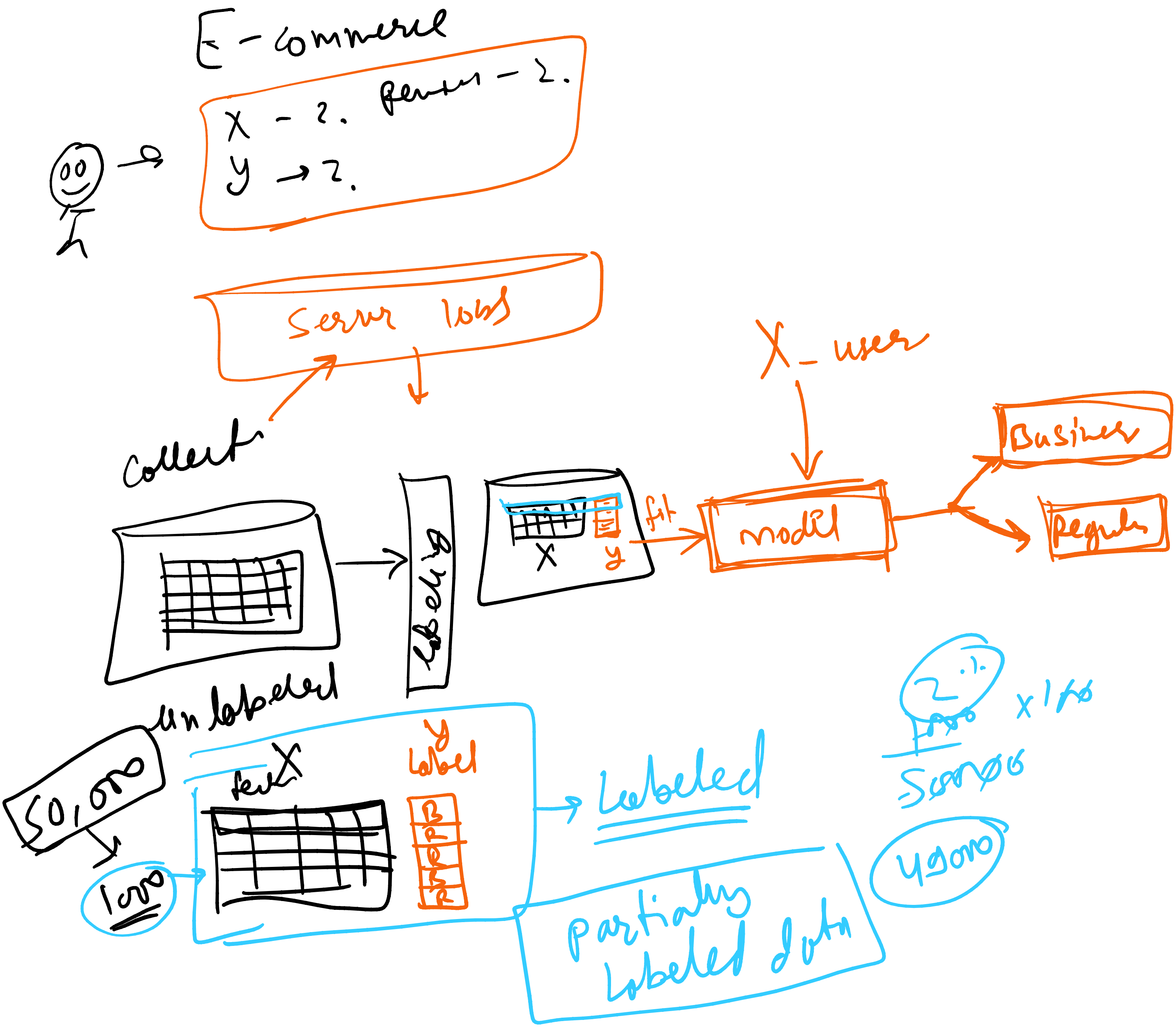
2.LinearRegressionModel

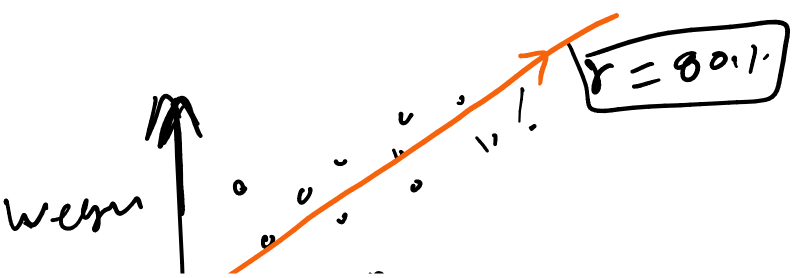
Sunday, July 31, 2022 1:05 PM

Data collection



X car → drive

y car → 20



Assemption

100%

92%

Dataset X

97% → 2%

LR	DT	RF	SVM	NB	NN
72	80	90	85	70	99

millions → 1 marks → model

assemption → LR, SVM

vector → numpy

Linear Regression

$H(X) = \hat{y}$

r = 70

$$\hat{y} = \hat{m}x + \hat{c}$$

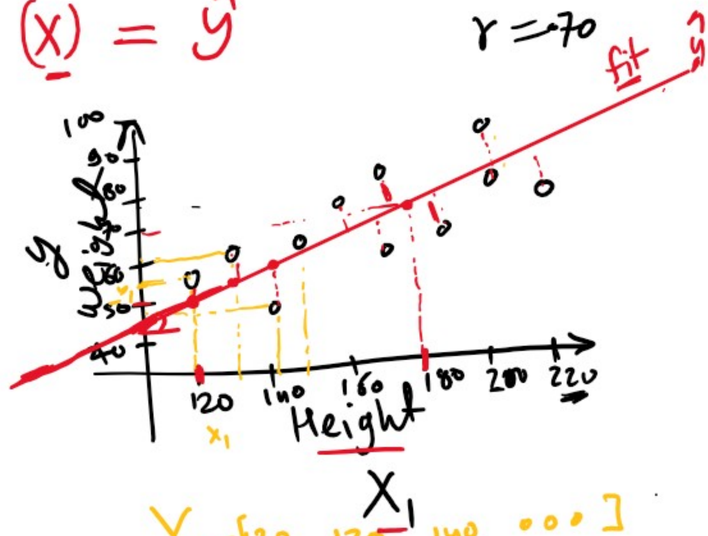
slope / coeff, intercept / bias

$$\hat{y} = \theta_0 + \theta_1 X_1$$

θ_0: bias, intercept; θ_1: coeff, slope

$$\hat{y} = b + w_1 x_1$$

$$\hat{y} = \theta^T \cdot X$$



X_1 = [120, 130, 140, ...]

y = [55, 60, 50, ...]

$\hat{y}$ = [50,

Training Data

$|y_i - \hat{y}_i|$

$X_0 = 1$

θ_0 = Bias

$$\boxed{\hat{y} = \theta \cdot X}$$ θ_0 = Bias

vector Notation

$$\boxed{\hat{y} = b + \underline{w} \cdot \underline{x}}$$

$\hat{\ }$ → estimated predicted

$$\Rightarrow \boxed{\hat{\text{weight}} = \hat{m} \cdot \text{height} + \hat{c}}$$

