Supervised Machine Learning (Since there are independent and dependent variables to work with, and data is ready to analyze)

Regression analysis:

Irrespective of the type of regression, there will be always one dependent variable.

1. Simple regression – One dependent variable and one independent variable.
2. Multiple regression – One dependent variable and many independent variables.
3. Ordinal regression – Logical (Yes/No)

Key terms:

* Difference between actual and predicted value is called residual.
* Linear is a straight line with the equation y = a + bx (or) mx + c. (m or b is slope, c or a is constant)
* R (Correlation). Example: R^2 = 0.789 for sales = 5\*adv+constant means that sales depends on adv and const by 78.9% (here sales is the dependent variable). R^2 always increases with a addition of a new variable.
* Adjusted R^2 can increase or decrease with an addition of a new variable, and the newly added variable can be described as effective or not respectively.

Assumptions:

* Normality of variables (Skewness,kurtosis,q-q test,shaprio wilk)
* Linearity (Using correlation we check for linearity)
* Independence of errors:

Derbin Watsal Test – 1.5 < d < 2.5 then no auto correlation and vice versa (autocorrelation refers to variables depending on themselves)

* Homoscedasticity ncv () – Variance of errors between independent variables are same.
* Multi collinearity vif() aka variance inflation factor – If vif() > 2, then multi collinearity exists. That is, two or more independent variables have high collinearity. Multicollinearity can lead to skewed or misleading results when a researcher or analyst attempts to determine how well each independent variable can be used most effectively to predict or understand the dependent variable in a statistical model.