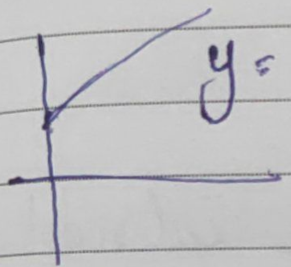
Date ....

(1) Linear Regression:-

→ Comes Under Supervised Learning (both X & Y)

→ LR is regression algm which is used to predict Continuous target variable.

→ Linear regression is used to find linear rltnshp b/w dependent (y) & independent variable (x) by fitting straight line.



$y = mx + c$

$m \to$ slope
$c \to$ y intercept.

slope $m = \frac{dy}{dx}$ → unit change in x results in change in y (i.e.) gradient.

Assumption:

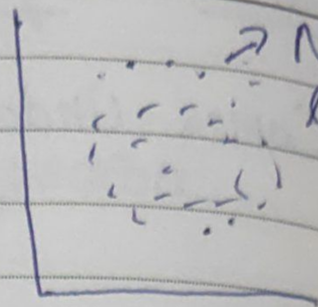
1. Linearity → corr should be high. (x & y)
2. Normal Distribution data
3. Little or no multicollinearity.
   ↳ There should not be any high correlation b/w independent variables / input variables
   eg: Area / sqft.

Date ....

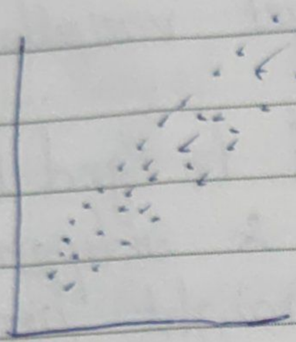
Note:

$x \xrightarrow{\text{corr high}} y$

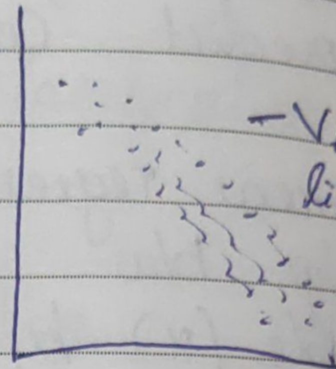
$x_1 \xrightarrow{\text{corr low}} x_2$



Non linearity



+ve linearity



-ve linearity

How LR works?

LR will find slope & intercept.

$m = ?$ $c = ?$

$$m = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})}$$

$$c = \bar{y} - m\bar{x}$$

$\bar{x}, \bar{y}$ are mean of x & y

Date ....

How do you say whether model is best or not?

→ Calculate error:-

Error/Loss → Actual Value - Predicted Value

→ $y - \hat{y}$ → (single value)

If Error is less then its ~~bad~~ ~~road~~ best model otherwise its bad model.

① How to calculate average error / Total error? (large no. of records)

1) Mean Squared Error (MSE)

$$MSE = \frac{1}{n} \Sigma (y_i - \hat{y}_i)^2$$

y_i - actual value

2) Mean absolute error (MAE)

$\hat{y}_i$ - predicted value

$$MAE = \frac{1}{n} \Sigma |y_i - \hat{y}_i|$$

n → no. of observation

3) Root mean squared error (RMSE)

$$RMSE = \sqrt{MSE}$$

Date ....

Notes:

→ MSE, MAE, RMSE are the error functions which tells about error made by the model.

→ Model is best only if total error is less.

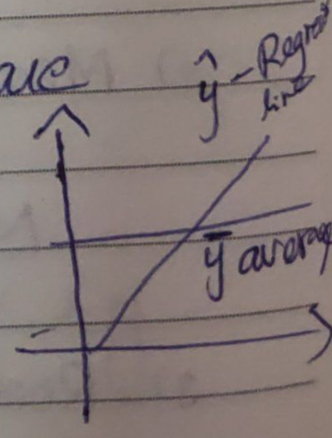
→ Error = actual − predicted.

② How to evaluate model?
How to measure performance of model?

