Simple Linear Regression

Supervised ML → Regression — O/p or Dependent feature → Continuous

Supervised ML → Classification — O/p " → Categorical

Dataset

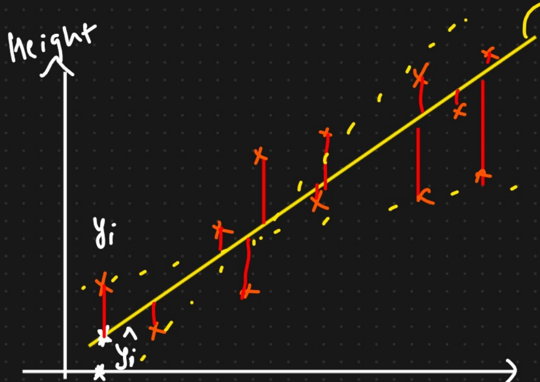
Independent feature ↓ — Dependent feature ↓ — O/p

Weight	Height
74	170
80	180
75	175.5
–	–
–	–
–	–

TRAIN DATASET ↓ Model

new Weight → Model → Height (prediction)

Model ↓ Acc ↑↑



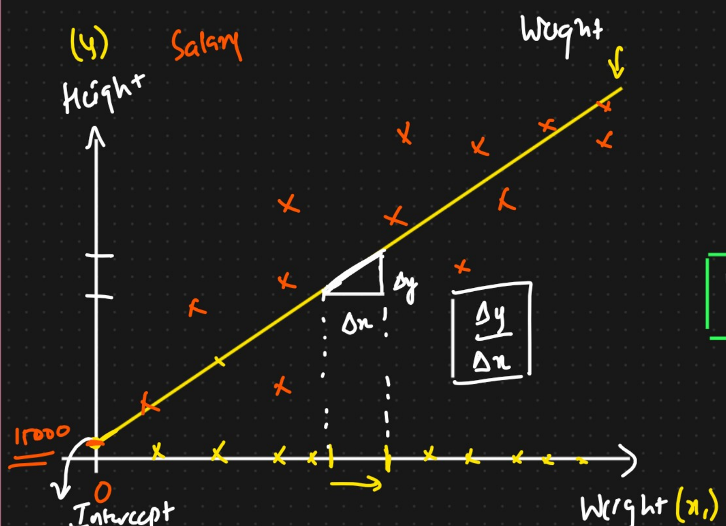
→ best fit line → predictions

Σ Error ↓↓ Minimal

Geometrical Intuition

Simple Linear Regression >>
Cost Func eg sum of square residual
Convergence Alg.
Multiple Linear Reg.
Polynomial Reg.
Performance Metrices >> R Squared and Adjusted R Square
Cost Func >> Error Calculation>> MSE, MAE,RMSE

Type of Linear Reg. >>
Simple Linear Regression
Multiple Linear Reg.
Polynomial Linear Reg
Ridge Linea Reg.
Lasso Linear Reg.
Elastic Net Linear Reg.



$\hat{y} = mx + c$

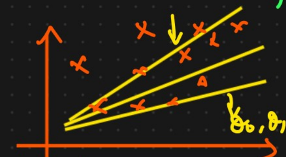
$\hat{y} = \beta_0 + \beta_1 x_1$

$\boxed{h_\theta(x) = \theta_0 + \theta_1 x_1}$

$\theta_0 \Rightarrow$ Intercept

$\theta_1 \Rightarrow$ Slope or Coefficient

Simple Linear Regression
⇑
Model
⇓
Hypothesis Testing



θ_0, θ_1 → change

Error ↓↓

Experience $x_1 \Rightarrow$ Datapoint

Cost function [Error].

$$J(\theta_0, \theta_1) = \frac{1}{n} \sum_{i=1}^{n} (\underset{\substack{\uparrow \\ \text{Actual}}}{y_i} - \underset{\substack{\uparrow \\ \text{predicted}}}{h_\theta(x)_i})^2$$ [Mean Squared Error]

