Machine Learning Foundations

Week 4

Lecture - 1 : Linear & Polynomial Regression

Feature matrix (A) :

$$A = \begin{array}{cccc} & \text{Age} & \text{Education} & \text{Experience} \\ \left[\begin{array}{c} 01 \\ 02 \\ 03 \end{array}\right. & \begin{array}{c} 35 \\ 45 \\ 35 \end{array} & \begin{array}{c} 15 \\ 15 \\ 16 \end{array} & \left.\begin{array}{c} 3 \\ 8 \\ 5 \end{array}\right] \end{array}$$

$$Y = \begin{array}{c} \text{Income} \\ \begin{bmatrix} 100000 \\ 150000 \\ 200000 \end{bmatrix} \end{array} \quad \begin{array}{c} \hat{\text{Income}} \\ \begin{bmatrix} 100100 \\ 149000 \\ 201000 \end{bmatrix} \end{array}$$

$$x_1 = \begin{bmatrix} 35 \\ 15 \\ 3 \end{bmatrix} \quad x_2 = \begin{bmatrix} 45 \\ 15 \\ 8 \end{bmatrix} \quad x_3 = \begin{bmatrix} 35 \\ 16 \\ 5 \end{bmatrix} \quad Y = \begin{bmatrix} 100000 \\ 150000 \\ 200000 \end{bmatrix}$$

$e_1 = 100100 - 100000 = 100$

$e_2 = 1000$

$e_3 = 1000$

$E = 100^2 + 1000^2 + 1000^2$

$= 10000 + 1000000 + 1000000 = 1010\cancel{100}$

$MSE = \frac{1}{n} \times E$

$= \frac{1}{3} \times$

(predicted) $\hat{y} = \theta_0 + x_1\theta_1 + x_2\theta_2 + x_3\theta_3$ ← Model (Linear Reg)

$\frac{1}{2} \times E$

Estimate: $\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix}$ → parameters

$\text{Error}^2 = (\text{actual} - \text{predicted})^2 = (\text{predicted} - \text{actual})^2$

$e_i^2 = (\hat{y}_i - y_i)^2$

$E = \frac{1}{2} \sum_{i=1}^{n} e_i^2$

$a \cdot a = a^T a$
$= a^2$

$E = \frac{1}{2}(A\theta - Y)^T (A\theta - Y)$

$A\theta$ → predicted
Y → actual

$(A\theta - Y) \cdot (A\theta - Y)$

To find best θ E should be minimum

$\Rightarrow \nabla E = 0$

We did this derivation in week 4 live session

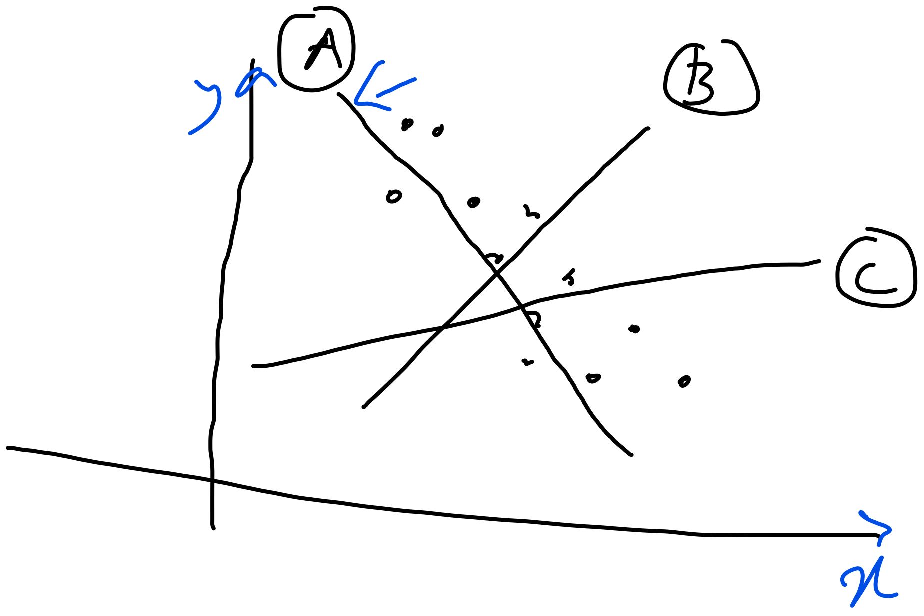
$A^T A\theta = A^T b$ $\qquad \boxed{A^T(A\theta - Y) = 0}$ $\qquad \nabla E = 0$

