



IIT ROORKEE



NPTEL ONLINE
CERTIFICATION COURSE

CLASSIFICATION & REGRESSION TREES PART-4

LECTURE 39

DR. GAURAV DIXIT

DEPARTMENT OF MANAGEMENT STUDIES



CLASSIFICATION & REGRESSION TREES

- CART
 - A data-driven method
 - Based on separating observations into homogeneous subgroups by creating splits on predictors
 - Used for both prediction and classification tasks
 - Model is represented by a tree diagram
 - Easy to interpret logical rules
 - CART algorithm grows binary trees
 - Adoption across domains



CLASSIFICATION & REGRESSION TREES

- Classification Trees
 - Recursive partitioning
 - About partitioning p -dimensional space of predictors using training partition, where p is no. of predictors
 - Pruning
 - About pruning the built tree using validation data



CLASSIFICATION & REGRESSION TREES

- CART example has been discussed in the lecture video



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...