$\uparrow$ $\hat{y}_i$

In order to get the best fit line. we need to minimize cost fun

n: no.of datapoints

y_i = Actual value

$h_\theta(x)$ = predicted value

Final Aim

[In order to get best fit line]

Minimize θ_0, θ_1 $\quad J(\theta_0, \theta_1) = \frac{1}{n} \sum_{i=1}^{n} (y_i - h_\theta(x)_i)^2 \downarrow\downarrow\downarrow$

Dataset

x	y	$h_\theta(x)$
1	1	
2	2	
3	3	

Minimization Error $\downarrow\downarrow$

$h_\theta(x) = \overset{0}{\overset{\downarrow}{\theta_0}} + \theta_1 x_1$

Lets consider $\theta_0 = 0$

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

Let $\theta_1 = 1$ | Let $\theta_1 = 0.5$ | Let $\theta_1 = 0$

Predicted

$x=1$ $\quad h_\theta(x) = 1(1) = 1$ $\quad h_\theta(x) = 0.5$ $\quad h_\theta(x) = 0$

$x=2$ $\quad h_\theta(x) = 1(2) = 2$ $\quad h_\theta(x) = 1.0$ $\quad h_\theta(x) = 0$

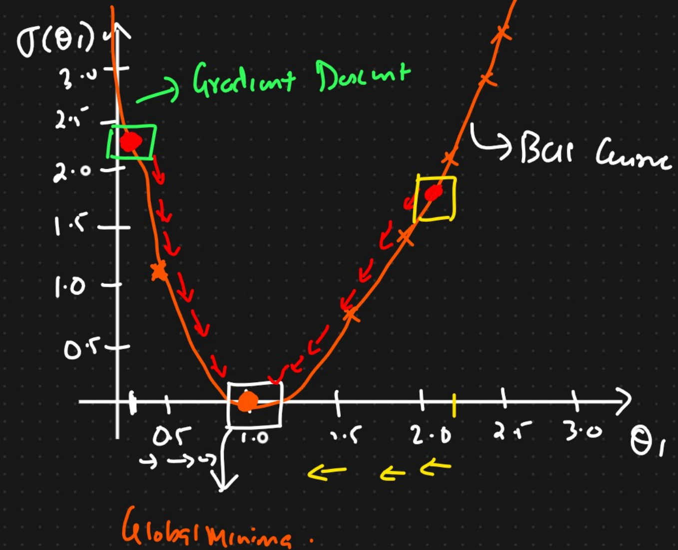
$x=3$ $\quad h_\theta(x) = 1(3) = 3$ $\quad h_\theta(x) = 1.5$ $\quad h_\theta(x) = 0$

Cost fn $\quad \theta_1 = 1$

$$J(\theta_1) = \frac{1}{n}\sum_{i=1}^{n}(y_i - h_\theta(x)_i)^2$$

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

$$J(1) = 0$$



Cost fn $\quad \theta_1 = 0.5$

$$J(\theta_1) = \frac{1}{3}\left[(1-0.5)^2 + (2-1)^2 + (3-1.5)^2\right]$$

$$J(\theta_1) = \underline{\underline{1.16}}$$

Cost fn $\quad \theta_1 = 0$

$$J(\theta_1) = \frac{1}{3}\left[(1-0)^2 + (2-0)^2 + (3-0)^2\right]$$

$$J(\theta_1) = 4.66$$

Convergence Algorithm $\quad$ {Optimize the change of θ_0, θ_1 to Global Minima}.

Repeat until Convergence

{ $\quad j = 0, 1$

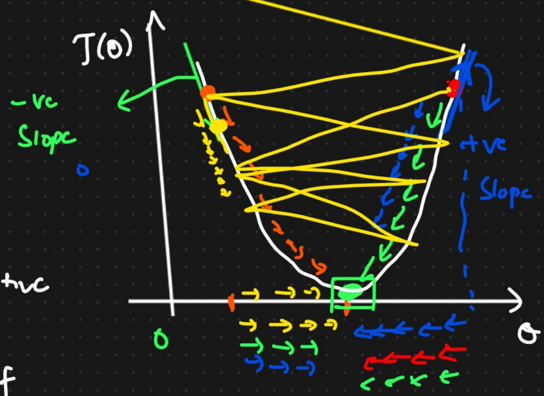
$$\theta_j : \theta_j - \alpha \boxed{\frac{\partial}{\partial \theta_j} J(\theta_j)} \Rightarrow \text{Derivative}$$

