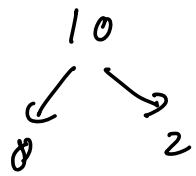
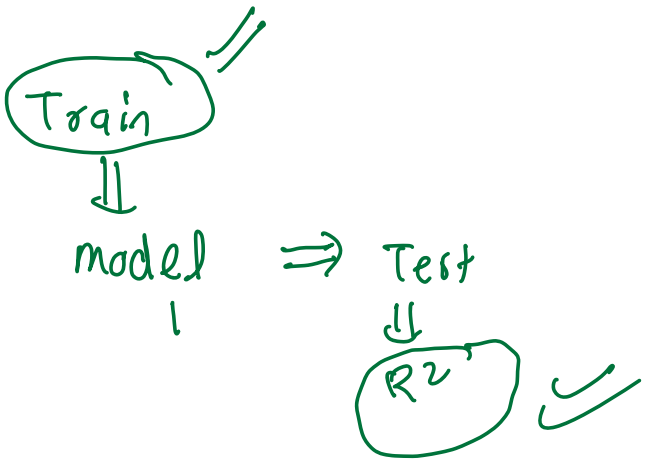
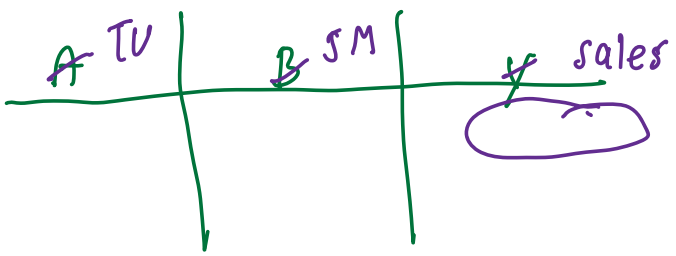
	Exp	Salary
Train	2	20K
Train	3	30K
Test	2	25K
T	—	—
T	—	—
Test	—	—
T	—	—
T	—	—
T	+	—



80 : 20

70 : 30





$$w_1 A + w_2 B + w_0 = \hat{y}$$

$$3 + 2 * TV + 3 * SM = y$$

$$\Rightarrow \quad 3 + 2\,TV - 3\,SM = y$$

TV	SM (million)	Sales
10,000	3	
1 L	21	
100,000		

$$w_1\,TV + w_2\,SM + w_0 = Sales$$

Min-max scaling : (0 - 1)

$$\frac{x_i - min}{max - min}$$

$x_i \simeq (min, max)$

$$\frac{min - min}{max - min} = 0$$

$$\frac{max - min}{max - min} = 1$$

Age	
20	0
25	0.33
30	0.67
28	0.6
35	1

(20 ⇔ 35)

$$\frac{35 - 20}{35 - 20}$$

Age	origin
20	America
25	India
26	India
28	Australia

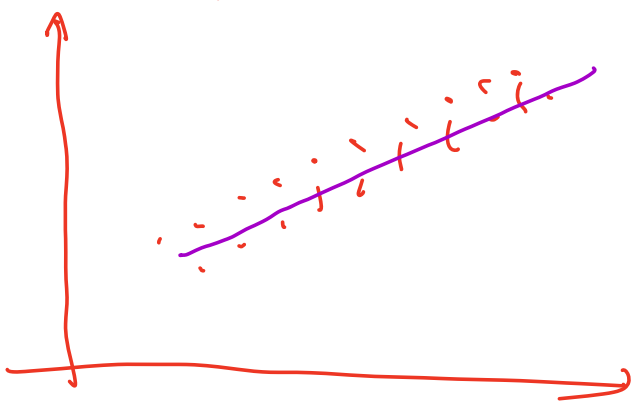
Age	America	India	Aus
20	1	0	0
25	0	1	0
26	0	1	0
28	0	0	1