$\nabla E = 0$

$$A^T A\theta - A^T Y = 0$$

$$\underline{\theta} = (A^T A)^{-1} A^T Y$$

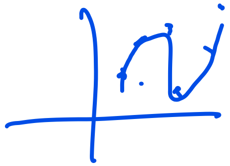
Maximum Likelihood X



→ which line is most likely to represent data points? A

→ How were we finding that line? → $\boxed{\text{least square}}$

Polynomial Regression : We do feature transformation depending on model we want.



For one feature : $\{1, x_1\}$, $\{1, x_1, x_1^2\}$

$y = c + m x_1 \rightarrow$

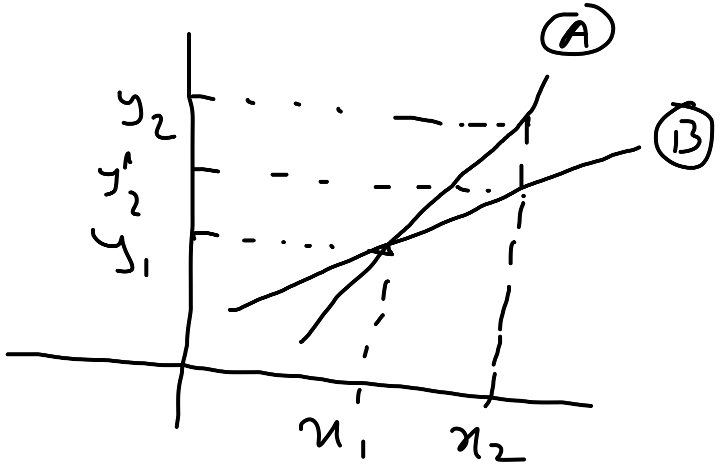
$y = a + b x_1 + x_1^2$

for two features : $\{1, x_1, x_2\}$, $\{1, x_1^2, x_2^2\}$, $\{1, x_1, x_2, x_1^2, x_2^2\} \rightarrow y =$

* First do transformation then apply Linear Regression

$y = \theta_0 + \theta_1 x_1 + \theta_2 x_1^2$

Regularization : To reduce sensitivity (wide variation) with input (layman)



→ which line is more sensitive to change in input?

→ why $(A^TA + \lambda I)$ is invertible even if A^TA is not

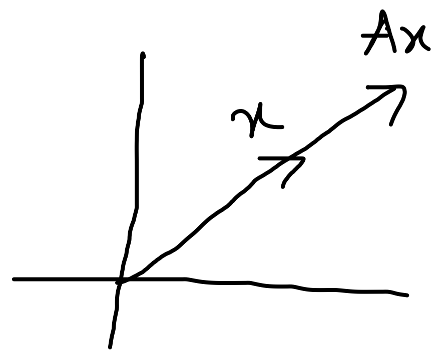
e.g. $A^TA = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 7 \\ 3 & 7 & 10 \end{bmatrix}$ $\quad A^TA + \lambda I = \begin{bmatrix} 3 & 2 & 3 \\ 2 & 7 & 7 \\ 3 & 7 & 12 \end{bmatrix}$ → Full rank invert

Lec 2: Eigen Value & Eigen Vector

* When you Apply a linear Transformation on a vector then if

$$\boxed{Ax = \lambda x}$$

where x = vector $\quad \lambda$ = scalar $\quad A$ = matrix



$x \rightarrow$ eigen vector $\quad \lambda \rightarrow$ eigen value

· Motivation was derived in Live session of week 4 #

$|(A-\lambda I)| = 0$ To find λ's

$Ax = \lambda x$ To find x's

Lec 3: Diagonalization

We did derivation in Live session

$n \times n$

Steps:

