Last Class – August 1

1) Quick recap
2) Review of R2 score
3) Model interpretability
4) Revision of Gradient Descent
5) Code for Linear Regression
6) Optimization
7) Implementing Gradient Descent
8) Plot loss function vs weights
9) How feature scaling helps in easier model training.

Today's class

1) Recap
2) Adjusted R^2
3) Intro to Stats Model
4) Assumptions of Linear Regression

$$\downarrow \quad 1 - \left[\underbrace{(1 - R^2)}_{1} \times \underbrace{\left(\frac{m-1}{m-d-1} \right)}_{2} \right] \uparrow \quad \uparrow$$

$d \uparrow \quad \frac{m-1}{m-d-1} \uparrow$

$(1-R^2) \frac{m-1}{m-d-1} \uparrow \quad 1 - (1-R^2) \times \frac{m-1}{m-d-1} \downarrow$

adj $R2 \downarrow$, if $d \uparrow$, $R2 \to$ constant

$d \uparrow$, $\frac{m-1}{m-d-1} \uparrow$ $R^2 \uparrow$, $(1-R^2) \downarrow$

$$1 - \left[(1-R^2) \downarrow \times \left(\frac{m-1}{m-d-1} \right) \uparrow \right]$$

$0.8 \to 0.7$ $\qquad$ $0.7 \to 0.8$

$1 - [0.8 \times 0.7]$

$1 - [0.7 \times 0.8]$

$0.8 \to 0.6$ $\qquad$ $0.7 \to 0.8$

$1 - [0.6 \times 0.8] = 1 - 0.48$

$= 0.52$

$1 - [0.8 \times 0.7] = 1 - 0.56$

$= 0.44$

$w_1 x_1 + w_2 x_2$ learnt

$w_1' x_1 + w_2' x_2 + w_3' x_3$

$w_1' = w_1$

$w_2' = w_2$

$x_2 \to 0.7$ to 0.1

$0.6x_1 + 0.7x_2$

$\Downarrow$

$0.5x_1 + 0.1x_2 + 0.8x_3$??

$0.7x_1 + 0.1x_2 + 0.8x_3$

$$1 - \left[\underset{1}{(1-R^2)} \times \underset{2}{\left(\frac{n-1}{n-d-1}\right)} \right]$$

$2 \geq 1$, $\quad \underline{1 \times 2} \geq (1-R^2)$

$(1-R^2) \times 2 \geq 1 \times (1-R^2)$

$\geq (1-R^2) \quad \ldots$ eq n (1)

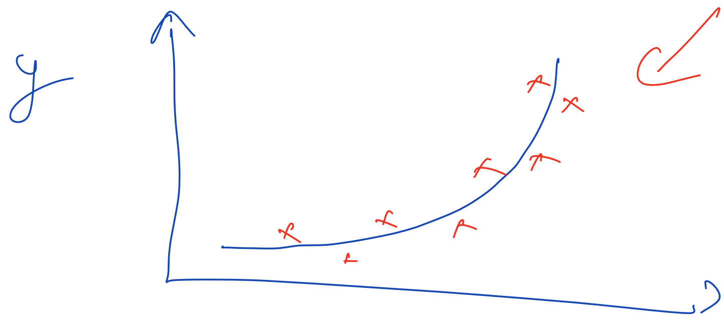
Subtracting both sides from 1

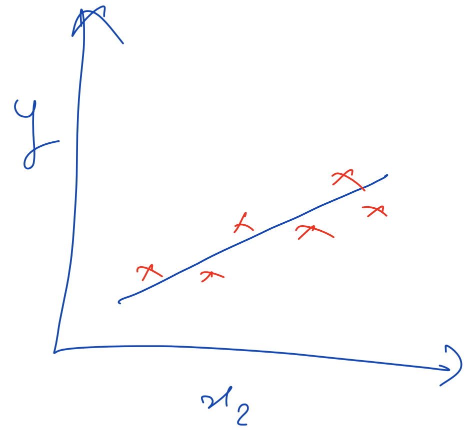
$$1 - \left[(1-R^2) \times 2\right] \leq 1 - (1-R^2)$$

$$1 - \left[(1-R^2) \times 2\right] \leq R^2$$

$$\text{adj } R^2 \leq R^2$$

$(y, \; x_1, \; x_2)$





x_1

$\downarrow$

$x_1^{1/2}, \quad x_1^{1/3},$

$(y, \; x_1, \; x_2, \; x_3)$

$$y = 1 + 0.5\,x_1 + 0.1\,x_2 + 0.7\,x_3 \quad (a)$$

$$W^* = (1, \; 0.5, \; 0.1, \; 0.7)$$

x_2 and x_3 are linearly related

$$\boxed{x_3 = 2\,x_2}$$

$$y = 1 + 0.5\,x_1 + 0.1\,x_2 + 0.7\,(2x_2)$$

$$= 1 + 0.5\,x_1 + 1.5\,x_2 + 0 \times x_3 \quad (b)$$

(a) $w' = (1, 0.5, 1.5, 2)$

(b) $w^* = (1, 0.5, 0.1, 0.7)$

Same linear reg model

$fI = w$ → messed up

x_1 → y

x_2, x_3 → i/p

$R^2_{x_1} \leftarrow \boxed{x_1} = \underline{\alpha_1} x_2 + \underline{\alpha_2} x_3 + \beta$

