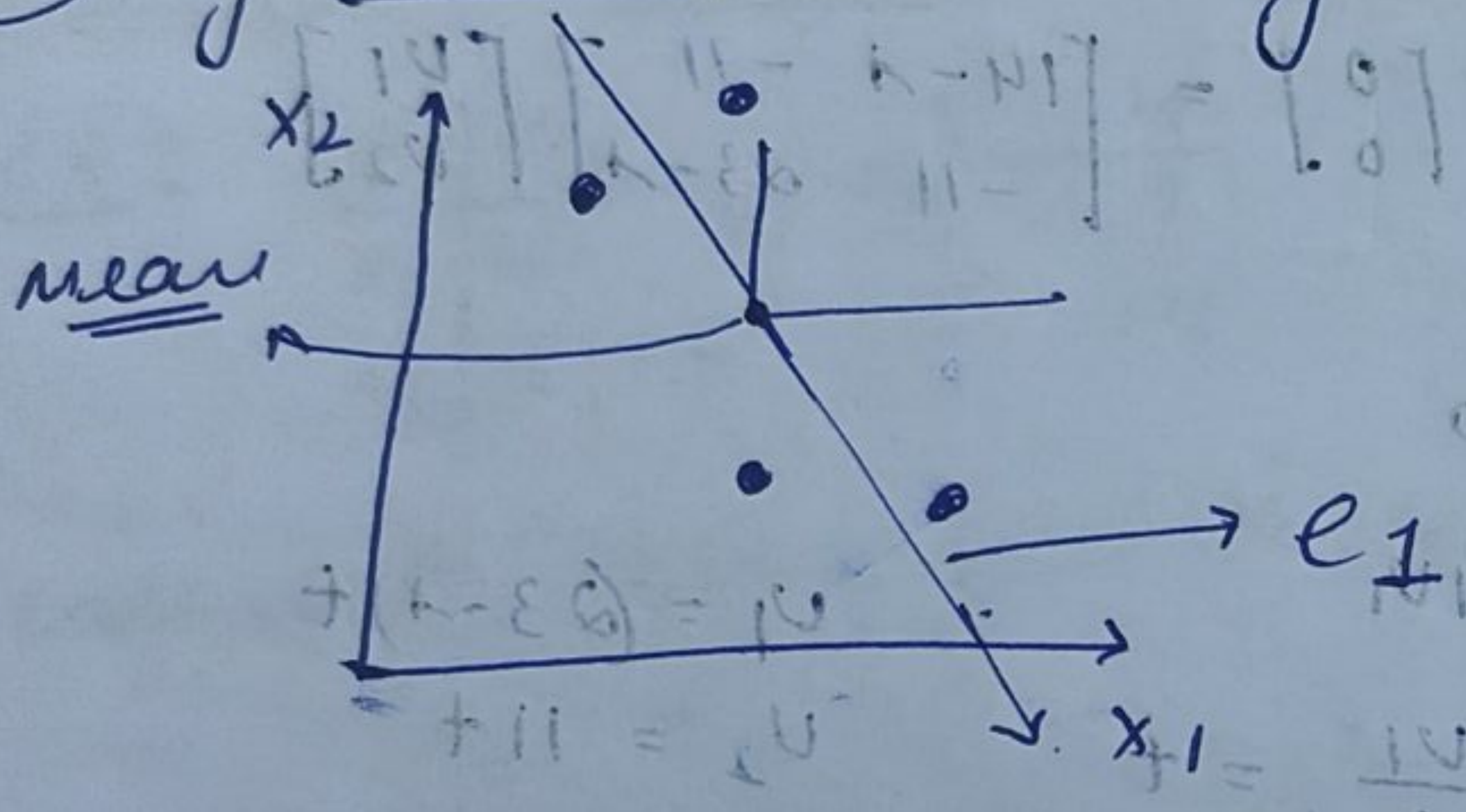
$$= -4.303$$

Note: just Take the 1st eqn always

(8)

5254 6254 7254 10254 11025 12254

(6) geometrical meaning



LDA Linear Discriminant Analysis:

Que: 2 classes

$\omega_1 = \{(4,2)\ (2,4)\ (2,3)\ (3,6)\ (4,4)\}$

$S_2 = \sum_{x \in \omega}$

(3) calc

$S_w =$

Chi-square Test χ^2 :-

Finding the critical value :-

1. Identify α, find $DOF = n-1$
2. For Right Tailed, (DOF, α)

   For left Tailed $(DOF, 1-\alpha)$

   Two Tailed $(DOF, \frac{1}{2}\alpha)$ & $(DOF, 1-\frac{1}{2}\alpha)$

STEPS :

