UNIT - IV

Dimension Reduction.

* Principle Component Analysis:-

	1	2	3	4
	Ex1	Ex2	Ex3	Ex4
X1	4	8	13	7
X2	11	4	5	14

* 8 datas points Convert in to Some Small values like 4.
* Which One is important data Point, have to be Noticed and Solve the value.
* It is unsupervised learning.

Step 1:- FIND Mean

$$\text{Mean} = \frac{\text{Sum of all Observation}}{\text{Total No of Obs.}}$$

$$\overline{X_1} = (4+8+13+7)/4 = 8$$

$$\overline{X_2} = (11+4+5+14)/4 = 8.5$$

Step2:- Find Co-variance Matrix. (S)

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

variable.

$$Cov(x_1, x_1) = \frac{1}{N-1} \sum_{K=1}^{N} (x_{1K} - \overline{x}_1)^2 \quad \therefore K = 1, 2, 3, 4$$

$N \rightarrow$ No of obs.

$\cancel{Cov(x_1, x_1) = \frac{1}{4-1} \sum \; \frac{1}{4-1}}$

$$Cov(x_1, x_1) = \frac{1}{4-1} \sum_{K=1}^{4} (x_{1K} - \bar{x}_1)^2$$

$$= \frac{1}{3} \sum_{K=1}^{4} (x_{1K} - \bar{x}_1)^2$$

$$= \frac{1}{3} \left[(x_{11} - \bar{x}_1)^2 + (x_{12} - \bar{x}_1)^2 + (x_{13} - \bar{x}_1)^2 + (x_{14} - \bar{x}_1)^2 \right]$$

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

$$= \frac{1}{3} \left[16 + 0 + 25 + 1 \right] = 14.$$

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

$$= \frac{1}{4-1} \left[(x_{11} - \bar{x}_1)(x_{21} - \bar{x}_2) + (x_{12} - \bar{x}_1)(x_{22} - \bar{x}_2) + (x_{13} - \bar{x}_1)(x_{23} - \bar{x}_2) + (x_{14} - \bar{x}_1)(x_{24} - \bar{x}_2) \right]$$

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

$$= \frac{1}{3} \left[-33 \right] = -11$$

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

$$= \frac{1}{4-1} \left[(x_{21} - \bar{x}_2)^2 + (x_{22} - \bar{x}_2)^2 + (x_{23} - \bar{x}_2)^2 + (x_{24} - \bar{x}_2)^2 \right]$$

$$= \frac{1}{3} \left[6.25 + 20.25 + 12.25 + 30.25 \right].$$

$$= 23$$

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

<u>Step 3:-</u> <u>Eigen values</u>.

$$\boxed{0 = \det(S - \lambda I)}$$ formula.

I = Identity matrix

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

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

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

$$\det(S - \lambda I) = (14-\lambda)(23-\lambda) - (-11)(-11)$$

$\therefore \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$ Determinant

$$= 322 - 14\lambda - 23\lambda - 121 + \lambda^2$$

main ↓ becomes

$U_1 = \Gamma_{11}$ ⌉

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

$$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \begin{matrix} \lambda_1 = 30.38 \\ \lambda_2 = 6.6151 \end{matrix}$$

Step 4:- Eigen vector's

$$U = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix}$$

$$\boxed{\begin{bmatrix} 0 \\ 0 \end{bmatrix} = (S - \lambda I)\, U}$$ formula.

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 14-\lambda & -11 \\ -11 & 23-\lambda \end{bmatrix} \begin{bmatrix} U_1 \\ U_2 \end{bmatrix}$$

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} (14-\lambda)U_1 - 11U_2 \\ -11U_1 + (23-\lambda)U_2 \end{bmatrix}$$

$$(14-\lambda)U_1 - 11U_2 = 0 \rightarrow (1)$$

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

$$\frac{U_1}{11} = \frac{U_2}{14-\lambda}$$

$$\frac{U_1}{11} = \frac{U_2}{14-\lambda} = t$$

↑ Relating with some variable

So ∴ $t = \frac{U_1}{11}$

$t = \frac{U_2}{14-\lambda}$

$$U_1 = 11t \quad ; \quad U_2 = (14-\lambda)t$$

main ↓ becomes

$$U_1 = \begin{bmatrix} 11 \\ (14-\lambda) \end{bmatrix}$$

$U_1 = 11t$; $U_2 = (14-\lambda)t$

Where $t = 1$

To find PCA.