⇓

Slope of

}

$\boxed{\alpha = 0.01}$



a point

Learning Rate :

① $\theta_1 = \theta_1 - \alpha(-ve)$

$\theta_{1_{new}} = \theta_{1_{old}} + (value)$

$\boxed{\theta_{1new} >> \theta_{1old}}$

Learning Rate => Speed of Convergence

② $\theta_{1new} = \theta_{1_{old}} - \alpha(+ve)$

$\theta_{1new} = \theta_{1old} - (value)$

$\boxed{\theta_{1new} << \theta_{1old}}$

In the context of machine learning and optimization, a "convergence algorithm" typically refers to an iterative optimization process that aims to find the minimum (or maximum) of a function. This function could represent a loss function to minimize in the case of training a machine learning model.

Here's a general overview of how a convergence algorithm might work:

Initialization: The algorithm starts with an initial guess or estimate of the optimal solution. This could be random or based on some heuristic.
Iterative Improvement: The algorithm iteratively updates the current solution in an attempt to improve it. This involves evaluating the function (e.g., the loss function) and its derivatives (if available) at each iteration to determine the direction and magnitude of the update.
Convergence Criterion: The algorithm continues iterating until a stopping criterion is met. This could be a predefined number of iterations, reaching a certain threshold of improvement, or some other condition.
Termination: Once the convergence criterion is satisfied, the algorithm terminates, and the current solution is considered the optimal solution (or an approximation of it).

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

Mean Squared Error

Conclusion

Repeat until Convergence

{

$$\theta_j : \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$$

}

Simple linear regression is a statistical method used to model the relationship between two continuous variables. It assumes that there is a linear relationship between the independent variable (usually denoted as
X) and the dependent variable (usually denoted as
Y). In simple linear regression, we are trying to fit a straight line to the data in such a way that it best represents the relationship between the two variables.
Aim : the sum of difference between predicted point and actual point should be minimal

Once the model is fitted to the data, it can be used to make predictions about the dependent variable based on new values of the independent variable.

The fitting process often involves techniques like the method of least squares

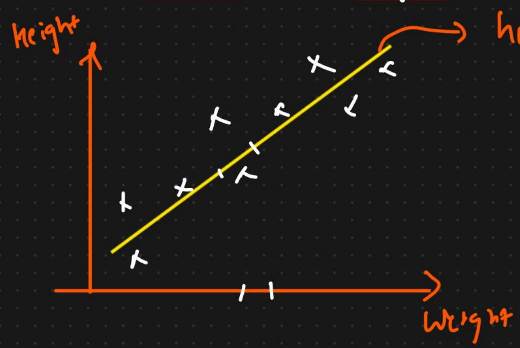
Multiple Linear Regression

Multiple linear regression is a statistical method used to model the relationship between two or more independent variables and a dependent variable. In contrast to simple linear regression, which involves only one independent variable, multiple linear regression allows for the consideration of several predictors simultaneously.