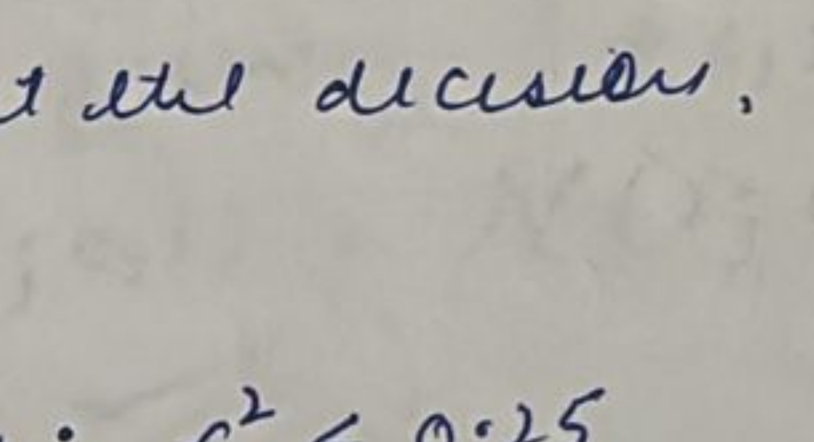
1. write H_0 & H_a
2. using α, & DOF find the critical values
3. Determine the rejection regions.
4. $$\boxed{\chi^2 = \frac{(n-1)s^2}{\sigma^2}}$$ s^2 → given, find the Test statistic, σ^2 → Hypothesised
5. If χ^2 is in the Rejection region, Then Reject H_0.
6. Interpret the decision.

EXAMPLE :

claim : $\sigma^2 \leq 0.25$

$H_0 = \sigma^2 \leq 0.25$

$H_a = \sigma^2 > 0.25$



Right Tailed Test.

$n = 41$

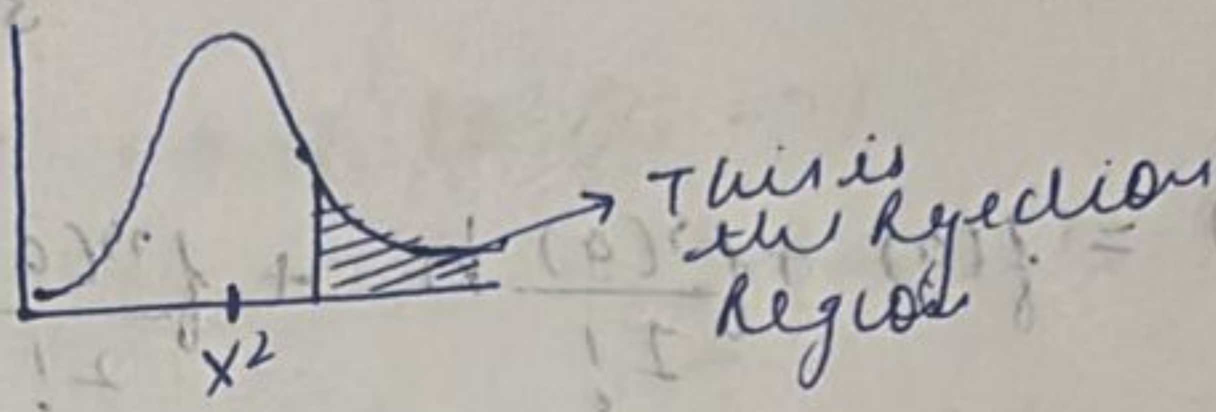
$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} \qquad \chi_0^2 = (40, 0.05)$$

$$= 55.758$$

$$= \frac{(41-1)(0.27)}{0.25}$$

$$= 43.2$$

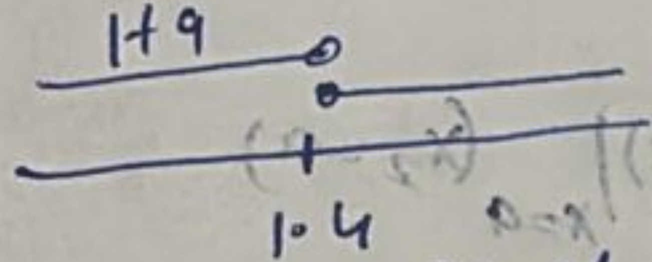
Rejection Region:



∴ Fail to Reject
Ho as χ^2 is Not in the
Rejection region.