$H(x)$ prediction Function

⇒ Cost function / Error function

Error function should be convex

because we want to calculate / estimate parameters of $H(x)$ such that those parameters minimize the cost function $C(X)$.

Mean Squared Error (MSE)

or

L2 Norm

$$\boxed{C(X) = \frac{\sum \left(\hat{y}_i - y_i \right)^2}{N}}$$

L2 Norm

$$\boxed{C(X) = \frac{\sum\left(\hat{y}_i - y_i\right)^2}{N}}$$

① prediction / Hypothetical function $H(x)$

$$\boxed{\hat{y} = m \cdot X + C} \quad \text{.....(i)}$$

② Cost function $C(x)$

$$\boxed{MSE = \frac{\sum\left(\hat{y}_i - y\right)^2}{N}} \quad \text{.....(ii)}$$

③ performance metric

MAE

mean Absolute Error

$$\frac{\sum\left|\hat{y}_i - y_i\right|}{N}$$

goal is to find value of m, C (parameters) such that they minimize the cost function $C(X)$

$$\text{minimize } MSE = \text{minimize } \frac{\sum\left(m x_i + C - y_i\right)^2}{N} \quad \text{---(iii)}$$

we need to differentiate eq (iii) wrt to m and C both, then need to calculate minimum of each function.