Dataset

I/P → Weight

O/P → Height



$h_\theta(x) = \theta_0 + \theta_1 x_1$

$\theta_0 \rightarrow$ Intercept

$\theta_1 \rightarrow$ Coefficient or Slope

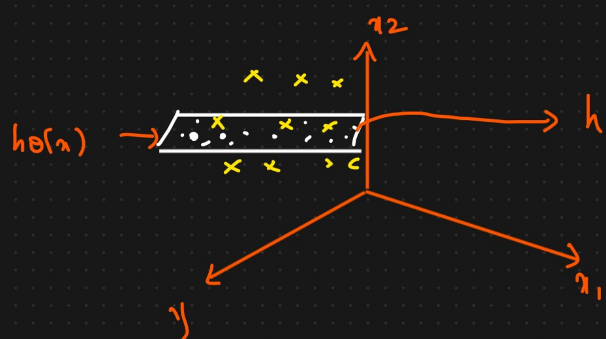
House Pricing Dataset

$\underbrace{\text{No. of Rooms} \qquad \text{Size of Room}}$

x_1 — No. of Rooms

x_2 — Size of Room

y — Price ← Dependent feature



$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2$

$\theta_1, \theta_2 \Rightarrow$ Slope or coefficient

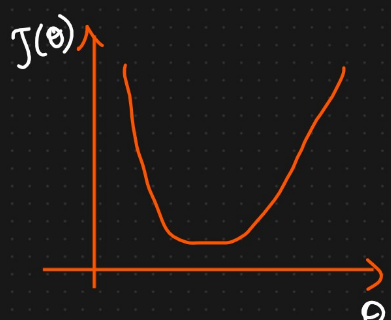
$\theta_0 \Rightarrow$ Intercept

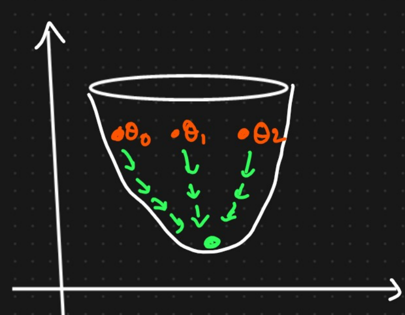
Simple Linear Regression

$$\boxed{h_\theta(x) = \theta_0 + \theta_1 x_1}$$

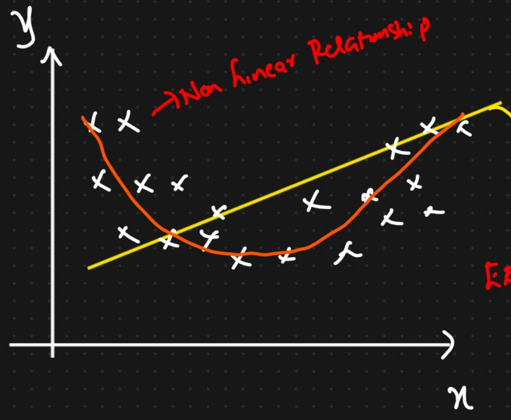
Generic Equation Multiple Linear Regression

$$\boxed{h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \cdots + \theta_n x_n}$$





Polynomial Regression



ERROR ↑↑

Simple Linear Regression

$h_\theta(x) = \theta_0 + \theta_1 x_1$

$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n$

↓

Multiple Linear Regression

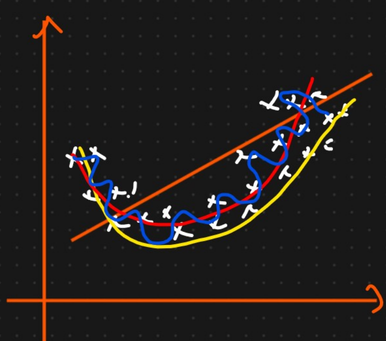
Polynomial Regression → Simple Polynomial Regr, → Multiple Polynomial Regr

⇓

Polynomial Degrees

$h_\theta(x) = \theta_0 + \theta_1 x_1$

Simple Polynomial Regression



polynomial degree = 0 $\quad h_\theta(x) = \theta_0 \times 1 = \theta_0 x_1^0 \Rightarrow$ Constant

polynomial degree = 1 $\quad h_\theta(x) = \theta_0 x_1^0 + \theta_1 x_1^1 \Rightarrow$ Simple Linear Regression