Consider Larger λ (eigen value)

$\lambda_1 = 30.3849$.

$$U_1 = \begin{bmatrix} 11 \\ 14-30.3849 \end{bmatrix} = \begin{bmatrix} 11 \\ 14-\lambda_1 \end{bmatrix}$$

↑ already found.

<u>Step 5:- Unit Eigen Vector (e_1)</u>

$$e_1 = \frac{U_1}{|U_1|}$$

$|U_1|$ ← length of U_1

To find $|U_1|$, <u>W.K.T</u> of Euclidian distance.

$$\boxed{= \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}}$$ formula.

W.K.T $U_1 = \begin{bmatrix} 11 \\ 14-\lambda_1 \end{bmatrix}$ → Consider x, → " y and Consider Co-ordinates (0,0).

$$|U_1| = \sqrt{(11-0)^2 + ((14-\lambda_1)-0)^2}$$

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

$$e_1 = \begin{bmatrix} 11/19.7348 \\ (14-\lambda_1)/19.7348 \end{bmatrix} = \begin{bmatrix} 0.5574 \\ -0.8303 \end{bmatrix}$$

↓ (30.384)

Cont:

Same step Instead of λ_1 Replace λ_2.

// $e_2 = \begin{bmatrix} 0.8303 \\ 0.5574 \end{bmatrix}$ Not Required here //.

<u>Step:-6</u> first Principle Component.

$$e_1^T \begin{bmatrix} x_{1K} - \bar{x}_1 \\ x_{2K} - \bar{x}_2 \end{bmatrix}$$ formula.

$$= \begin{bmatrix} 0.5574 & -0.8303 \end{bmatrix} \begin{bmatrix} x_{11} - \bar{x}_1 \\ x_{21} - \bar{x}_2 \end{bmatrix}$$ If $K=1$

$$= \begin{bmatrix} 0.5574 & -0.8303 \end{bmatrix} \begin{bmatrix} 4 & -8 \\ 11 & -8.5 \end{bmatrix}$$

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

$$= 0.5574(-4) + (-0.8303)(2.5)$$

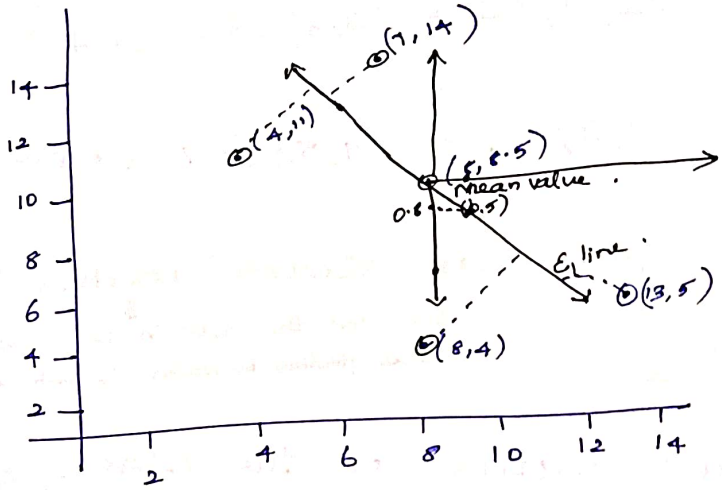
$$= -4.30535$$

	Ex_1	Ex_2	Ex_3	Ex_4
X_1	4	8	13	7
X_2	11	4	5	14
	-4.3052	3.7361	5.6928	-5.1238
	If $K=1$	If $K=2$	If $K=3$	If $K=4$

8 data Points Convert in to 4 data points.

to a Smaller
Class Discrimination.

Graph Representation



1) from mean point draw x, y axis.

2) W.K.T

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

3) Project all values in E_1 line.

— x —

Linear Discriminant Analysis:- (LDA)

Subdivision CA

* LDA is a Dimensionality Reduction technique used as a pre-processing step for pattern-Classification and ML applications.

* [Dimension means features of the data, When we have large No of i/p with us, Each data have large No of feature in it. So ML Not possible to tackle it. So we need to reduce No of dimension].

* LDA Similar to PCA, but LDA in addition finds the axes that maximizes the Seperation between multiple Classes.
(Only reduce dimension, Not Category (Class A, B.) to which belongs to.)

Goal:- To project feature space (N-dimensional data) on to a Smaller Subspace K ($K \leq n-1$) while maintaining the Class Discrimination.

Ex :-

2-D data.

