LINEAR REGRESSION

1. It is a supervised learning algorithm.
2. Regression analysis is a form of predictive modelling technique which investigates the relationship between dependent and independent variable.
3. Linear regression assumes a linear or straight line relationship between the input variables (X) and the single output variable (y).
4. Output (y) can be calculated from a linear combination of the input variables (X). When there is a single input variable, the method is referred to as a simple linear regression.

|  |  |  |
| --- | --- | --- |
|  | **LINEAR REGRESSION** | **LOGISTIC REGRESSION** |
| CORE CONCEPT | The data is modelled using a straight line | The probability of some obtained event is presented as a linear function of a combination of predictor variables |
| USED WITH | Continuous variables | Categorical variables |
| OUTPUT | Value of the variable | Probability of the occurrence of the variable |
| ACCURACY AND GOODNESS OF FIT | Measured by loss , R squared, Adjusted R squared | Accuracy, Precision, Recall, F1 score, ROC Curve, Confusion Matrix |

1. Formula for Simple linear regression = y = mx+b
   1. Y = Dependent Variable
   2. X = Independent Variable
   3. M = coefficient (slope of the line)
   4. B = constant (intercept)
2. Regression is a method of modelling a target value based on independent predictors. This method is mostly used for forecasting and finding out cause and effect relationship between variables. Regression techniques mostly differ based on the number of independent variables and the type of relationship between the independent and dependent variables.
3. **Simple linear regression** is a type of regression analysis where the number of independent variables is one and there is a linear relationship between the independent(x) and dependent(y) variable.
4. Use of Regression:
   1. Determine the strength of the predictors
   2. Trend Forecasting
   3. Forecasting and effect
5. Linear regression is sensitive to outliers
6. Not computationally expensive
7. Easy to understand
8. Used for evaluating trends and sales estimates
9. Analyzing impact of Price changes
10. The best fit line is the one which as least error. This means less difference between estimated and actual value.
11. R-squared value is a statistical measure of how close the data are to the fitted regression line. It is also known as coefficient of determination or the coefficient of multiple determinations.
12. Generally more the R-square value, more accurate is the model.
13. Low values of R-square is not always bad.