1. Find eigen values $(\lambda_1, \lambda_2, \lambda_3)$ 3×3
2. Find eigen vector (x_1, x_2, x_3)
3. $S = \underbrace{\begin{bmatrix} | & | & | \\ x_1 & x_2 & x_3 \\ | & | & | \end{bmatrix}}$

$S^{-1}AS = \Lambda$

4. $S^{-1}AS = \Lambda \rightarrow$

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}$$

$$\begin{bmatrix} | & | & | \\ x_1 & x_2 & x_3 \\ | & | & | \end{bmatrix}^{-1} \begin{bmatrix} & A & \end{bmatrix} \begin{bmatrix} | & | & | \\ x_1 & x_2 & x_3 \\ | & | & | \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}$$

Imp property: $\quad S^{-1}AS = \Lambda \quad$ —— ①

eqn① × eqn① $\quad S^{-1}AS\,S^{-1}AS = \Lambda^2$

$\Rightarrow \quad S^{-1}A^2S = \Lambda^2$

eqn① × eqn① . . . . k times $\quad S^{-1}A^kS = \Lambda^k$

Lec4: Fibonacci Sequence:

* Derived completely in live session
* Not as imp from exam point of view

Lec5: orthogonally diagonalizable matrices

* when A is real symmetric
* we know $S^{-1}AS = \Lambda$

$$A = S\Lambda S^{-1}$$

* for real symmetric case we have

$$SS^T = S^TS = I$$

i.e. S is orthogonal

$\therefore \quad S^T A S = \Lambda$

or $\boxed{A = S \Lambda S^T}$ $\quad \boxed{A = Q \Lambda Q^T}$

Steps :
1. Find eigen values
2. Find eigen vectors
3. ** Normalize eigen vectors.
4. Rest steps same as diagonalization.

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} \qquad \lambda_1 = 1 \qquad \lambda_2 = 3$$

For $\lambda_1 = 1$

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

$$\begin{bmatrix} 1-1 & 2 \\ 0 & 3-1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 0 & 2 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$0 \cdot x_1 + 2x_2 = 0 \qquad x_2 = 0 \qquad v_1 = \begin{bmatrix} k \\ 0 \end{bmatrix}$$

$$v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

For $\lambda_2 = 3$

$$\begin{bmatrix} -2 & 2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \quad -2x_1 + 2x_2 = 0$$

$$x_1 = x_2$$

$$v_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$S = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \qquad S^{-1} = \frac{1}{1-0} \begin{bmatrix} 1 & \\ & \end{bmatrix}$$

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

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

$$S^{-1}AS = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}$$

$$\underline{\underline{AA^{-1}}} = \underline{\underline{I}}$$

$$\boxed{A^{-1} = \frac{1}{|A|} \operatorname{adj} A}$$

$$y = mx + c$$

$$m = 0 \qquad c = 2$$

$$y = 0x + 2$$

$$y = 2$$

residual

x	y	$\hat{y}$	$\hat{y} - y$	$(\hat{y} - y)^2$
1	2	2	0	0
2	4	2	-2	4
3	1	2	-1	1
4	3	2	-1	1

Sum of residual $= 6$

$$SSE = \frac{6}{2} = 3 \qquad MSE = \frac{6}{4}$$

6 //

x_1 2nd degree model fit

$$y = \theta_0 + \theta_1 x + \theta_2 x^2$$

y	x
4	1
16	2
19	4

$$y_1 = \theta_0 + \theta_1 x_1 + \theta_2 x_1^2$$

$$y_2 = \theta_0 + \theta_1 x_2 + \theta_2 x_2^2$$

$$y_3 = \theta_0 + \theta_1 x_3 + \theta_2 x_3^2$$

$$4 = \theta_0 + \theta_1 + \theta_2$$

$$16 = \theta_0 + 2\theta_1 + 4\theta_2$$

$$19 = \theta_0 + 4\theta_1 + 16\theta_2$$

$$\begin{bmatrix} 4 \\ 16 \\ 19 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 4 & 16 \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{bmatrix}$$

$$\begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 4 & 16 \end{bmatrix}^{-1} \begin{bmatrix} 4 \\ 16 \\ 19 \end{bmatrix} = \begin{bmatrix} \quad \end{bmatrix}$$