$C_1 \Rightarrow x_1 = (x_1, x_2) = \{(4,1), (2,4), (2,3), (3,6), (4,4)\}$

$C_2 \Rightarrow x_2 = (x_1, x_2) = \{(9,10), (6,8), (9,5), (8,7), (10,8)\}$

Step 1 :- Compute With in class Scatter matrix (S_W)

(How well the data in the class C_1, C_2, that finding by mean of each Categorys)

$$S_W = S_1 + S_2$$

S_1 is the Co-variance matrix for the Class C_1 & S_2 is for C_2.

$$S_1 = \sum_{x \in C_1} (x - \mu_1)(x - \mu_1)^T$$

$(x - \mu_1)$ → Data Present in Class 1

μ_1 → mean of class C_1

Compute

$$\mu_1 = \left\{ \frac{4+2+2+3+4}{5}, \frac{1+4+3+6+4}{5} \right\}$$

$$\boxed{\mu_1 = [3.00, 3.60]}$$

|||ly $\boxed{\mu_2 = [8.4, 7.60]}$.

$$[x_1 - \mu_1] = \begin{array}{c} \begin{array}{ccccc} x - \mu_1 & x - \mu_1 & x - \mu_1 & \cdot & \cdot \\ (4-3) & 2(-3) & (2-3) & & \end{array} \\ \begin{bmatrix} 1 & -1 & -1 & 0 & 1 \\ -2.6 & 0.4 & -0.6 & 2.4 & 0.4 \end{bmatrix} \\ \begin{array}{ccccc} (1-3.6) & (4-3.6) & (3-3.6) & (6-3.6) & (4-3.6) \\ x_2 - \mu_1 & x_2 - \mu_1 & x_2 & & \end{array} \end{array}$$

Now for Each x, we going to Calculate $(x - \mu_1)(x - \mu_1)^T$. So, We will have "5" Such matrices.

$$\begin{bmatrix} 1 \\ -2.6 \end{bmatrix} \begin{bmatrix} 1 & -2.6 \end{bmatrix} = \begin{bmatrix} 1 & -2.6 \\ -2.6 & 6.76 \end{bmatrix} \rightarrow (1)$$

↓ first matrix

|||ly for Rest.

$$\begin{bmatrix} -1 \\ 0.4 \end{bmatrix} \begin{bmatrix} -1 & 0.4 \end{bmatrix} = \begin{bmatrix} 1 & -0.4 \\ -0.4 & 0.16 \end{bmatrix} - (2)$$

$$\begin{bmatrix} -1 \\ -0.6 \end{bmatrix} \begin{bmatrix} -1 & -0.6 \end{bmatrix} = \begin{bmatrix} 1 & 0.6 \\ 0.6 & 0.36 \end{bmatrix} - (3)$$

$$\begin{bmatrix} 0 \\ 2.4 \end{bmatrix} \begin{bmatrix} 0 & 2.4 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 5.76 \end{bmatrix} - (4)$$

$$\begin{bmatrix} 1 \\ 0.4 \end{bmatrix} \begin{bmatrix} 1 & 0.4 \end{bmatrix} = \begin{bmatrix} 1 & 0.4 \\ 0.4 & 0.16 \end{bmatrix} - (5)$$

Adding (1) + (2) + (3) + (4) + (5) and take average, We will get Co-variance matrix S_1

$$S_1 = \begin{bmatrix} \overset{4/5 \rightarrow 1+1+1+0+1/5}{0.8} & -0.4 \\ -0.4 & 2.6 \end{bmatrix}$$

|||ly for Class 2.

$$S_2 = \begin{bmatrix} 1.84 & -0.04 \\ -0.04 & 2.64 \end{bmatrix}$$

$$S_W = S_1 + S_2$$

$$S_W = \begin{bmatrix} 2.64 & -0.44 \\ -0.44 & 5.28 \end{bmatrix}$$

Step 2:- Compute Between Scatter Class matrix (S_B).

$$S_B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T$$

$$= \begin{pmatrix} -5.4 \\ 4 \end{pmatrix} \begin{pmatrix} -5.4 & 4 \end{pmatrix} = \begin{pmatrix} 29.16 & 21.6 \\ 21.6 & 16.00 \end{pmatrix}$$

Step 3:- find Best LDA Projection Vector.

* ||ly like PCA find Eigen vectors.