$R^2_{x_2} \leftarrow x_2 = k_1 x_1 + k_2 x_3 + \gamma$

$R^2_{x_3} \leftarrow x_3 = c_1 x_1 + c_2 x_2 + c_3$

Variance Inflation Factor

$$(VIF) = \frac{1}{1 - R^2_{x_j}} \quad , x_j = x_1, x_2, x_3$$

$$\rightarrow \boxed{R^2_{x_1} = 0}$$

$$VIF_{x_1} = \frac{1}{1-0} = \boxed{1}$$

$$\rightarrow R^2_{x_2} = \boxed{1}$$

$x_2 \rightarrow$ o/p

$x_1, x_3 \rightarrow$ i/p

$$VIF_{x_2} = \frac{1}{1-1} = \boxed{\infty}$$

$x_1, x_2, x_3, \ldots x_{10}$

VIF_{x_1}, VIF_{x_2}, ... $VIF_{x_{10}}$

$\times$ $\boxed{x_5} \rightarrow \boxed{100}$ ✓

$x_2 \rightarrow 99$ $\qquad x_2 \rightarrow 95$

$x_{10} \rightarrow 98$ $\qquad x_{10} \rightarrow 75$

⋮

$x_j \rightarrow VIF_{x_j} <= 5 \rightarrow$ keep

$x_j \rightarrow VIF_{x_j} > 10 \rightarrow$ drop it

$$R^2_{x_2} = -1$$

$$VIF_{x_2} = \frac{1}{1-(-1)} = \frac{1}{1+\cancel{1}} = 0.5$$

$$R^2_{x_2} = -\infty$$

$$VIF_{x_2} = \frac{1}{1-(-\infty)} = \frac{1}{\infty} = 0$$

$x_1, x_2, x_3, \ldots x_{10}$

$$^{10}C_2 = \frac{10 \times 9}{2} = 45$$

100 features $= {}^{100}C_2 = \frac{100 \times 99}{2} = 99 \times 50$

$= 4950$

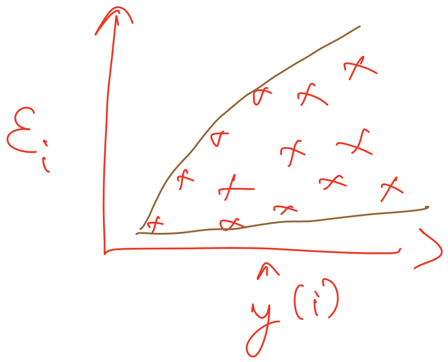
$(x_1, x_3) \rightarrow 0.9$

$(x_3, x_5) \rightarrow 0.89$

10:30 → Resume

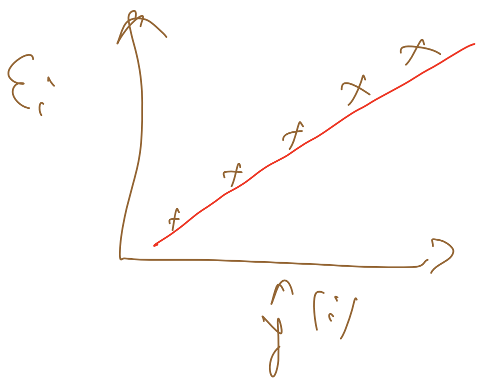
$$\varepsilon_i = y^{(i)} - \hat{y}^{(i)}$$

error　　gd　　pred

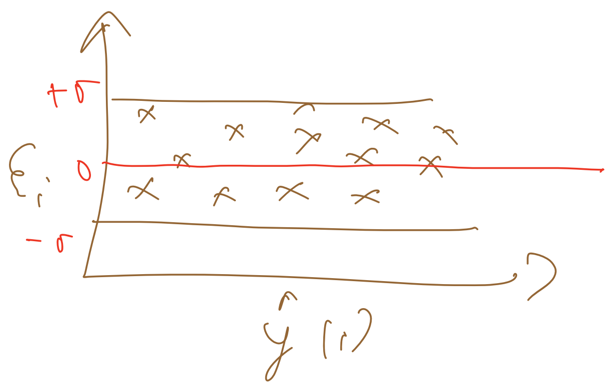


$$\varepsilon_i \propto \hat{y}^{(i)}$$

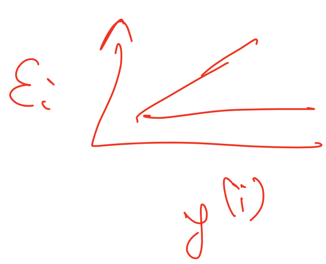
$$\varepsilon_i = c_1 \hat{y}^{(i)} + c_2$$



