IIT ROORKEE

NPTEL ONLINE CERTIFICATION COURSE

REGRESSION TREES

LECTURE 45

DR. GAURAV DIXIT

DEPARTMENT OF MANAGEMENT STUDIES

CLASSIFICATION & REGRESSION TREES

- Regression Trees
  - Outcome variable should be numerical
  - Steps to build tree model are similar to that of classification trees
  - Prediction step, impurity measures and performance metrics are different
- Prediction step
  - Value of a leaf node is predicted value for a new observation that fell in that leaf node
  - Value of a leaf node is computed by taking average of training partition records constituting that leaf node

CLASSIFICATION & REGRESSION TREES

- Impurity Measures
  - Sum of squared deviations from mean of leaf node
    - Equivalent to squared errors since mean value of leaf node is predicted value
  - Lowest impurity is zero when all the observations that fell in a leaf node have same actual value of outcome variable

CLASSIFICATION & REGRESSION TREES

- Further Comments on CART
  - Can be used as a variable selection approach
  - No variable transformation is required
  - Robust to outliers
  - Non-linear and non-parametric technique
  - Handle missing values

  - Sensitive to sample data changes
  - Predictor's strength as a single variable is modeled and not as part of a group of predictors

CLASSIFICATION & REGRESSION TREES

- Further Comments on CART
  - Might not fit linear structures or relationships between predictors
    - New predictors based on hypothesized relationships can be used
  - Require a large dataset
  - High computation time

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…