



IIT ROORKEE



NPTEL ONLINE
CERTIFICATION COURSE

CLASSIFICATION & REGRESSION TREES PART-2

LECTURE 37

DR. GAURAV DIXIT

DEPARTMENT OF MANAGEMENT STUDIES



CLASSIFICATION & REGRESSION TREES

- Impurity Measures
 - Gini index and Entropy measure
- Gini Index

For an outcome variable with m classes, Gini impurity index for a rectangular part is defined as

$$gini = 1 - \sum_{k=1}^m P_k^2$$

Where P_k is the proportion of rectangular part observations belonging to class k

CLASSIFICATION & REGRESSION TREES

- Gini Index
 - Gini values lie in the range $\{0, (m-1)/m\}$ for m -class scenario and $\{0, 0.5\}$ for two-class scenario
- Entropy Measure

For an outcome variable with m classes, entropy for a rectangular part is defined as

$$Entropy = - \sum_{k=1}^m P_k \log_2(P_k)$$

CLASSIFICATION & REGRESSION TREES

- Entropy Measure
 - Entropy values lie in the range $\{0, \log_2(m)\}$ for m-class scenario and $\{0, 1\}$ for two-class scenario
- Open RStudio
- Tree diagram or tree structure
 - Each split of p-dimensional space into two parts can be depicted as a split of a node in a decision tree into two child nodes
 - First split creates branches of root node



CLASSIFICATION & REGRESSION TREES

- Two types of nodes in tree structure
 - Decision node: Depicted with a circle
 - Terminal or leaf node: Depicted with a rectangle
 - Correspond to Final rectangular parts
- Steps to classify a new observation using tree based models
 - New observation to be classified is dropped down the tree starting from root node
 - At each decision node, the appropriate branch is taken until we reach a leaf node



CLASSIFICATION & REGRESSION TREES

- Steps to classify a new observation using tree based models
 - At leaf node, majority class is assigned to the new observation
 - For class of interest scenario, proportion of records belonging to the class of interest is compared with the user specified cut off value for the same
- 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...

