Linear Regression

Variance → Co-var. → Co-relation → Regression.

(one var.)

(2 var. & dirⁿ of Rel.)

dirⁿ + Strength

dirⁿ strength + (How change in one var. attributes to change in another)

Historical data → predict → Future Data

eg. → from historical data I've found out that Price of House is dependent upon its size.

$P \propto S$

$P = kS$

1000 sq. ft.

8000

80 Lac.

80,00,000

If I give you a new size ⇒ Price?

1500 sq. ft.

1500 × 8000

1,20,00,000

1-2 crores.

$P \propto$ Size ✓

$P \propto$ No. of Rooms ✓

$P \propto$ No. of Balcony ✓

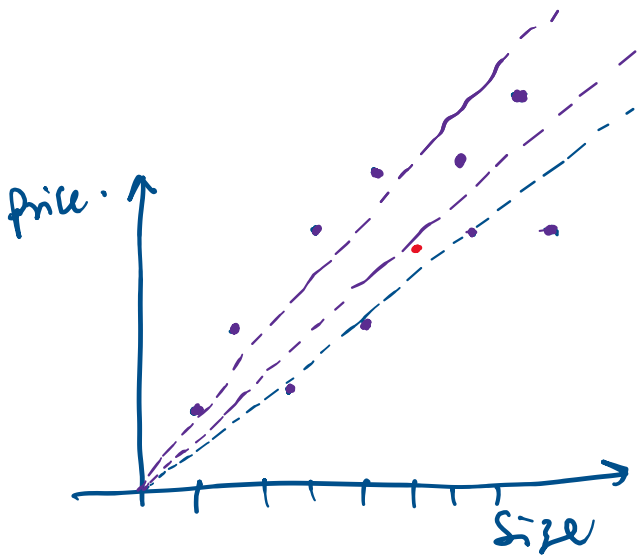
$\propto$ No. of washrooms ✓

Mathematics -

Multiple lines.

↓

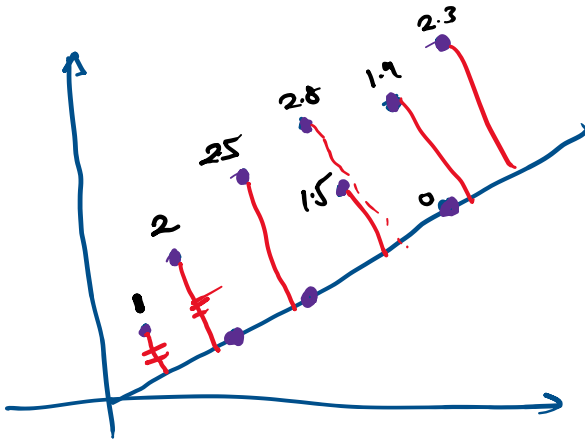
Best fit line



Case I

Error = dist^n betn pt. & line.

$1 + 2 + 2.5 + 1.5 + 2.8 + 0 + 1.9 + 2.3$

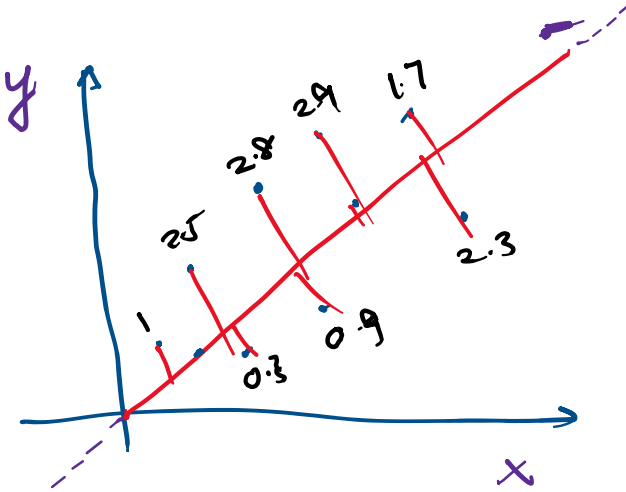


Aim = Best fit line

↓

Error = Min^m.

