



IIT ROORKEE



NPTEL ONLINE  
CERTIFICATION COURSE

# CLASSIFICATION & REGRESSION TREES

## LECTURE 36

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

- Recursive Partitioning
  - Partitioning p-dimensional space of predictors into non-overlapping multi-dimensional rectangles
  - The partitioning process is recursive in nature
    - Applied on the results of previous partitions
- Steps for Recursive Partitioning
  - An optimal combination of one of the predictors,  $x_i$  and its value  $v_i$  is selected to create first split of p-dimensional space into two parts
    - Part I:  $x_i \leq v_i$
    - Part II:  $x_i > v_i$



# CLASSIFICATION & REGRESSION TREES

- Steps for Recursive Partitioning
  - Step 1 is applied again on the two parts and process continues to create more rectangular parts
  - The partitioning process continues till we reach pure homogeneous parts
    - All the observations in the part belong to just one of the classes
- Open RStudio



# 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...