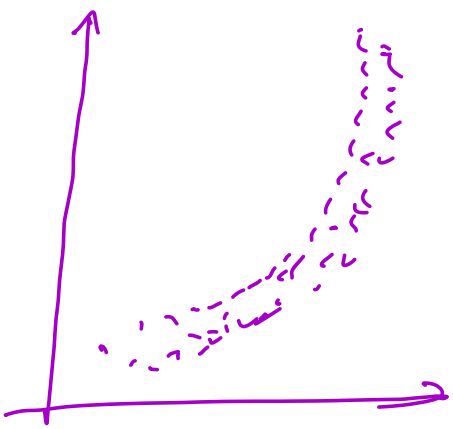
Outliers

Assumptions

Break : 10:16 PM

① Assumptions of Linear reg





② Features are not multi-colinier

x_1 | x_2 | x_3 | x_4 / y

$$y = f(x)$$

$$x_2 = w_1 x_3 + w_2 x_4$$

VIF (variance Inflation factor)

x_1, x_2, x_3, x_4

x_1 x_2 x_3 x_4

model

R^2

$x_1 = w_1 x_2 + (w_2 x_3 + w_3 x_4)$

$= w_1 x_2 + w_3 x_4$

x_2 x_1 x_3 x_4

$$VIF = \frac{1}{1 - R_i^2}$$

VIF > 10 : v High Multicolc

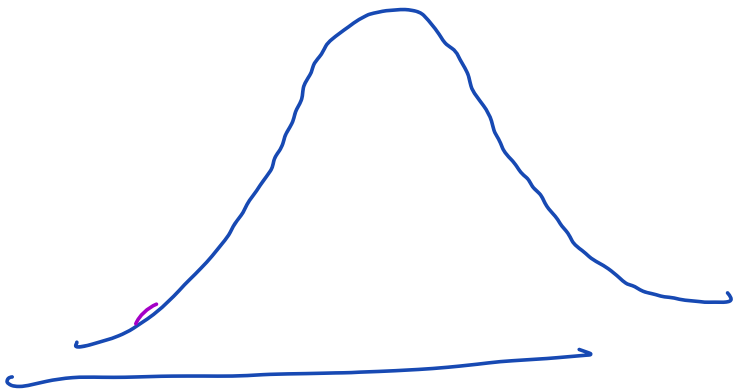
VIF (5, 10)

VIF

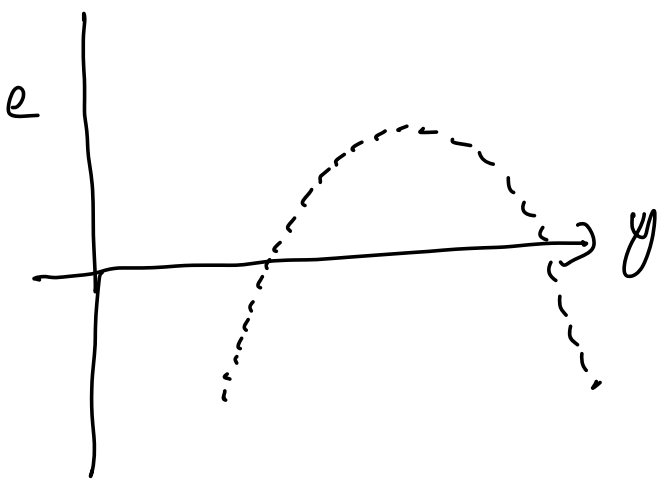
⇒ variance_inflation_factor(x)