(2) $H_a: \sigma < 1.4$ $\qquad \chi^2 = \frac{(25-1)(1.1)^2}{(1.4)^2} = \frac{24 \times 1.1 \times 1.1}{1.4 \times 1.4}$

$H_0: \sigma \geq 1.4$ $\qquad = 14.81$

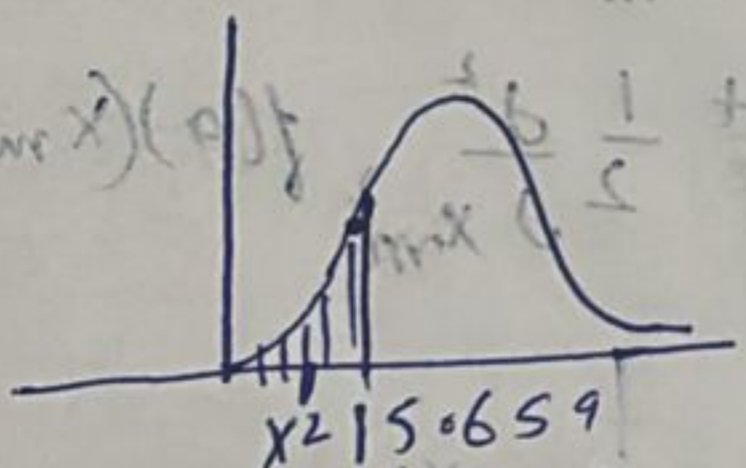


$$\chi_0^2 = (24, 1-0.1)$$

left tailed

$$= (24, 0.9)$$

$$= 15.659$$



∴ H_0 is Rejected

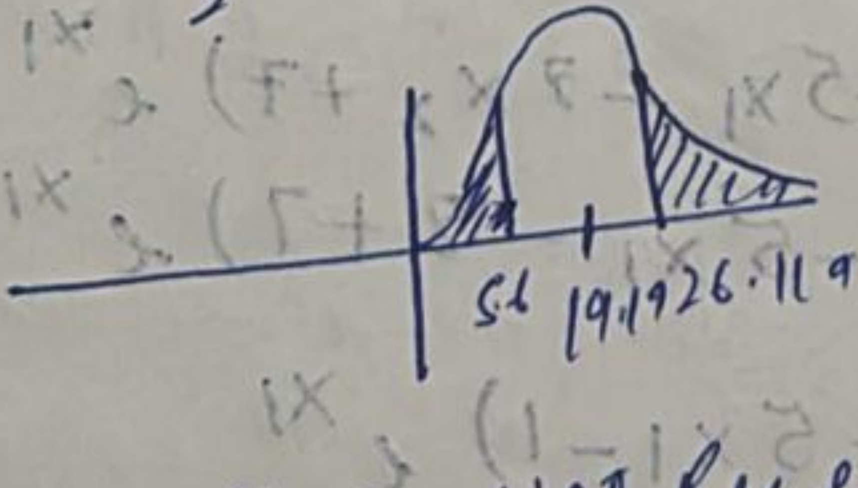
(3) claim : $\sigma^2 = 15.9$ Ho $\qquad \chi^2 = \frac{(15-1)\,21.8}{15.9}$

$\sigma^2 \neq 15.9$ Ha $\qquad = 19.19$

$\alpha = 0.05$

$\frac{\alpha}{2} = 0.025$ $\qquad \chi_0^2 = (DOF, 0.025)$

(DOF, 0.975)



5.629, 26.119

Ho is Not Rejected

Regression Analysis

$$Y = B_0 + B_1 X_1 + B_2 X_2 + \varepsilon$$

Y → outcome VAR

B_1 → coeff. to be estimated

X_2 → Predictor VAR.

ε → Random error

Regression Analysis → Estimating values of Y based on diff values of X

* $h_\theta(x) \rightarrow y$

Training set
↓
Learning Algo
↓
size of house → $\boxed{h}$ → Estimated Price

$h_\theta(x) = -20 + 0.25x$

e.g. $\begin{bmatrix} \text{Data} \\ \text{Matrix} \end{bmatrix} \begin{bmatrix} -20 \\ .25 \end{bmatrix} = h_\theta(x)$

cost funct: $J(\theta_0, \theta_1)$

$$= \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^i) - y^i)^2$$

Ques: Find Best fit line through (1 1) (2 3) (3 3) (4 5) using least square.

$Y = x_1 + m x_2$

$Y = \begin{bmatrix} 1 \\ 3 \\ 3 \\ 5 \end{bmatrix}$ $A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}$ $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$\boxed{A^T A \cdot X = A^T Y}$ $\begin{bmatrix} 4 & 10 \\ 10 & 30 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 12 \\ 36 \end{bmatrix}$

B. $f_\theta(x) = \sum_{j=0}^{1} \theta_j x_j$ $\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{bmatrix}$ $Y = \begin{bmatrix} 5 \\ 25 \\ 22 \\ 18 \end{bmatrix}$

Data Matrix D:

** suppose a val for θ

Ques: Build a linear Prediction Model :-
wind speed PPL ENR Req.

