Machine learning Algorithms

In ML Intuition is very important for the interviews.
Understand the maths behind all algorithms.

ML Algorithms
- Supervised ML → Regression, Classification
- Unsupervised ML → Clustering

Lets talk about Regression and classification problem

Regression Problem

→ House price prediction.

No. of Rooms → Input
Total size → Input
Location → Input
→ Independant variables

price → Output → Dependant variables

↳ 1.5m, 1.66m, 1.725m
↳ Output is continuous in regresion problem

Classification Problems

① Binary Classification
② Multiclass Classification

Example for binary classification

No. of play hour	No. of sleep hours	No. of Study Hours	Pass/ Fail
5	8	2	F
6	8	1	F
0	7.5	5.5	P
3	9	2	P
4	5	7	P

→ Binary Classification

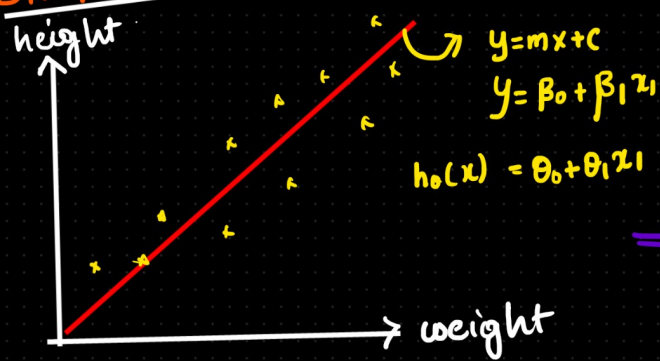
Regression Algorithm:

1. linear Regression [Regression]
2. Ridge Regression [—»—]
3. Lasso Regression [—»—]
4. Elasticnet Regression [—»—]
5. Logistic Regression [classification]

6. Decision tree
7. Random forest
8. Adaboost
9. XG Boost

→ Both Regression and Classification Problem.

Simple Linear Regression



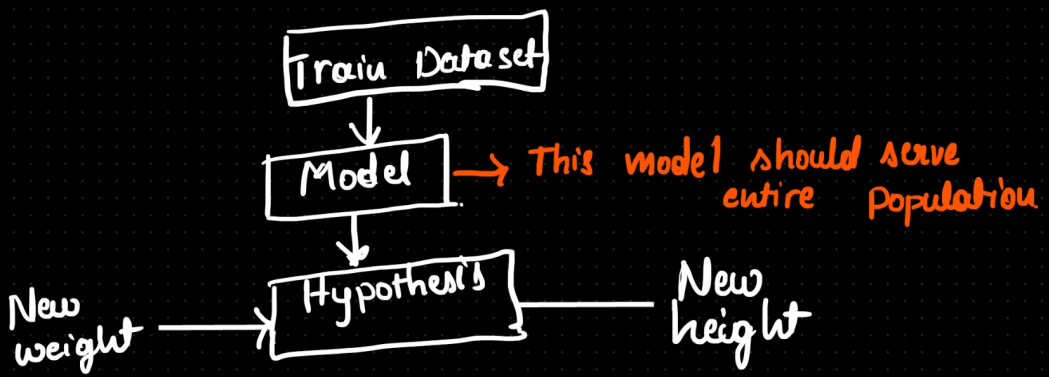
weight → I/P

height → O/P.

⇒ If we have 2 feature we will call as simple linear Regression

⇒ If we have more than 2 feature we will call as multiple linear Regression

Goal of linear Regression is to create the best fit line.

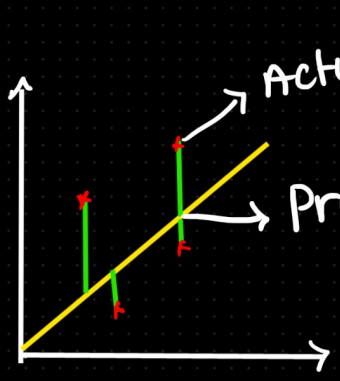


Equation of Straight line

$h_\theta(x) = \theta_0 + \theta_1 x$ → Equation using for hypothesis testing

θ_0 → Intercept

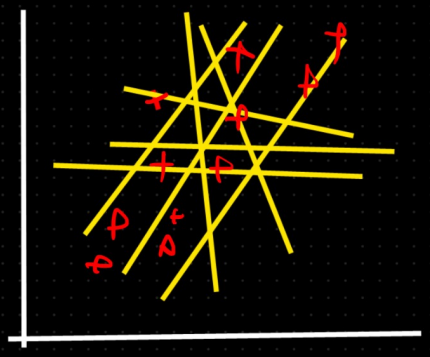
θ_1 → Slope/Coefficient



Error = Predicted Point − Actual Point

Error = $\hat{y} - y_i$

1) we should change the value of θ_0 and θ_1 and we need to find best fit line with minimal error.



