**LINEAR REGRESSION**

**Table of contents:**

1.introduction to linear regression

2.assumptions of linear regression

3. advantages of linear regression

4. disadvantages of linear regression

5.types of linear regression

6.mathematical intuition to linear regression

7.metrics

8.gradient descent

9.ridge regression

10.Lasso regression

**LINEAR REGRESSION**:

 It is a statistical method that is used for predictive analysis.

Linear regression algorithm shows a linear relationship between a dependent (y) and one or more independent (y) variables, hence called as linear regression. Since linear regression shows the linear relationship, which means it finds how the value of the dependent variable is changing according to the value of the independent variable.

**ASSUMPTIONS in LINEAR REGRESSION**:

There are four assumptions associated with a linear regression model

1 linearity: the relationship between x and y is linear

2 homoscedasticity: the variance of residuals is the same for the any value of x

3 independence: observations are independent to each other

4 normality: for any fixed value of x, y is normally distributed

**Advantages:**

1. Linear regression performs exceptionally well for linearly separable data
2. Easy to implement and train the model
3. It can handle overfitting using dimensionlity

**Disadvantages:**

1. Sometimes Lot of Feature Engineering Is required
2. If the independent features are correlated it may affect performance
3. It is often quite prone to noise and overfitting

**Types of linear regression:**

**1 simple linear regression**

it is used to the mathematical relationship between one dependent variable and one independent variable

y=mx+c

2 **multiple linear regression:**

it is used to the mathematical relationship between one dependent variable and more than one independent variable

y=m1x1+m2x2+m3x3+………….mnxn+c

3 **polynomial regression:**

instead of straight line if we try to fit a line with the curve then it is called polynomial

regression

**MATHEMATICAL INTUTION BEHIND THE LINEAR REGRESSION:**

Step 1: Plot of the Independent & Dependent Variables. Draw the best fit line (Approx.).

Best Fit Line is determined based on the sum of errors to be minimum.

Step 2: Calculate the individual errors.

Error is defined based on the actual & the predicted value. Below is the formula to calculate the error for individual points.

In our Sample Example, we have 3 points which are not in line with our best fit line. So we need to calculate the error function.

Error=y-y^

The calculation of error can conclude the status of selecting the best fit line.

Step 3: Calculating the minimum sum of squares of errors or Ordinary Least-squares.

As we have calculated the individual Error for all the points, we are now going to sum & consider the min value to evaluate the best fit line by using the below formula.

Where

Once we have our Best Fit Line on the Linear Model, we are having the freedom to predict the dependent values based on the given independent variables

**METRICS:**

**1 residual:**

The error between the actual value and predicted value.

We know that linear regression tries to fit a line that produces the smallest difference between predicted and actual values, where these differences are unbiased as well. This difference or error is also known as residual.

Residual=actual value -predicted value

On x- axis we will take the independent variable and residuals are on the y-axis

 if the residuals are randomly s scattered, then your model may perform well.

**2. mean square error:**

it is the average of the squared difference between the predicted and actual value. Since it is differentiable and has a convex shape, it is easier to optimize.

MSE =1/N (y pred -y actual)^2

If the error the is equal to (or) near to zero then it called as a good model

MSE penalizes large errors.

**3.mean absolute errors:**

This is simply the average of the absolute difference between the target value and the value predicted by the model. Not preferred in cases where outliers are prominent.

MAE=1/N |y pred -y actual|

If the error the is equal to (or) near to zero then it called as a good model

MAE does not penalize large errors.

4. **R squared:**

This metric represents the part of the variance of the dependent variable explained by the independent variables of the model. It measures the strength of the relationship between your model and the dependent variable

If the data points are very close to the regression line, then the model accounts for a good amount of variance, thus resulting in a high R² value

If R² is high (say 1), then the model represents the variance of the dependent variable

If R² is very low, then the model does not represent the variance of the dependent variable and regression is no better than taking the mean value, i.e. you are not using any information from the other variables

**5.Root mean squared error(RMSE):**

This is the square root of the average of the squared difference of the predicted and actual value.

R-squared error is better than RMSE. This is because R-squared is a relative measure while RMSE is an absolute measure of fit (highly dependent on the variables — not a normalized measure).

 RMSE is just the root of the average of squared residuals.

RMSE penalizes large errors

6**.Adjusted r squared:**

The main difference between **adjusted R-squared**and R-square is that **R-squared** describes the amount of variance of the dependent variable represented by every single independent variable, while **adjusted R-squared** measures variation explained by only the independent variables that actually affect the dependent variable.

R² tends to increase with an increase in the number of independent variables. This could be misleading. Thus, the adjusted R-squared penalizes the model for adding furthermore independent variables (k in the equation) that do not fit the model.

If adjusted r squared is nearer to one then we say it as good model

**GRADIENT DESCENT:**

Gradient descent is an iterative optimization algorithm to find the minimum of a function.

In linear regression, the model targets to get the best-fit regression line to predict the value of y based on the given input value (x). While training the model, the model calculates the cost function which measures the Root Mean Squared error between the predicted value (pred) and true value (y). The model targets to minimize the cost function.   
To minimize the cost function, the model needs to have the best value of θ1 and θ2. Initially model selects θ1 and θ2 values randomly and then iteratively update these value in order to minimize the cost function until it reaches the minimum. By the time model achieves the minimum cost function, it will have the best θ1 and θ2 values. Using these finally updated values of θ1 and θ2 in the hypothesis equation of linear equation, the model predicts the value of x in the best manner it can.

**Ridge regression:**

it is used to reduce the overfitting and helps to get the less variance, less bias.

In ridge regression the slope values will shrink and never reaches to zero.

**Lasso regression:**

Lasso regression is also used to reduce the overfitting and addition to it.

It also helps in feature selection. when the any features slope value is nearer to zero it will remove that feature considering it as unimportant feature.