(or) $\Rightarrow \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = S_W^{-1} (\mu_1 - \mu_2)$

$\Downarrow$

$$S_W^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Ans

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} 0.91 \\ 0.39 \end{bmatrix}$$

So W.K.T

$$S_W = \begin{bmatrix} 2.64 & -0.44 \\ -0.44 & 5.28 \end{bmatrix}$$

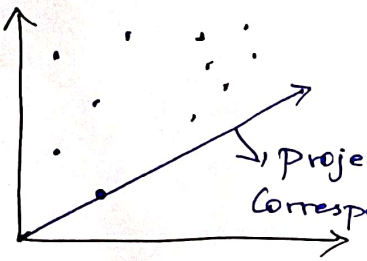
$$S_W^{-1} = \frac{1}{13.74} \begin{bmatrix} 5.28 & 0.44 \\ 0.44 & 2.64 \end{bmatrix} = \begin{bmatrix} 0.384 & 0.032 \\ 0.032 & 0.192 \end{bmatrix}$$

Step 4:- Dimension Reduction.

$y = W^T x \rightarrow$ I/p data Samples.

$\downarrow$

Projection vector



—x—

Least Square Method

* It is used to learn linear model for Regression and Classification.

* Regression problem is finding the fn Estimator

$$\hat{f} : x \rightarrow R \text{ from } (x_i, f(x_i))$$

Example

The diff between actual and Estimated fn values is called Residual (ϵ)

$$\epsilon_i = \underbrace{f(x_i)}_{\text{Actual value}} - \underbrace{\hat{f}(x_i)}_{\text{Estimated value}}$$

* The Sum of Squares of Residuals is to be minimized.
This is called LSM

(i.e) The difference must be low, → We can accurately Estimate the value.
If High → Estimation fail

* Suppose it is represented in the format

$$\sum_{i=1}^{n} \epsilon_i^2 \quad \text{[Sum of square of Residuals]}.$$

Univariate Linear Regression :

* Consider Two feature weight and height

W, h

W.K.T Slope Intercept = $\boxed{y = a + bx}$. x & y are Two points.

Slope

Intercept

* We have to find the Relation between W, & h

Cont.

* We are providing (n) values $\underbrace{[w_i, h_i]}_{\text{Original values}}$.

* Assume Univariant linear fn.

$$\boxed{\underset{\text{Estimated value}}{w} = a + bh}$$

from this equ we have to find a, b.

Such that some of square of residual to be minimized.

Convert equ

$$\sum_{i=1}^{n} \left(\underset{\text{Original}}{w_i} - \underset{\text{Estimated}}{(a+bh_i)}\right)^2 \rightarrow (1)$$

To get solution we moving on partial Derivatives with equal to '0'

$$\frac{\partial}{\partial a} \sum_{i=1}^{n} \left(w_i - \underset{\text{Constant}}{(a+bh_i)}\right)^2 = -2 \sum_{i=1}^{n} \left(\underset{\text{Contain set of values}}{w_i} - (a+bh_i)\right) = 0$$

$w_i - a - bh_i = 0$

$\boxed{w_i - bh_i = a.}$

Represented $\hat{f}$, $\parallel^{ly}$ $\hat{a}, \hat{b}$ $\Rightarrow$ $\boxed{\hat{a} = \bar{w} - \hat{b}\bar{h}}$

Intercept

Mean of all values Considering

Mean.

$$\frac{\partial}{\partial b} \sum_{i=1}^{n} \left(w_i - (a + \underbrace{bh_i}_{\text{hi along with b}})\right)^2 = -2 \sum_{i=1}^{n} \left(w_i - (a+bh_i)\right) h_i = 0$$

So include h_i

$$= -2 \sum_{i=1}^{n} \left(w_i - (\bar{w} - \hat{b}\bar{h} + bh_i)\right) h_i$$

Required answer is $\hat{b}$

$$\underset{\text{Slope/Regression Co-efficient}}{\hat{b}} = \sum_{i=1}^{n} \frac{(w_i - \bar{w})(h_i - \bar{h})}{(h_i - \bar{h})}$$

↳ B common term.

$W = \hat{a} + \hat{b}h$ This is our Solu.

Sub $\hat{a}$ and $\hat{b}$ in above equ.

Regression Co-efficient $\hat{b}$ is Ratio of $\left\{ \dfrac{\text{Covariance b/w } h \text{ \& } W}{\text{Variance } h} \right.$.

*. Assume feature x, target variable is y

Then what is $\hat{b}$?

$$\hat{b} = \frac{\cancel{n}\ \overset{\text{variance}}{\sigma_{xy}}}{\underset{n \text{ times}}{\cancel{n}}\ \sigma_x} = \frac{\sigma_{xy}}{\sigma_x}$$

- Suppose $\hat{a}$ the Regression line passes through mean $(\bar{x}, \bar{y})$, if we are adding Constant to all x values, it does not affect Regression Co-efficient $(\hat{b})$. Because all the values deviated from mean.