Error = $1 + 2.5 + 2.8 + 2.9 + 1.7$
$(+0.3 + 0.9 + 2.3)$



$y = mx + C$

m → Slope

C → Intercept.

Sales $\propto$ Adv.

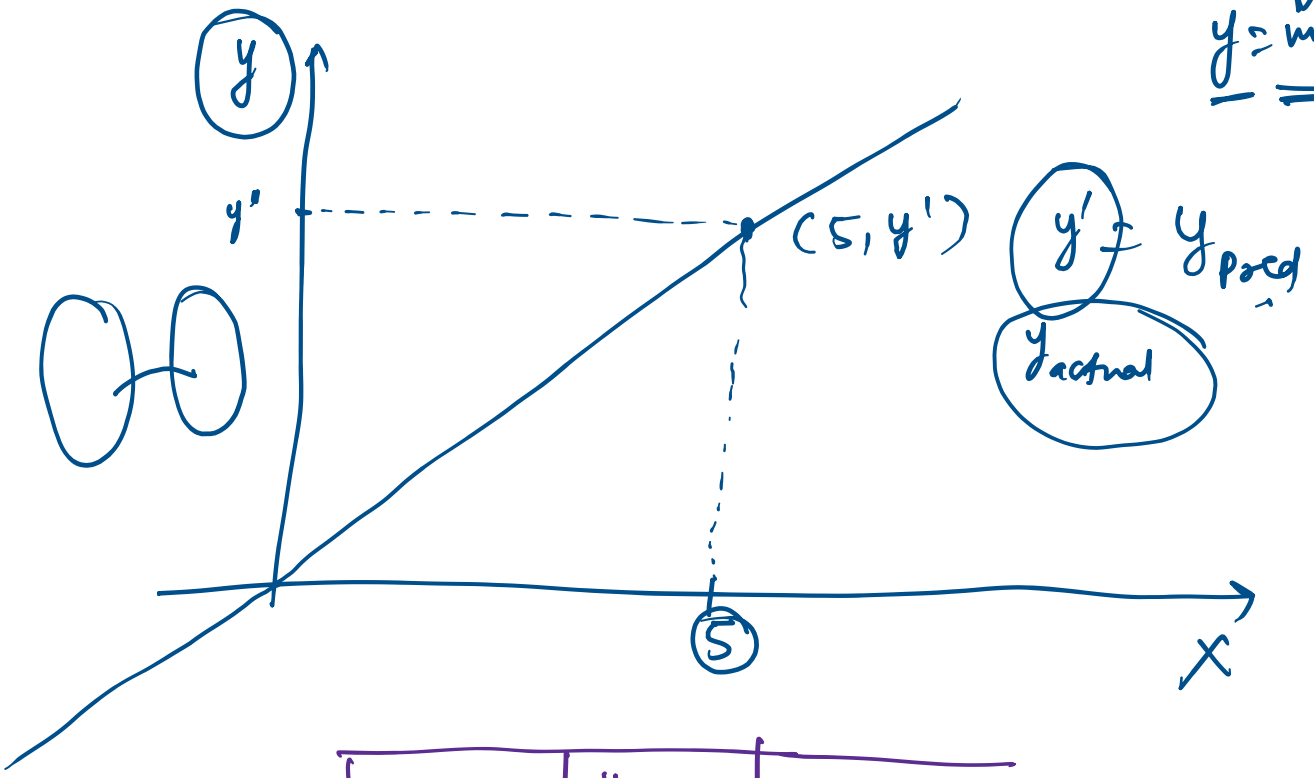
Sales = $K(Adv) + C$

If Adv = 0, sales = 0.

Our goal is to find statistically significant values of slope (m) & Intercept (C) that minimizes

diff betn y_{pred} & y.