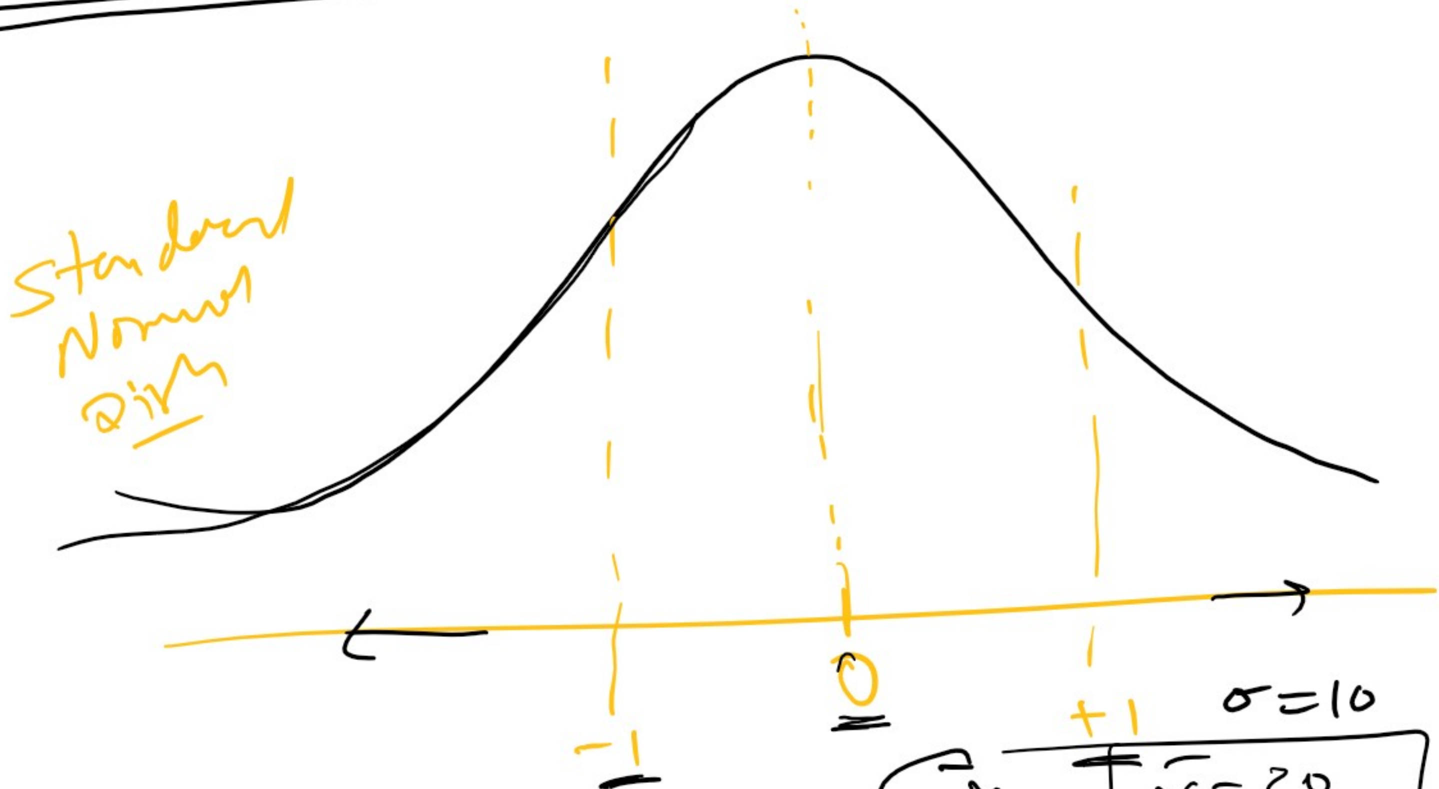
function :

Direct method to calculate params

$$\underline{m} = \frac{\sum_i (\bar{y} - y_i)\cdot(\bar{X} - X_i)}{\sum_i (\bar{X} - x_i)^2}$$

