IIT ROORKEE

NPTEL ONLINE CERTIFICATION COURSE

NAÏVE BAYES

LECTURE 31

DR. GAURAV DIXIT

DEPARTMENT OF MANAGEMENT STUDIES

NAÏVE BAYES

- Complete or Exact Bayes for classification
  - Search for records in training partition having same predictors' values as the new observation to be classified
  - Find the most prevalent class of the outcome variable among the records
  - Assign this class to the new observation
- Class of interest
  - User specified cut off value for the class of interest

NAÏVE BAYES

- Class of interest
  - Search for records in training partition having same predictors' values as the new observation to be classified
  - Find the probability of a record belonging to the class of interest among the records
  - If computed probability value > cut off value, assign the new observation to the class of interest

NAÏVE BAYES

- Concept of conditional probability
  - For an outcome variable with m classes $\{C_1, C_2, \ldots, C_m\}$ and p predictors $\{x_1, x_2, \ldots, x_p\}$, we are interested in the following probability value:

$$P(C_i|x_1, x_2, \ldots, x_p) = \frac{P(x_1, x_2, \ldots, xp|C_i)P(C_i)}{P(x_1, x_2, \ldots, xp|C_1)P(C_1) + \ldots + P(x_1, x_2, \ldots, xp|Cm)P(Cm)}$$

  - Assign the new observation to the class with highest probability value
  - Or, if the probability value for the class of interest > cut off value for the same, assign the new observation to the class of interest

NAÏVE BAYES

- Bayes Model for classification
  - Predictors should also be categorical
  - Numerical variables will have to be converted into categorical variables through binning
- Open Excel
- Complete or Exact Bayes Limitations
  - For a model even with small no. of predictors, many new observations to be classified might not get exact matches
  - Probability of a match might reduce significantly on adding just one variable to the set of predictors

Key References

- Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data by EMC Education Services (2015)
- Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner by Shmueli, G., Patel, N. R., & Bruce, P. C. (2010)

Thanks…