$y = mx + c$



y_{pred}	y_{actual}	Error
65	68	3
60	60	0

80 !: ∞

To

Tom

Ways to think abt the error,

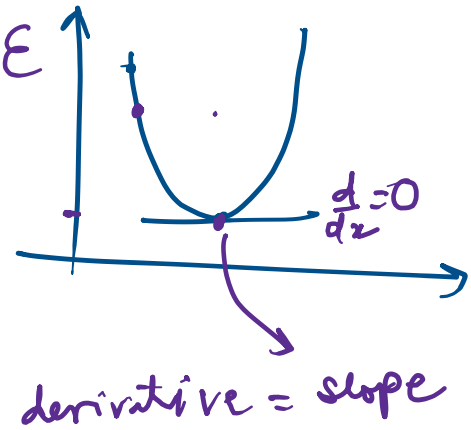
1. Absolute
2. Square

✗ (3) 4th Power ✗ (more complex)

✗ (4) 6th Power ✗ (more complex)

Error needs to be minimised

$$\frac{d}{dx}(\text{error}) = 0$$



Eq of error.

$$\frac{d}{dx} x^2 = 2x = 0$$

$x = 0$

$$\frac{d}{dx}(x) = 1$$

$$\frac{d}{dx}(x^2) = 2x$$

Error.

SSR $= \Sigma (y - \hat{y})^2$ ⟶ Reduce this error.

Sum of Sq. residuals

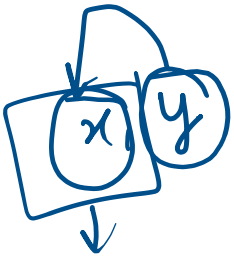
Algorithms

1. ordinary Least Square → Best fit Line.
2. Gradient descent. → Step wise.

1. OLS:

Least Sq. $= \dfrac{\text{Cov.}(x, y)}{\text{var}(x)}$ → var. of y wrt x

Co-Var. of x wrt. itself ↓ var(x)



$$= \sum_{1}^{n} \frac{(x - \bar{x})(y - \bar{y})}{n} \Big/ \frac{\Sigma (x - \bar{x})^2}{n}$$

$$= \frac{\Sigma (x - \bar{x})(y - \bar{y})}{\Sigma (x - \bar{x})^2}$$

How would you become great / best

Become Better.

Best

2, 3

(OLS)

Best fit line

Minimize your fault.

(Gradient Descent)

Minimizing the error (step wise)

Gradient Descent:

Aim is to minimize the error.

Cost

Cost function

$$SSR = \sum (y - \hat{y})^2$$

$(wx+b)$

$$h(\theta) = (y - (wx+b))^2$$
$$= (y - wx - b)^2$$

$$\frac{dh(\theta)}{dw} = 2(y-wx-b)\cdot(-x)$$
$$= -2(y-wx-b)(x)$$
$$\Downarrow$$
dw

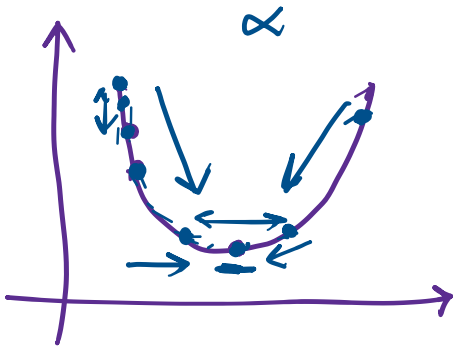
$$\frac{dh(\theta)}{db} = 2(y-wx-b)\cdot(-1)$$
$$= -2(y-wx-b)$$

db ⟸

HW?

$$w_{next} = w - \alpha(dw)$$

$$b_{next} = b - \alpha(db)$$

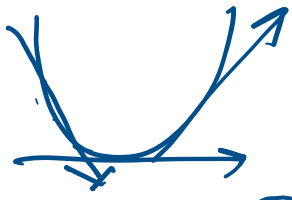


eg. → $w_{next} = 5 - 0.1(-0.25)$

$= 5 +$ [ ] → fwd dirⁿ

$W_{next} = 5 - 0.1(+0.25)$

$= 5 - \square$



backward direction θ = -ve

$+\theta$ = +ve

$+\infty = 0$

Accenture

"ACCY"