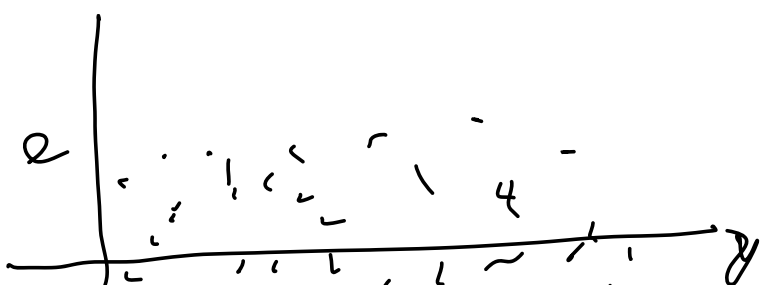
(3) Errors are normaly distribifute

e_i

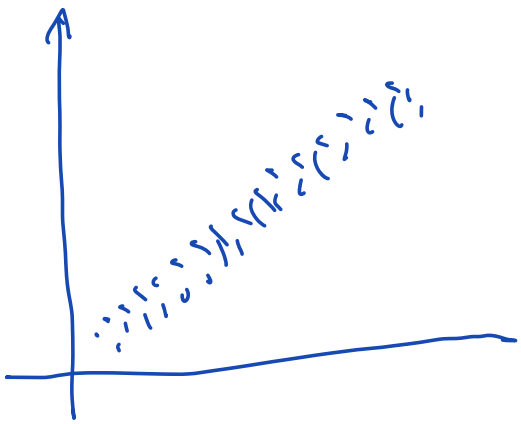


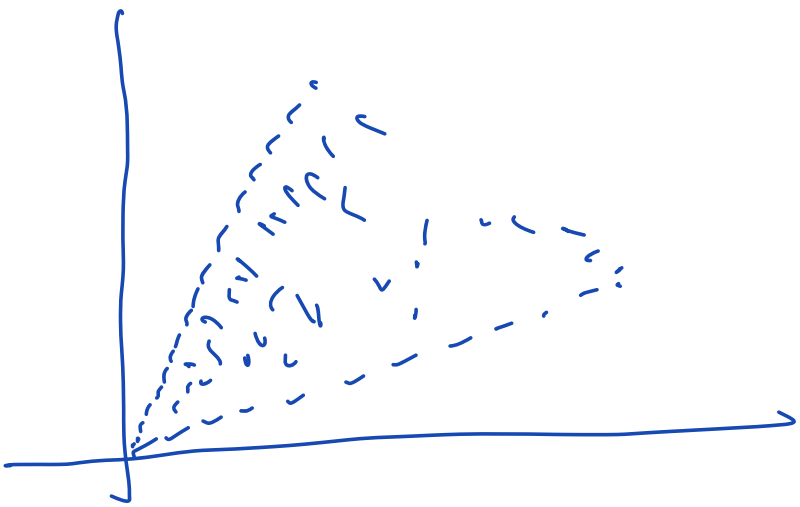
(4) Error terms must be independent





⑤ Homoscedasticity : Error terms have constant variance



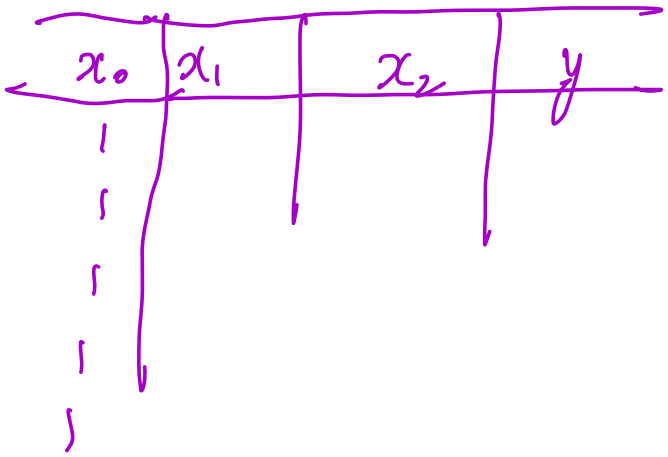


OLS model:

$w_0 x_0 + w_1 x_1 + w_2 x_2 = y$

$x_0 = 1$

OLS(x) → x_0, x_1, x_2



$w_1 x_1 + w_2 x_2 + w_0 x_0 = y$

LR.fit (X, y)

OLS (X, y)

x_0 x_1 x_2

Ho : all irrevalent

p-value

x_1 0.10 → 90%

x_2 0.03 → 97%

$\alpha = 5\%$