→ wrong approach

2) we cannot find best fit line by selecting random values. we need to choose a value and we need to go slowly towards best fit line.

3) To achieve best fit line we need to use the following 2 important concepts

⇒ COST FUNCTION

⇒ GRADIENT DESCENT

$h_\theta(x) = \theta_0 + \theta_1 x_1$ ⇒ Now Error means $\hat{y} - y$

1

= COST FUNCTION $J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x_i) - y^i)^2$ ($\hat{y}$ = $h_\theta(x_i)$)

MEAN SQUARED ERROR

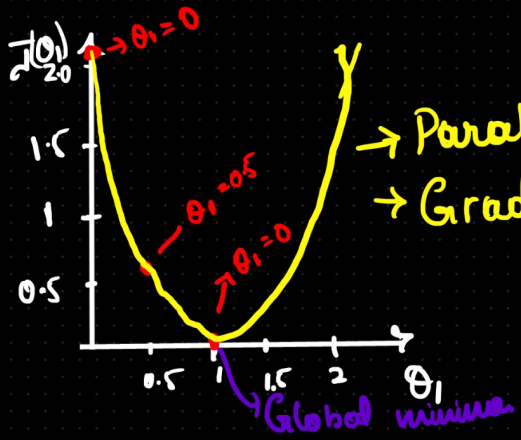
what we need to solve?

we need to make sure that

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^i) - y^i)^2$$

(A) $h_\theta(x) = \theta_0 + \theta_1 x_1$ if my $\theta_0 = 0$ my straight line

$h_\theta(x) = \theta_1 x_1$



→ Parabola

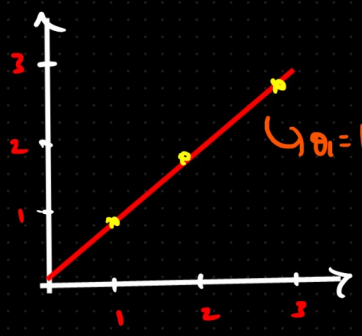
→ Gradient Descent

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

Quadratic Equation

$ax^2 + bx + c = 0$

If $\theta_1 = 1$

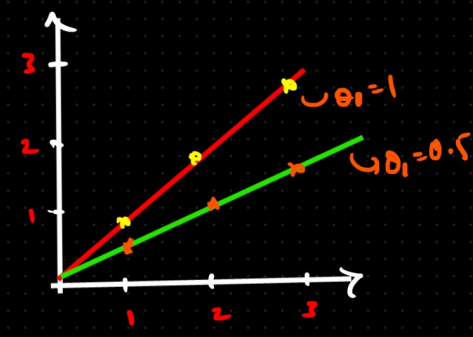


$J(\theta_1) = \frac{1}{2m} (h_\theta(x^i) - y^i)^2$

$= \frac{1}{2\times3} \left[(1-1)^2 + (2-2)^2 + (3-3)^2 \right]$

$= 0$

If $\theta_1 = 0.5$



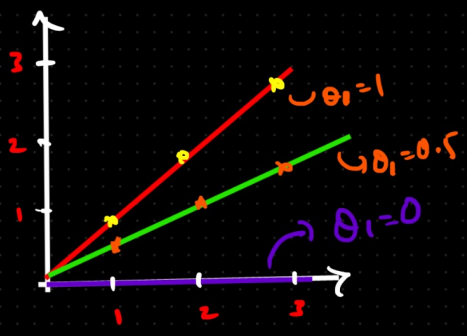
$J(\theta_1) = \frac{1}{2m} \left(h_\theta(x_i) - (y^i) \right)^2$

$= \frac{1}{2\times3} \left((1-0.5)^2 + (2-1)^2 + (3-1.5)^2 \right)$

$= \frac{1}{6} (0.25 + 1 + 2.25)$

$= 0.58$

If $\theta_1 = 0$



$$J(\theta_1) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x) - y^i)^2$$

$$= \frac{1}{2\times 3}\left((0-1)^2 + (0-2)^2 + (0-3)^2\right)$$

$$= \frac{1}{6}(1+4+9)$$

$$= 2.33$$

Gradient Descent (Convergence Algorithm)

repeat until

{

$$\theta_j := \theta_j - \alpha \frac{\partial(J\theta_j)}{\partial(\theta_j)}$$

}

$$\theta_{new} = \theta_{old} - \alpha \frac{\partial(J(\theta_{old}))}{\partial\theta_{old}}$$