MLR

y = x_1 = Interest rates

x_2 = GDP

x_3 = value of NIFTY 50 Index

x_4 = Spend. on Ad.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$$

Project

Evaluation Metrices

* Error Metrices

(1) Mean Absolute Error : (MAE)

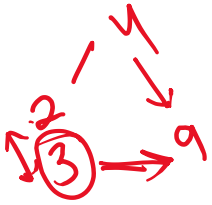
Mean of the absolute value of errors.

$$\frac{1}{n}\sum_{i}^{n} |y - \hat{y}|$$ ✓

(2) Mean Squared Error (MSE):

Mean of squared Errors.

$$\frac{1}{n}\sum_{i=1}^{n} (y - \hat{y})^2$$ ✓ ✗



Absolute
%
%ile
Rank

(3) Root Mean Squre Error (RMSE)

Sq. Root of Mean of Squred Error.

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n} (y - \hat{y})^2}$$

MAE V/S RMSE

E	|E|	E^2
2	2	4
2	2	4
2	2	4
2	2	4
2	2	4
2	2	4
2	2	4
2	2	4
2	2	4
2	2	4

T-1

- MAE = 2
- RMSE = 2

Evenly distributed errors

E	|E|	E^2
1	1	1
1	1	1
1	1	1
1	1	1
1	1	1
3	3	9
3	3	9
3	3	9
3	3	9
3	3	9

T-2

MAE = 2
RMSE ≈ 2.25

E	|E|	E^2
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
20	20	400

T-3

MAE = 2 ✓
RMSE = 6.32

Since RMSE gives relatively high weights to large errors hence should be used when ~~large~~ large errors are undesirable.

- Interpretability = MAE > RMSE > MSE
- Sensitive to outlier = MSE > RMSE > MAE

R^2 & Adjusted R^2

R^2 = Coeff. of determination

$-1 \leq r \leq 1$

$0 \leq R^2 \leq 1$

$$R^2 = 1 - \frac{SS_{res}}{SS_{total}}$$

$$= 1 - \frac{\sum (y - \hat{y})^2}{\sum (y - \bar{y})^2}$$

1 or 100%

0 or 0%

68%

$$R^2 = 1 - \frac{\text{unexplained variance}}{\text{Total variance}}$$

Adjusted R^2

Iteration 1: 5 var.

	R^2	Adj R^2
	68%	
+1 → 6	68.9%	
+1 → 7	69.2%	

y

Adj R^2 can decrease also.

R^2

32

EDA select variables

Adj. — R^2

77.8%

X_1, X_2, X_3 R^2

R^2 = 78%

77.8

77.0%

X_1, X_2, X_3, xy y

R^2 = 79.5%

x

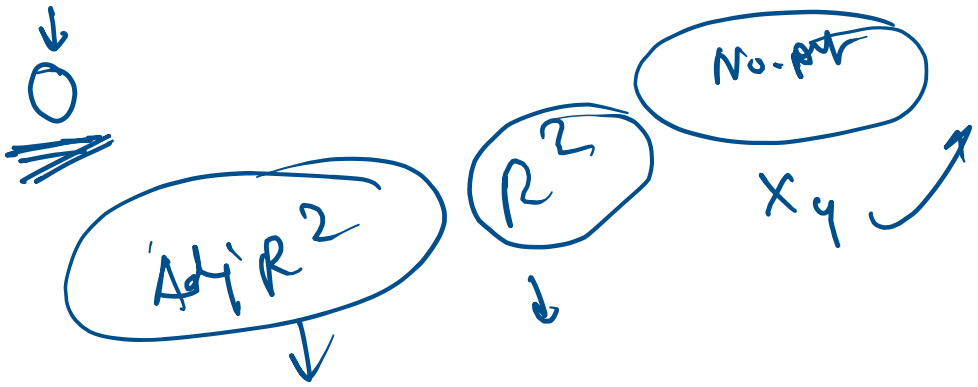
$$\bar{R}^2 = 1 - \frac{SS_{res}/df_e}{SS_{total}/df_t}$$

x ✓

1.5

Adj R^2 Penalises for every increase in variable.