$$y^{(i)} - \hat{y}^{(i)} = c_1 \hat{y}^{(i)} + c_2$$



$$\varepsilon_i = y^{(i)} - \hat{y}^{(i)}$$



x

$[1, 1, 1]$ $[3]$

$[1, 2, 1]$ $[5.1]$

$[5, 5, 5] \rightarrow [25]$

$[0.1, 0.5, 7.1]$ $[100]$

$[-1.7, 2.0, -100]$ $[5.0]$

$$y^{(i)} - \hat{y}^{(i)} = \varepsilon_i = \frac{d\hat{y}^{(i)}}{dx^{(i)}}$$

$$\varepsilon_i = k\hat{y}^{(i)} + c$$

$$\boxed{\frac{d\hat{y}^{(i)}}{dx^{(i)}} = k\hat{y}^{(i)} + c}$$

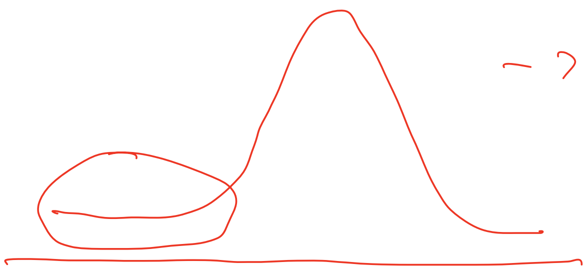
$$\underline{\hat{y}^{(i)}} = c_1 \cdot e^{x^{(i)}} + c_2$$

x_1, x_2, x_3

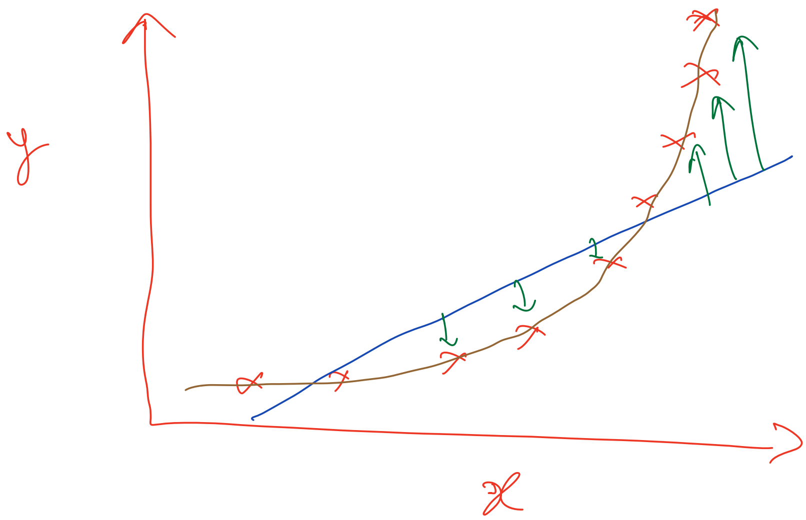
$x_2 = x_1$^2

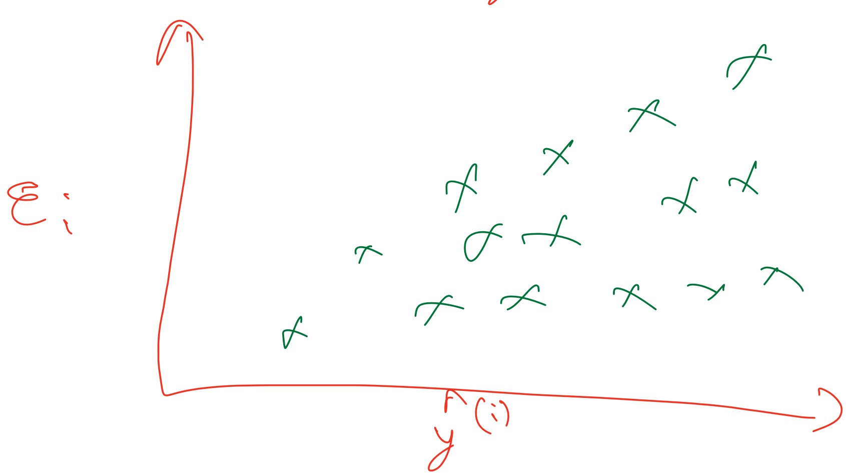
x_1, x_2, x_3

$x_2 = \sqrt{x_1}$



→ errors are not normally distributed





$$L = MSE = \frac{1}{m}\sum_{i=1}^{m}\left(y^{(i)} - \hat{y}^{(i)}\right)^2$$

$$\frac{\partial L}{\partial w_d} = \boxed{\frac{\partial L}{\partial \hat{y}^{(i)}}} \cdot \boxed{\frac{\partial \hat{y}^{(i)}}{\partial w_d}} \rightarrow x_d$$

$$\hat{y}^{(i)} = w_0 + w_1 x_1 + w_2 x_2 + \cdots w_d x_d$$