Suppose x_1, x_2, ⋮, x_n adding Constant to all values

mean Not going to Change.

To make calculation simple.

* for that Purpose we have to make Zero intercept.

The line passes through $\bar{x}, \bar{y}$

* Subtract $\bar{x}$ from each x_i, $\left\{ \begin{array}{l} \text{automatically Zero intercept becomes} \\ \text{when Subtract } \bar{x}. \end{array} \right.$

* The line passes through Zero instead of $\bar{x}$. (mean)

* The line passes through Zero.

- ✓ Suppose if replace x_i with $x_i' = \dfrac{x_i}{\sigma_{xx}}$ → all x_i values are derived by variance

Cont.

✓ Suppose Replace $\bar{x}$ with $\bar{x'} = \dfrac{\bar{x}}{\sigma_{xx}}$.

* feature in to normalized feature
* How mean & variance.

This is called Normalized feature $(x_i', \bar{x'})$ are.

* Take Covariance between Normalize feature $(x_i', \bar{x'})$ and Target value y

$$\hat{b} = \frac{1}{n} \sum_{i=1}^{n} \underbrace{\left(x_i' - \bar{x'}\right)}_{\text{Normalized feature}} \left(y_i - \bar{y}\right)$$

// $(w_i - \bar{w})$ feature $(h_i - \bar{h})$ feature //.

Sum of Residuals.

To get answer equate to zero.

from equ ① we wrote ↓

$\therefore$ W.K.T $w = a + bh$.

$y_i = a + bx_i$

$$\sum_{i=1}^{n} \left(y_i - (\hat{a} + \hat{b} x_i)\right)$$

$$= n\left(\underbrace{\bar{y}}_{\text{Sum of all terms}} - \hat{a} - \hat{b}\underbrace{\bar{x}}_{\text{Sum}}\right) = 0.$$

$$\boxed{\hat{a} = \bar{y} - \hat{b}\bar{x}}$$

Because of this Overfitting Possible. in LSM.

L
Reg

Multivariate Linear Regression

* If arbitary feature, we must choose matrix notation.

* How Univariate Converted in to matrix form?

W.K.T $y = a + bx$

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} a + \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} b + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix} \longrightarrow (1)$$

Target Value — Vector — feature — Scalar — Suppose if Contain Residual.

If we have d feature $x \longrightarrow n \times d$ matrix

$b \rightarrow$ (d- vector).

To Simplify the Solution generally using homogeneous Co-ordinates. (Different type of equ represented in Single format).

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ 1 & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

This is homogeneous (1) equ Converted in to to single format.

$$y = x^0 w + \epsilon$$

✓ $x^0 = n \times (d+1)$

$n \rightarrow$ Row, $(d+1) \rightarrow$ Columns, (*) The 1st Column is One, Remaing n Columns are Columns of x.

✓ $w = \begin{pmatrix} a \\ b \end{pmatrix}$ — (*) 1st Row Only Intercept, Remaining Row all Regression Co-efficient.

* To make Zero Intercept

find Co-variance bet every feature and Target.

Assume an Expression $\boxed{x^T y}$ $\xrightarrow[\text{J tis}]{}$ n vector, j^{th} entry in this Vector are Product of jth row of x^T.

$(x_{1j} \ldots x_{nj})$ with $(y_1 \ldots y_n)$

* Suppose

$$(x^T y)_j = \sum_{i=1}^{n} x^T_{ij} \cdot y_i$$

for j entry.

$\Downarrow$

$(x_{1j} \ldots x_{nj})$ with $(y_1 \ldots y_n)$.

If same writting in univariate $\Rightarrow \sum_{i=1}^{n} (x_{ij} - \mu_j)(y_i - \bar{y}) + n\mu_j\bar{y}$

μ_j: j^{th} column mean all

$n\mu_j\bar{y}$: due to matrix representation added. Mean of $(\mu_j \bar{y})$.

$$= n(\sigma_{jy} + \mu_j \bar{y})$$

Achieve Normalization by using matrix

* We are using a diagonal matrix with all entries of $\frac{1}{\sigma_{xy}}$; if j^{th} Column $\Rightarrow \frac{1}{\sigma_{jj}}$

* Assume $S = n \cdot \sigma_{jj}$

$S^{-1} = \frac{1}{n \cdot \sigma_{jj}}$ $\Leftarrow$ if scaling occurs

$\hat{w} = S^{-1} x^T \cdot y$ [normalization is applied for each & Every feature

$\Downarrow$

$= (x^T x)^{-1} x^T \cdot y$

* Given below are ..