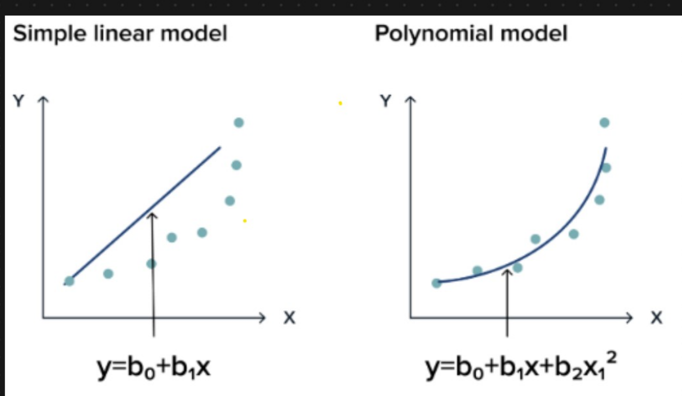
polynomial degree = 2 $\quad h_\theta(x) = \theta_0 x_1^0 + \theta_1 x_1^1 + \theta_2 x_1^2$

$h_\theta(x) = \theta_0 x_1^0 + \theta_1 x_1^1 + \theta_2 x_1^2 + \theta_3 x_1^3$

⋮

polynomial degree = n

$$\boxed{h_\theta(x) = \theta_0 x_1^0 + \theta_1 x_1^1 + \theta_2 x_1^2 + \dots + \theta_n x_1^n}$$



Polynomial regression is a form of regression analysis in which the relationship between the independent variable X and the dependent variable Y is modeled as an n-th degree polynomial function.

② Multiple Polynomial Regression

$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3$ {Multiple Linear Regression}

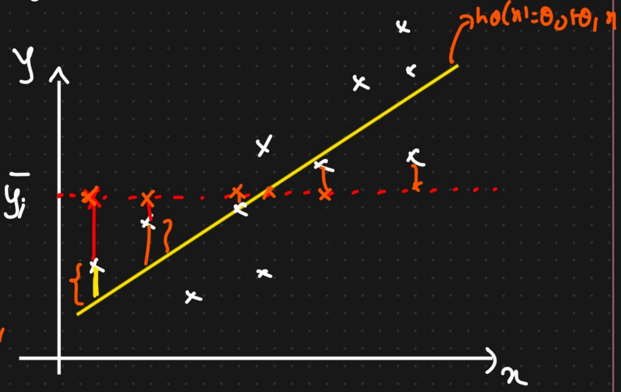
Polynomial Degree = 2

$$\boxed{h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_1^2 + \theta_5 x_2^2 + \theta_6 x_3^2}$$

To check the accuracy of simple linear regression model

Performance Metrics Used In Regression

① R Squared ② Adjusted R Squared



① R Squared

$$R\text{squared} = 1 - \frac{SS_{Res}}{SS_{Total}}$$

SS_{Res} → Error

SS_{Total} {Average of y_i line} → Error

SS_{Res} = Sum of Square Residual {Error}

SS_{Total} = Sum of Square Total

R-squared (coefficient of determination) is a statistical measure used to assess the goodness of fit of a regression model. It indicates the proportion of the variance in the dependent variable that is explained by the independent variables in the model.

$$R\text{ Squared} = 1 - \frac{\overbrace{\sum_{i=1}^{n} (y_i - h_\theta(x))^2}^{SS_{Res}} \Rightarrow \text{Small}}{\sum_{i=1}^{n} (y_i - \bar{y}_i)^2 \Rightarrow \text{Big}}$$

$$\boxed{R\text{squared} = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}}$$

Adjusted R-squared is a modified version of R-squared that adjusts for the number of predictors in a regression model. While R-squared measures the proportion of the variance in the dependent variable that is explained by the independent variables in the model, adjusted R-squared penalizes for the number of predictors in the model, making it a more useful metric when comparing models with different numbers of predictors.

R Squared ranges between 0 to 1

Price

f_0 f_1 f_2 — f_3 f_4 → O/P

R squared = 75% = 0.75

R Squared = 80%

R Squared = 90% ↑↑↑

R squared = 92% ↑↑

② Adjusted R squared

$\downarrow x_1$ $\downarrow x_2$ $\downarrow y$

Gender Size of the house No. of rooms Location → Price

R - squared = 85% = 0.85

R - squared = 90% = 0.90

R - squared = 91% = 0.91

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

N = no. of datapoints

R^2 = R squared

P = No. of Independent features.

