# Classification and Regression Trees

Lauren Flanakin

# Outline

- Development of CART
- Definition
- CART Steps
- Visual Explanation
- Advantages/Disadvantages
- Examples
- Review

# Development of CART

- Leo Breiman- as an Applied Statistician, he discovered tree-based methods of Classification that later became machine learning
- Wrote *CART: Classification and Regression Trees* with Jerome Friedman and Richard Olshen in 1984

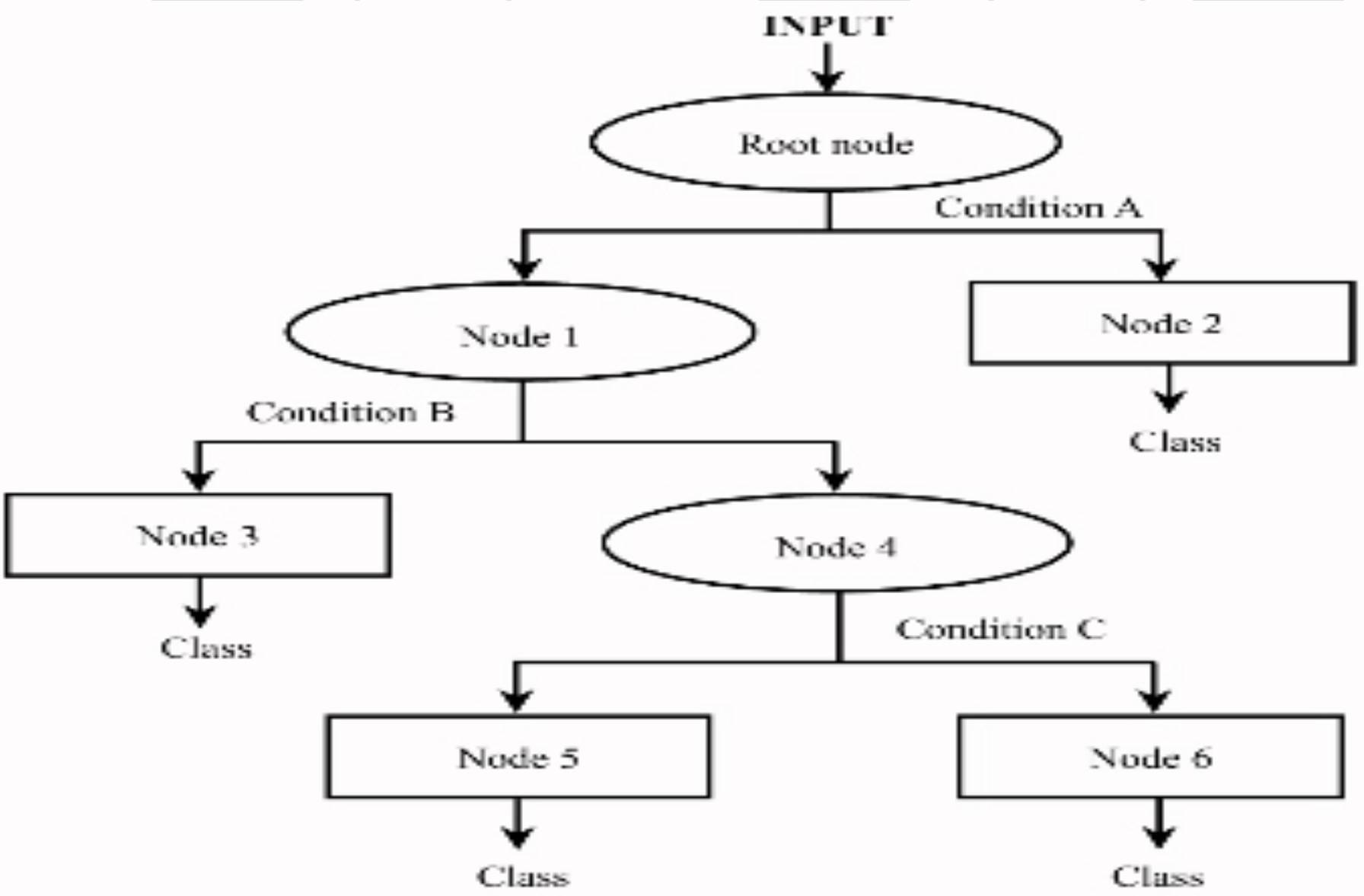
# Definition of CART

- CART: builds classification or regression trees for numeric attributes (regression) or categorical attributes (classification)