$\theta_{new} \simeq \theta_{old} - \alpha$ (-ve)

$\theta_{new} \simeq \theta_{old} +$ (+ve)

$\theta_{new} >> \theta_{old}$

Learning Rate: Learning used to determine the speed of Convergence. Usually learning rate should be very small 0.001

In case of multi-linear Regression your formula will be

$$h_\theta(x) = \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_4 + \theta_0$$

Advantage of MSE

1) It is having one global minima.
2) It is differentiable

Disadvantage of MSE

1) Not robust to outliers
2) MSE Penalizes the error.

Mean Absolute Error

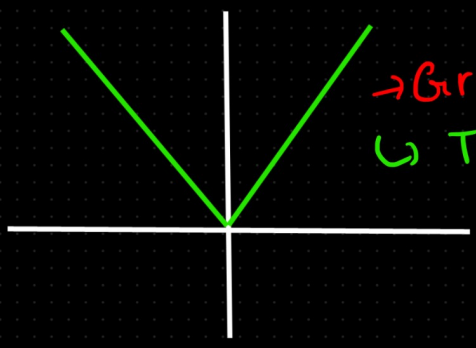
$$J(\theta_0, \theta_1) = \frac{1}{m} \sum_{i=1}^{m} |\hat{y} - y_i|$$

Note: If there is outlier then you can use MSE.

If there is outlier we have 2 option

→ Use MAE

→ Remove Outlier use MSE



→ Gradient Descent for MAE

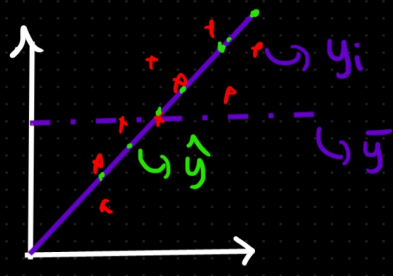
↳ This we called as subgradient

Performance Matrix

There are 2 performance matrix we are using

1) R^2 2) Adjusted R^2

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



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

$= 1 -$ lower value / Higher value

So always performance matrix R^2 lies between 0 and 1.

Important Interview Question:-

Can R^2 be negative value?

Yes it can. and this means the best fit line is worst than average line

Adjusted R^2

House Pricing → Lets Consider

No.of.Room	Total Size	Location	Gender	Price.
–	–	–	–	100

Correlated to Pricew

Not Correlated to price

→ In R^2 it will reduce the performance if there is any un relavant features are there

$$R^2 \text{ Adjusted} = \frac{1-(1-R^2)(N-1)}{N-P-1}$$

N → No. of data point

P → No. of . Predictors

P=2 ≃ R^2=90%. . R^2 Adjusted=86%.

P=3 ≃ R^2=91%. R^2 Adj = 82%

(N-P-1) ↓ denominator.

⇒ If P=2 denominator value will be higher

⇒ If P=3 denominator value will decrease

Lets take an example when P=2

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

$$R^2 adj = 1 - \frac{(1-0.9)(100-1)}{100-2-1}$$

$$= 1 - \frac{0.1 \times 99}{97}$$

$$= 1 - \frac{9.9}{97}$$

$$= 1 - 0.102$$

$$= 0.8979$$

Lets take an example when P= 3

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

$$R^2 adj = 1 - \frac{(1-0.9)(100-1)}{100-3-1}$$

$$= 1 - \frac{0.1 \times 99}{96}$$

$$= 1 - \frac{9.9}{96}$$

$$= 1 - 0.1031$$

$$= 0.8969$$

So always $R^2 >> Adjusted\ R^2$

Conclusion:- we need to check always R^2 and Adjusted R^2 both

If my difference between R^2 and Adjusted R^2 is higher then it means there are many features that are highly correlated.

Linear Regression Assumptions:-

1. There is a Linear Relationship between X & Y
2. Independent features should have Normal distribution
3. Always take care multi colinearity
4. Homoscedacity = All the features have same variance
5. Feature Scaling required

RMSE [Root mean Squared Error] Gradient Descent

$$MSE = \frac{1}{2m} \sum_{i=1}^{m} \sqrt{(h_\theta(x) - y_i)^2}$$

$$RMSE = \sqrt{MSE}$$