$R^2 = 0.8$ N = 11 p = 2

P = 3 → feature

$$\text{Adjusted R squared} = 1 - \frac{[(0.2)(10)]}{11-2-1} = 0.75$$

$$\boxed{R^2 >> \text{Adjusted } R^2}$$

R^2 = 80% Adjusted R^2 = 75%

P=3 R^2 = 85% Adjusted R^2 = 78%

$p = 4$ $\uparrow\uparrow R^2 = 87\%$ Adjusted $R^2 = 76\%$ $\downarrow\downarrow$

$\Downarrow$

Independent feature is not that important

Different type of cost function

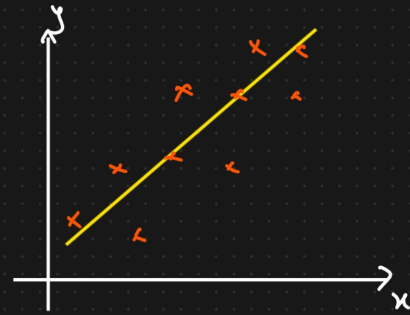
To know about the error

MSE, MAE, RMSE [Cost function]

① Mean Squared Error (MSE) ✓

② Mean Absolute Error (MAE)

③ Root Mean Squared Error (RMSE)



$$J(\theta_0,\theta_1) = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - h_\theta(x)_i\right)^2$$

⇓

Mean Squared Error

① Mean Squared Error (MSE)

$$MSE = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - h_\theta(x)_i\right)^2$$

$$\boxed{MSE = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2} \longrightarrow \text{Quadratic Equation}$$

⇓ → Convex function

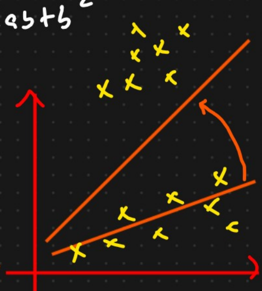
Non Convex function

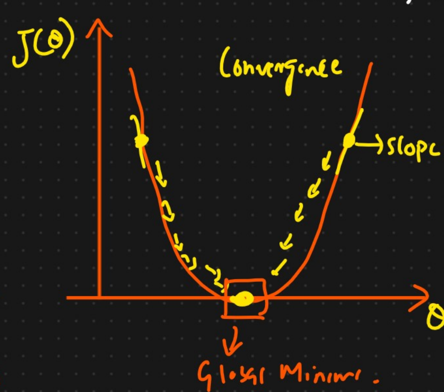
$(a-b)^2 = a^2 - 2ab + b^2$

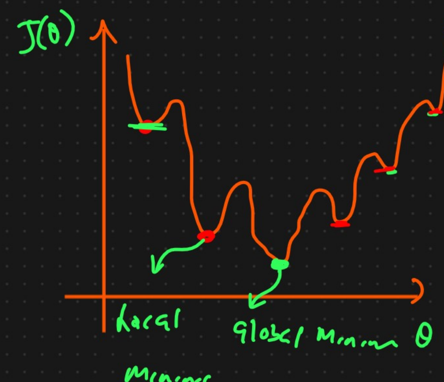
$(\text{ERROR})^2$ ↑↑↑

⇓

penalize the Error







Advantage

① Equation is differentiable

② It has only one local or global minima.

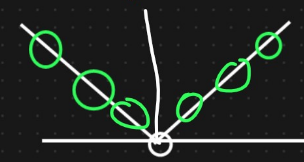
Disadvantage

① Not Robust to outliers

② It is not in the Same unit

② Mean Absolute Error (MAE)

$$\boxed{MAE = \frac{1}{n} \sum_{i=1}^{n} |y - \hat{y}|}$$



Advantage

① Robust to outlier

② It will be in the same unit

Disadvantage

① Convergence usually takes more time

③ RMSE (Root Mean Squared Error)

$$RMSE = \sqrt{MSE}$$

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

Advantages

① Same Unit

② Differentiable

Disadvantage

① Not Robust to outliers.

Note: Linear Regression

Performance Metrics = R^2 and Adjusted R^2 ⇒ Acc of model

Cost function → Error → MSE, MAE, RMSE