Linear Regression

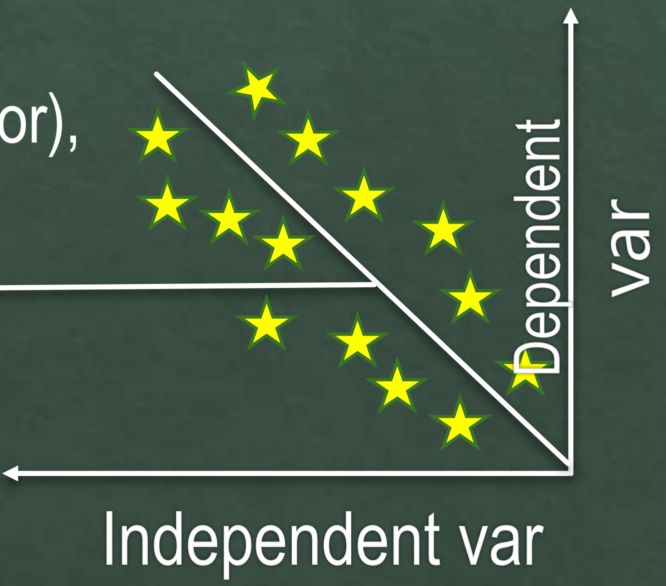
- Linear Regression is a Supervised Machine Learning Algorithm that shows a relationship between a dependent (Target) variable and independent (predictor) variables.
- Linear Regression makes predictions for continuous/real or numeric variables.
- It shows the sloped straight line representing the relationship between the variables.

- Mathematical Equation: Y = m * X + C

  Where Y = Dependent Var (Target) , X = Independent Var (Predictor), m = Slope , C = Intercept

- Types of Linear Regression:
  - Simple Linear Regression.
  - Multiple Linear Regression.

Line of Regression



Dependent var

Independent var

Google Developer Student Clubs
Jamia Hamdard

Simple Linear Regression

When a single independent variable (predictor) is used to predict the value of a numerical dependent variable then such a Linear Regression algorithm is called Simple Linear Regression.

Mathematical Equation : $Y = m * X + C$

Project: Predict the house price from the area.

GitHub Link: https://github.com/AmeenUrRehman/Machine-Learning-Projects-/tree/up-pages/Simple%20Linear%20Regression

Multiple Linear Regression

When more than one independent variable (predictor) is used to predict the value of a numerical dependent variable then such a Linear Regression algorithm is called Multiple Linear Regression.

Mathematical Equation : $Y = m1 + m * X + C$

Project: Predict the house price from the area.

GitHub Link: https://github.com/AmeenUrRehman/Machine-Learning-Projects-/tree/up-pages/Multiple%20Linear%20Regression

SAY LINEAR REGRESSION
ONE MORE TIME
memegenerator.net

Finding the best-fit line

When working with Linear Regression our main goal is to find the best-fit line which means the error between predicted values and actual values should be minimum. The best-fit line will have the least error.

To find the best-fit line we use the cost function here for linear regression we use the mean squared error (MSE) cost function in which the average squared error occurred between the predicted values and actual values.

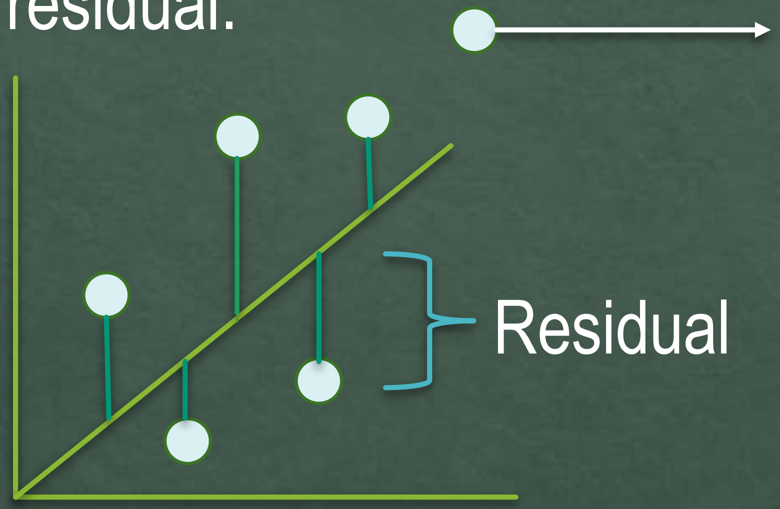
$$MSE = 1 / N * \sum_{i=1}^{n} * ((Ya - Yp)) \wedge 2$$

Where Ya = Actual Value and Yp = (m * Xi + C) = Predicted value

GitHub Link: https://github.com/AmeenUrRehman/Machine-Learning-Projects-/tree/up-pages/Gradient%20Descent

Residuals

The distance between the actual value and the predicted values is called residual.



Outliers: These are the data points that are significantly different from the rest of the dataset.

Optimization: The process of finding the best model out of various models is called Optimization.

Gradient Descent

Gradient Descent is an optimization algorithm used to minimize the Mean Squared Error by calculating the gradient of the cost function.

It is done by random selection of the value of the coefficient and then iteratively updating the values to reach the minimum cost function.

How to achieve gradient descent?

There are multiple methods but we are going to focus on R – Squared for Linear Regression algorithms as it gives the best result for this.

R – Squared method = R – Squared is a statistical method that determines the goodness of fit. It Output result between 0 – 100 Percent so more the value in % less difference between predicted and actual values and hence represent a good model.

Assumptions of Linear Regression :

- Linearity: The Output predicted must have a linear relationship.
- Homoscedasticity: Multiple Linear Regression assumes that the amount of error on the residual is similar at each point of the linear model.
- Non-Multicollinearity: The independent variables should not be highly correlated with each other.

Advantages of Linear Regression Algorithm

- Linear Regression is easy to train and implement.
- It performs well to find the nature of the relationship among the different variables.
- It handles overfitting pretty well using dimensionally reduction techniques, regularization, and Cross-Validation.

Disadvantages of Linear Regression Algorithm

- It assumes linearity between dependent and Independent variables which is rarely represented in real-world data.
- It is sensitive to outliers it is essential to pre-process the data set and remove the outliers before applying linear regression to the data.
- It is prone to multicollinearity.