→ R-squared / R2-score / R^2-score

$$R^2 = 1 - \frac{RSS}{TSS}$$

RSS - Residual sum of square



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

TSS - Total sum of square

$$TSS = \sum (y_i - \bar{y}_i)^2$$

Date ....

R^2 Compares Regression line with average line.

Range of $R^2 = [-1, 1]$

$R^2 = 90\%$ → better model.

$R^2 = 1$ → Best model

$R^2 = 0$ → Bad model

How to handle outlier:-

(1) 3-Sigma Rule:-

This method is used when data is normal.

Lower Limit $= \mu - 3\sigma$

Upper Limit $= \mu + 3\sigma$

Value $\in [\mu - 3\sigma, \mu + 3\sigma]$

Note:-

* Handle Outlier only if percentage of outliers is less than 5%.

Date ....

* If its above 5% dont do anything

Heatmap Analysis:-

Linearity:-

→ We include columns which has correlation with target.

→ We drop columns which has less correlation with target.

Little or no multicollinearity:-

→ We include columns which has less correlation among each other.

→ We drop one of the column if two input variable have high correlation

Date ....

Adjusted R^2-Score :

R^2 score increases as the number of independent of variables increases which has very less relationship with target variable.

To overcome above issue we use adjusted R^2

Adjusted R^2 score will measure the performance of the model by ignoring columns which has very less relationship with the target.

$$adj\,R^2 = 1 - \frac{(1-R^2)(N-1)}{N-P-1}$$

R^2 - R^2 score

N - No. of observation (test_size), P - No. of independent Variable

Note:

Adj R^2 < R^2 score, then we say its good model.

Date

How to select the best line?

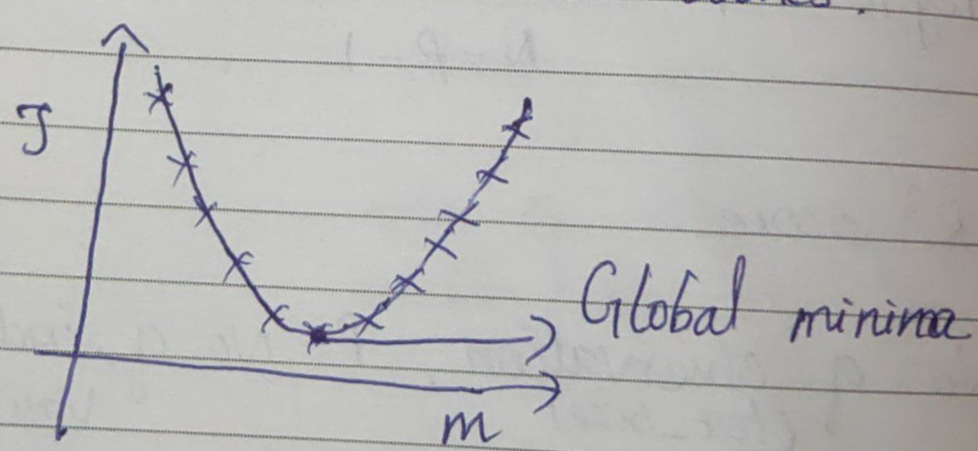
Cost function $J = \frac{\sum (y_i - \hat{y}_i)^2}{n}$

* what if error is very high?

Gradient Descent

It is one of the optimization technique which minimizes error/loss by choosing optimal value for slope & intercept.

* How Gradient Descent works?



It starts with random ~~sp~~ slope & works iteratively to reach global minima.

Date ....

Global minima is a point where error is less.

$$\left.\begin{aligned} m_{new} &= m_{old} - \eta \frac{\partial J}{\partial m} \\ c_{new} &= c_{old} - \eta \frac{\partial J}{\partial m} \end{aligned}\right\}$$ Eqns to update slope & intercept

$\eta \rightarrow$ learning rate $\Rightarrow$ How many steps model has taken to reach global minima

If η is small $\rightarrow$ It will take more time to reach global minima (not good option)

If η is large $\rightarrow$ It will overshoot that means it will never reach global minima.

How to choose η?

Initially take larger steps and take smaller steps when near to global minima.