$\Rightarrow f_\theta(x) = \sum_{j=1}^{d} \theta_j x_j$ Data Matrix = [Freq wind speed PPL]

Now, [Data Matrix] [X] = [Y] Y = [Energy Req]

** Ques: $A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{bmatrix}$ $Y = \begin{bmatrix} 1 \\ 1 \\ 2 \\ 2 \\ 4 \end{bmatrix}$ $A^TAX = A^TY$

$A^TY = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 2 \\ 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 10 \\ 37 \end{bmatrix}$

1+2+6+8+20

$A^TA = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{bmatrix} = \begin{bmatrix} 5 & 15 \\ 15 & 55 \end{bmatrix}$

2×5 5×2

1+4+9+16+25
55

$\begin{bmatrix} 5x_1 + 15x_2 \\ 15x_1 + 55x_2 \end{bmatrix} = \begin{bmatrix} 10 \\ 37 \end{bmatrix}$ $\begin{bmatrix} 5 & 15 \\ 15 & 55 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 10 \\ 37 \end{bmatrix}$

$x_1 + 3x_2 = 2$

$x_1 = 2 - 3x_2$

$15x_1 + 55x_2 = 37 \Rightarrow 15(2 - 3x_2) + 55x_2 = 37$

$\Rightarrow 30 - 45x_2 + 55x_2 = 37$

$10x_2 = 7$

$x_2 = 0.7$

$x_1 = 2 - 2.1 = -0.1$

$\boxed{X = \begin{bmatrix} -0.1 \\ 0.7 \end{bmatrix}}$

2nd Method :-

Prediction Eq^n = $\hat{Y} = \boxed{\hat{B}_0 + \hat{B}_1 x}$

$$B_1 = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}$$

Note: Mean all rounded off

$B_0 = \bar{Y} - B_1 \bar{X}$

$\bar{X}$ → mean of x values

$\bar{Y}$ → Mean of Y values

Interpretation: slope (B_1): sales volume (y) is expected to inc by 0.7 by each Rs 1 invested in advertising.

B_0: Avg value of sales is -0.1 when advertising is 0.

→ Diff to explain

→ Expect some sales without advertising

calculating SSE : $\Sigma (Y_i - \hat{Y}_i)^2$

* Y_i → given values of Y
* $\hat{Y}$ → calc for each by $\hat{Y} = B_0 + B_1 \hat{x}$

Ques: $T = at + b$

t	0.5	1.1	1.5	2.1	2.3
T	32	33	34.2	35.1	35.7

$$A = \begin{bmatrix} 1 & 0.5 \\ 1 & 1.1 \\ 1 & 1.5 \\ 1 & 2.1 \\ 1 & 2.3 \end{bmatrix}$$

$$A^T A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0.5 & 1.1 & 1.5 & 2.1 & 2.3 \end{bmatrix} \begin{bmatrix} 1 & 0.5 \\ 1 & 1.1 \\ 1 & 1.5 \\ 1 & 2.1 \\ 1 & 2.3 \end{bmatrix} = \begin{bmatrix} 5 & 7.5 \\ 7.5 & 13.41 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0.5 & 1.1 & 1.5 & 2.1 & 2.3 \end{bmatrix} \begin{bmatrix} 32 \\ 33 \\ 34.2 \\ 35.1 \\ 35.7 \end{bmatrix} = \begin{bmatrix} 170 \\ 259.43 \end{bmatrix}$$

$A^T A X = A^T Y$

$$\begin{bmatrix} 5 & 7.5 \\ 7.5 & 13.41 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 170 \\ 260 \end{bmatrix}$$

$$x = \begin{bmatrix} 30.52 \\ 2.32 \end{bmatrix} = \begin{bmatrix} b \\ a \end{bmatrix}$$

$5x_1 + 7.5x_2 = 170$

$\Rightarrow x_1 + 1.5x_2 = 34$

$x_1 = 34 - 1.5x_2$

$7.5(34 - 1.5x_2) + 13.4x_2 = 260$

$\Rightarrow 255 - 11.25x_2 + 13.4x_2 = 260$

$2.15x_2 = 5$

$\underline{x_2 = 2.32}$

$\underline{\underline{x_1 = 30.52}}$

Matrix Representation

$\Rightarrow T = A \cdot X$

OR

$$T = \begin{bmatrix} \text{Data} \\ \text{Matrix} \end{bmatrix} \begin{bmatrix} \theta \end{bmatrix}$$