= $x^T x$ is uncorelated feature

uncorrelated feature
⇓
If we are find the variance bet any pair of different features, the value is zero.

Regularized Regression:-

* If Regression model is unstable, (Chances of Overfitting the data).

* In that Situation, adding some more Constraints in to the weight, to avoid Overfitting.

* Also Pearform Shrinkage. (Avg of weights small)

* Regularized Problem in to optimization

$$w^* = \underset{w}{\operatorname{argmin}} (y - xw)^T (y - xw)$$

To make optimization

$$w^* = \underset{w}{\operatorname{argmin}} (y - xw)^T (y - xw) + \lambda |w|^2$$

λ — amount of Regularization

$|w|$ = Squared norm vector for w.

$$\boxed{\hat{w} = (x^T x + \lambda I)^{-1} \cdot x^T y}$$ ⇒ Ridge Regression L2 norm.

identity matrix

Lasso - Least absolute Shrinkage and Selection Operator. L1-norm.
Regression

Solution.

Problem using LSM

* Given below are the figures of production of a factory. fit a straight line trend & plot these fig on a graph and show the trend line.

X year	Y Profit.	x	x^2	xY	Yc
2010	60	-3	9	-180	61.42
2011	72	-2	4	-144	66.28
2012	75	-1	1	-75	71.14
→ 2013	65	0	0	0	76
2014	80	1	1	+80	80.86
2015	85	2	4	+170	85.72
2016.	95	3	9	+285	90.58.
	532				

Question.

$\boxed{Yc = a + bx}$ formula.

$$a = \frac{\Sigma y}{N} = \frac{532}{7} = 76 \; ; \quad b = \frac{\Sigma xy}{\Sigma x^2} = \frac{136}{28} = 4.86$$

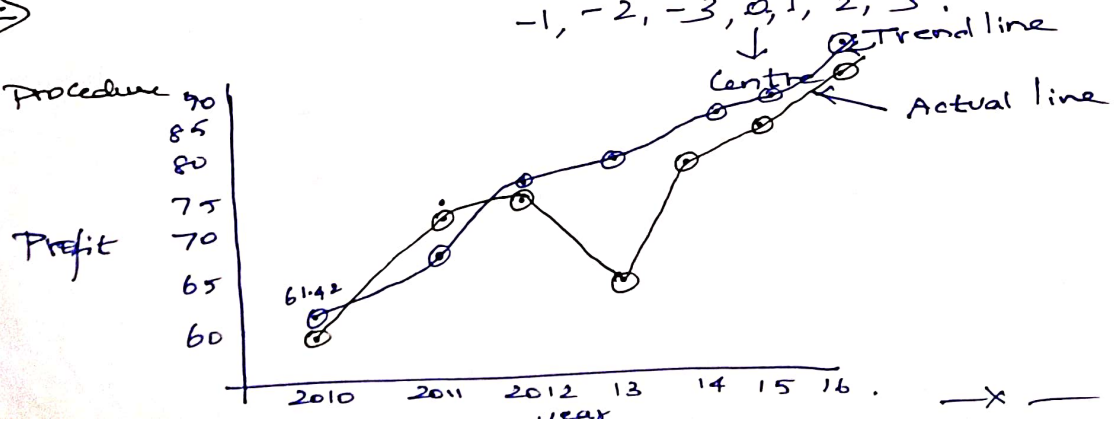
To find x, we have to calculate deviation.

* Count is 7 so odd ⇒ Deviation ⇒ from middle point make it as Zero, and strech both end

-1, -2, -3, 0, 1, 2, 3.

↓ Centre

* If Even → Different Procedure



REMARKS ON LAZY AND EAGER ALGORITHM:-

Lazy Learning:-

1) Simply Stores training data and waits until it gets a test tuple. (It just store memorize, wait for New Ex). So lazy

2) Less training time, more prediction Time.

3) Ex. All instance based learning algorithms.

Eager Learning:-

1) When we give a training Set, it Constructs a model for classification Before getting New Example. (It will Start define generalization before new Ex)

2) More Training time, less prediction Time

3) Ex Decision Tree, Naive Bayes, ANN etc.

— x —