$$\underline{c} = \bar{y} - m\cdot\bar{X}$$

Standard scaling



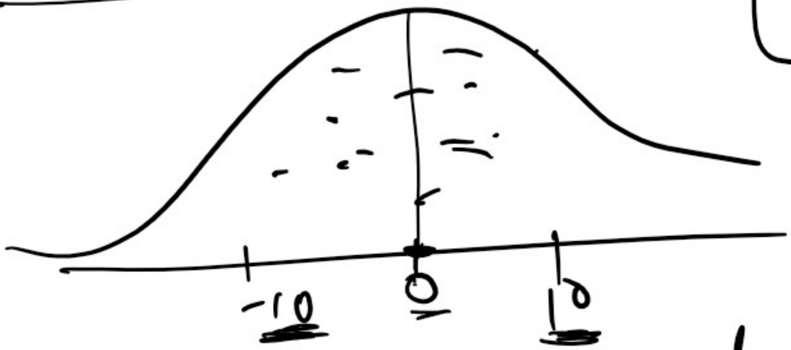
① shift mean to zero

$$\hat{X} = X - \bar{X}$$

X	$\bar{X}=20$
10	-10
20	→ 0
10	-10
20	10

$$X = X - X$$

10	10
30	10
30	

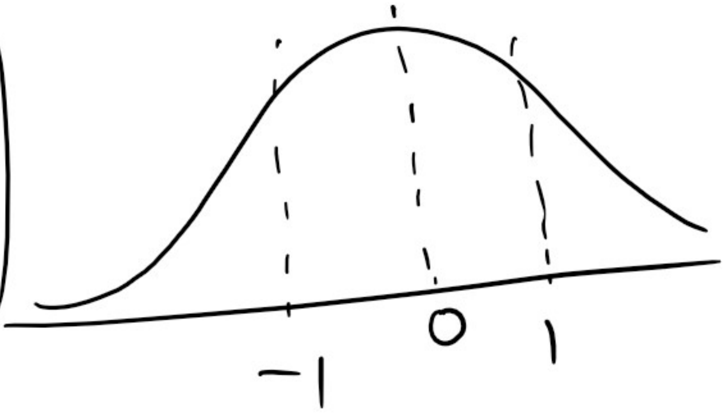


$\rightarrow \bar{X} = 0$

$\rightarrow \sigma = 1$

② make standard deviation to 1 unit

$$X\_tr = \frac{\hat{X}}{\sigma_X}$$



$$X\_scaled = \frac{X - \bar{X}}{\sigma_X}$$

10 20 30

-1 0 1

Generalize form

$$\hat{y} = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \cdots + \theta_n X_n$$

$$MSE = \Sigma(\hat{y} - y_i)^2$$

$$\boxed{MSE = \frac{\sum (\hat{y}_i - y_i)^2}{n}}$$

$$\boxed{\hat{\theta}_j = \frac{\sum_{i=1}^{N} (\bar{y} - y_i) \cdot (\bar{X}^{(j)} - X_i^{(j)})}{\sum (\bar{X}^{(j)} - X_i^{(j)})^2}}$$

$$X = \begin{array}{c} \begin{array}{cccc} X_1 & X_2 & X_3 & X_4 \end{array} \\ \left[\begin{array}{c} [X_{10}\ X_{20}\ X_{30}\ X_{40}] \\ [X_{11},\ X_{21},\ X_{31},\ X_{41}] \\ \vdots \\ [X_{1m}\ X_{2m}\ X_{3m}\ X_{4m}] \end{array} \right] \end{array}$$

observation

$m \times n$

feature

$$\underline{[\theta_0\ \theta_1\ \theta_2]}_{1\times 3} * \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & \cdots \\ X_{10} & X_{11} & X_{12} & X_{13} & X_{14} & \cdots \\ X_{20} & X_{21} & X_{23} & X_{22} & \cdots \end{bmatrix}_{3\times m}$$

$\Theta \rightarrow$ parameter $\qquad X \rightarrow$ matrix

$$\underline{[\theta_0\ \theta_1\ \theta_2]} \cdot \begin{bmatrix} X_0 \\ X_1 \\ X_2 \end{bmatrix}$$

$$\hat{y} = [\theta_0 \cdot x_0 + \theta_1 x_1 + \theta_2 x_2]$$

y = weight

$$\begin{array}{cccc} & & x_1 \text{ height} & x_2 \text{ Age} \\ X = & \left[\begin{array}{c} (1) \\ (2) \\ (3) \end{array}\right. & \begin{array}{c} 140 \\ 160 \\ 190 \end{array} & \left.\begin{array}{c} 20 \\ 60 \\ 30 \end{array}\right] \end{array}$$

weight $= b + w_1$ height $+ w_2$ Age

$$\boxed{\hat{y}_1 = \theta_0 + \theta_1 x_1 + \theta_2 x_2}$$

$$\Theta = [\theta_0 \quad \theta_1 \quad \theta_2] \quad 1 \times 3$$

$$X = \begin{bmatrix} 1 & 140 & 20 \\ 1 & 160 & 60 \\ 1 & 190 & 30 \end{bmatrix} \quad m \times 3$$

$$\boxed{\hat{y} = \Theta \cdot X^T}$$

$$\begin{array}{cccc} & 1 & 2 & 3 \; \dots \; m \\ [\theta_0 \quad \theta_1 \quad \theta_2] \cdot & \left[\begin{array}{c} 1 \; x_0 \\ 140 \; x_1 \\ 20 \; x_2 \end{array}\right. & \begin{array}{c} 1 \; x_0 \\ 160 \; x_1 \\ 60 \; x_2 \end{array} & \left.\begin{array}{c} 1 \; x_0 \\ 190 \; x_1 \\ 30 \; x_2 \end{array}\right] \end{array}$$

$$\hat{y} = [\underbrace{\theta_0 + \theta_1 \cdot 140 + \theta_2 20}_{y_1}, \; \underbrace{\theta_0 + \theta_1 \cdot 160 + \theta_1 \times 60}_{y_2}, \; \underbrace{\theta_0 + \theta_1 190 + \theta_2 \cdot 30}_{y_3}]$$