# CART Steps

- 1. Start with root node (all data in dataset)
- 2. Split the node with max purity with “Gini”
  - Recursive process
- 3. Assign nodes with predicted classes
- 4. Missing data: program uses best available info to replace missing data (based on a variable that is relative to the outcome variable)
- 5. Stop tree building: when every aspect of the dataset is visible in decision tree
- 6. Tree Pruning: based on miscalculation rate
- 7. Optimal Selection: best tree that fits dataset with a low percentage of error

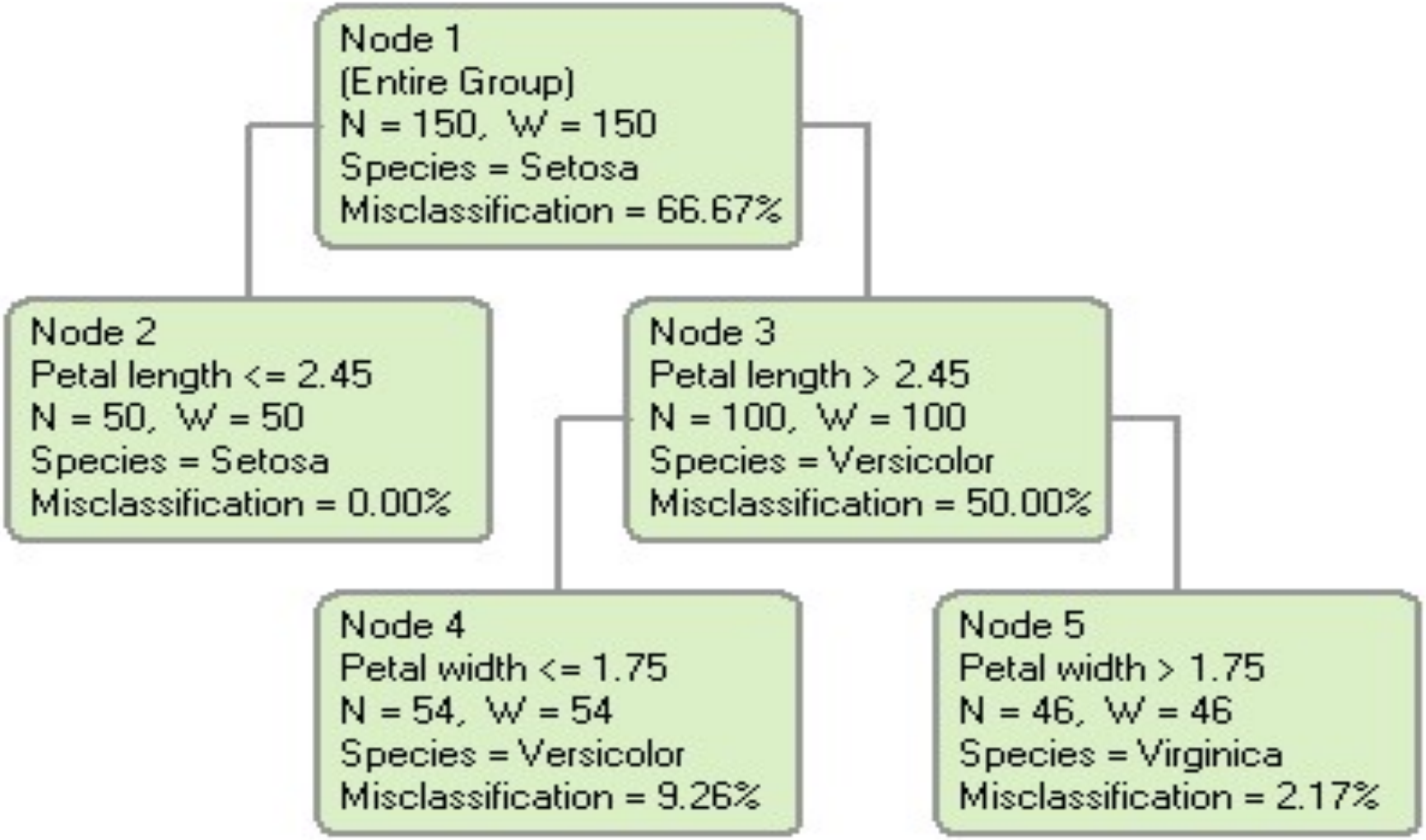
# Visual Example



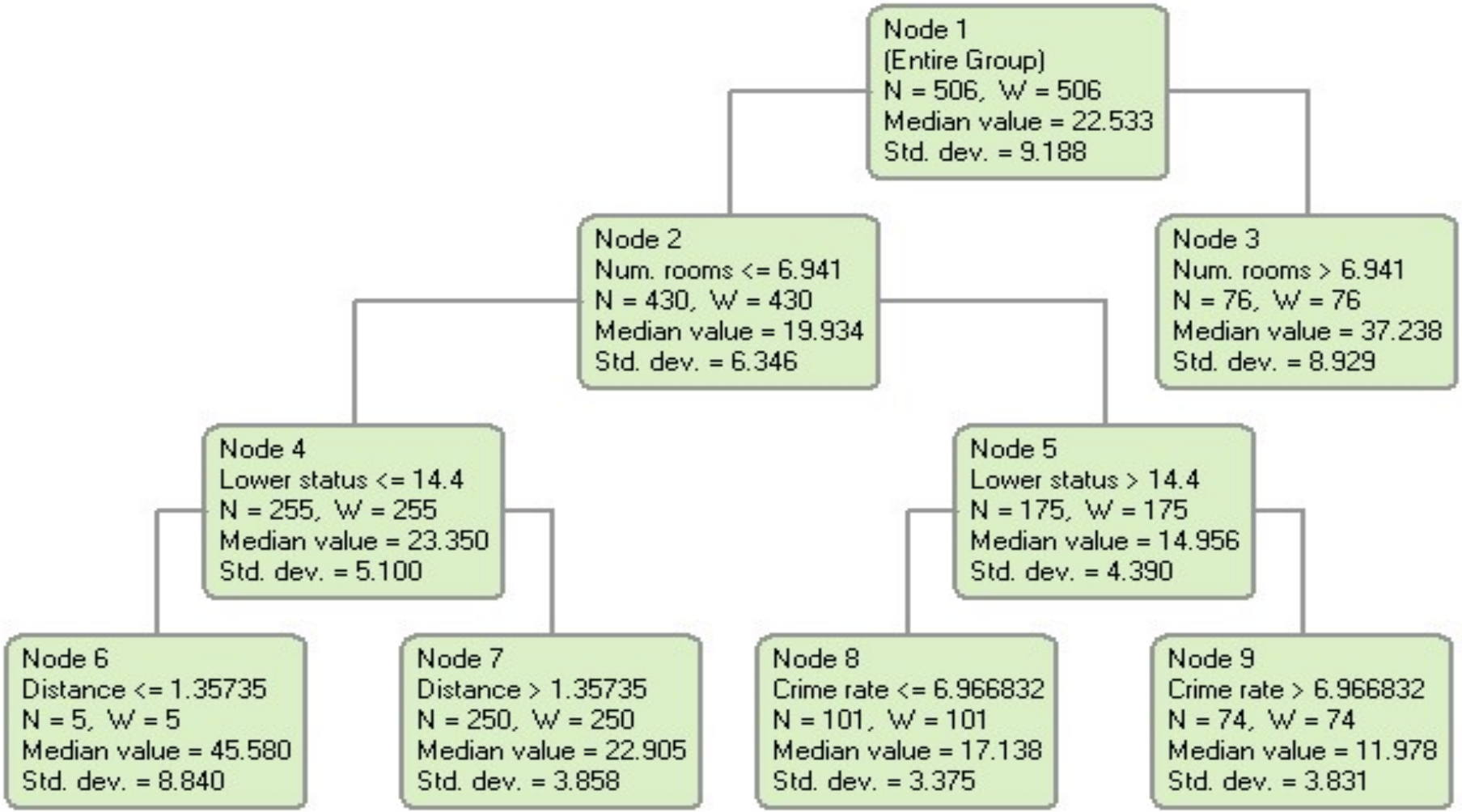
# Advantages and Disadvantages

- Advantages
  - handles data with any structure
  - Machine learning-little input from analyst
  - Final results can be summarized in logical if-then conditions
- Disadvantages
  - Knowing when to stop splitting
  - Computations are complex in determining best split conditions

# Example of Classification Tree



# Example of Regression Tree



# Review

- Development of CART
- Definition
- CART Steps
- Visual Explanation
- Advantages/Disadvantages
- Examples
- Review

# References


- Classification and Regression Trees [Graph illustration of classification and regression trees] DTREG Retrieved from http://www.dtreg.com/classregress.htm
- Statsoft (2008). Classification and Regression Trees. Retrieved from http://www.statsoft.com/TEXTBOOK/stcart.html
- www.wikipedia.com