Classification

1. Predicting category (input features (independent), predictive(dependent) output
   1. Binary classification (two classes)
      1. Loan approval (Yes / No)
      2. Student pass or fail (Yes / No)
   2. Multi Class Classification (one feedback can be assigned to only one category)
      1. Feedback >> Content, faculty, understanding, duration, support
         1. Eg. Feedback from one student
   3. Multi Label Classification: One single message/Feedback can contain info about all categories.
      1. Content, faculty, understanding, duration, overall.

Classification

1. Logistic regression
   1. If dependent variable is not a number
   2. Still have all independent variables
   3. Classify the dependent variable into classes/groups (0,1), True or False) = binary
   4. Eg. Bank approval
      1. Build model whether to give loan or not
         1. Relationship between variables (affect)
         2. Threshold of the relationship = 0.5
         3. Logistic function is then applied to get probabilities whether he can get approved or not.
            1. Logistic regression = p = 1/1+e-y
         4. Log of the probability of YES vs NO
         5. Pr(y) = event occurring = YES or 1
         6. Pr(~y) = non-event = NO or 0
         7. Goal of Log Regression is to get the best relationship between independent vs dependent (maximum likelihood) to fit sigmoid curive
         8. Advantages
            1. Affect/influence

* Bayes theorem (conditional probabilities) (probability of one event occurring given that the other even already occurred)
* In ML Classification – NB -0 Naiive Bayes Classification
* Condition in stats, in ML it’s Data, more data we have, we get better accurate results.
  + Eg. Student ages in class
  + Data:evidence = X = 18,24,26,35,38,40,46,47,50,56,57,51
* Independence between every pair of features
* Age group vs num of cards
* Age Group vs num of loans
* Age group vs num of defaulters

Metrics

TP( True Positive)

TN (True negative)

FP (False positive) = model predicted wrongly positive

FN ( False Negative) = model predicted wrongly negative

CLASSIFICATION METRICS

1. Accuracy
2. Precision
3. Recall