Linear Regression

$$\boxed{\hat{y} = H_\theta(x) = \theta \cdot X}$$

$$\boxed{\hat{\theta} = \left(X^T \cdot X\right)^{-1} \cdot X^T y}$$

direct method

$$\boxed{MSE(X, H_\theta) = \frac{1}{m}\sum_{i=1}^{m}\left(\theta^T x^{(i)} - y^{(i)}\right)^2}$$

→ Gradient Descent (optimization techniques)

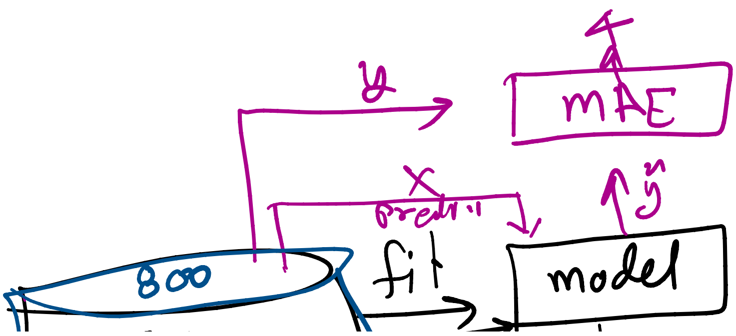
→ Hyperparameter Tuning

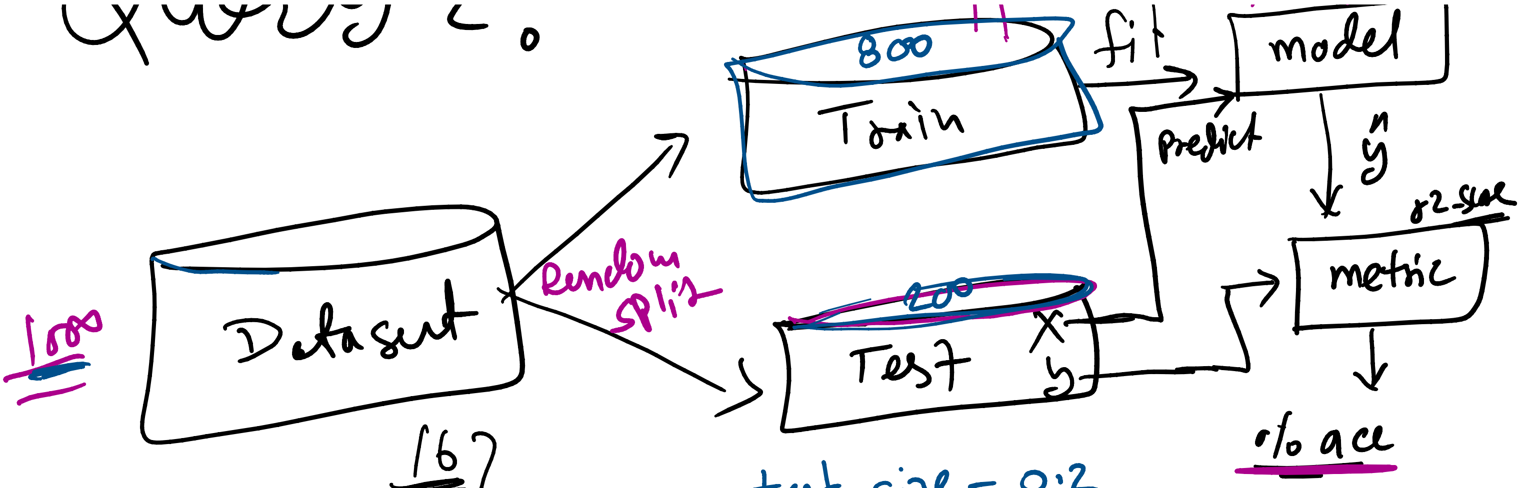
Housing

Assignment – $\boxed{\text{CH-1}}$, $\boxed{\text{CH-2}}$

$\boxed{\text{CH-4}}$ Regression

Query 2.





mpg = 16, 18, 22

test_size = 0.2

20%

0 – 8

cyls

hum

4560

$$\underline{y} = \theta_0 + \theta_1 \underline{X_1} + \theta_2 \underline{X_2} + \dots \theta_n X_n$$