R^2 ↓



100%

Q. What are the evaluation metrics you have used for your Regression Model?

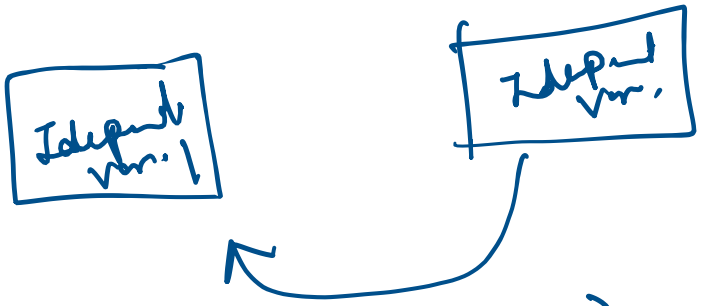
1. R^2
2. Adj-R^2
3. MSE
4. MAE
5. RMSE

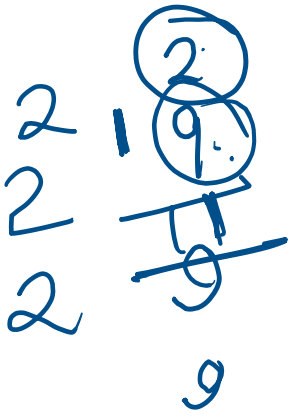
Multicollinearity & How to Avoid it?

Independent variable

Dependent variable

what happen if one independent var. is highly correlated with one or more Independent Variable?



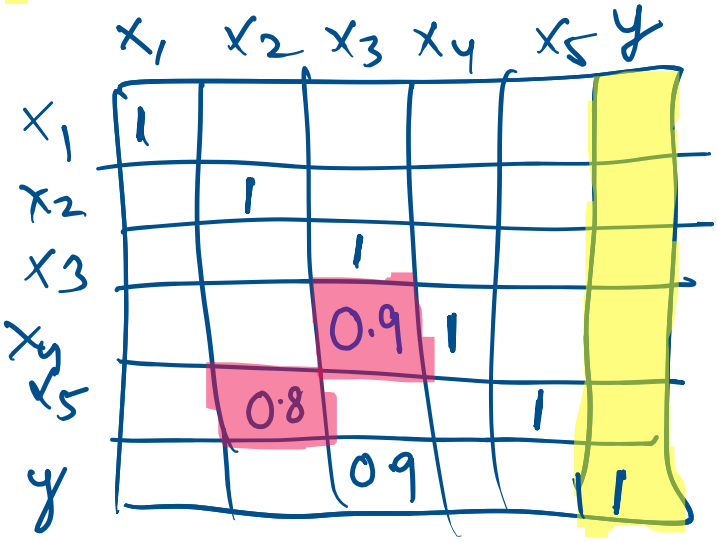


Model (generalisation)

Linear (")

Detection :

(1) Correlation Matrix



② Variance Inflation factor (VIF) :-

VIF tells how much the variance of an estimated Regression Co-eff increases if your predictors are correlated.

$$VIF = \frac{1}{1-R^2} \qquad VIF = (1, \infty)$$

VIF (now about 5 or more).

Prevent

① drop one of it.

② Feature Engineering. (X_3 & X_4)

⇓

Combinata

Weigh Heigt

⇓

BMI

X_3	X_4	Y
.	→	0.9 ✓
	→	0.8 ✗

EOF >>>