

CIS-483/683 First Midterm Exam  
Important Concepts and Algorithms

**Concepts**

1. decision tree
2. entropy, information gain, split attribute
3. methods for discretizing attributes with numerical values
4. methods for building a decision tree (compare and contrast them)
  - information gain
  - gain ratio
  - Gini method
5. evaluation:
  - training, tuning, and testing sets
  - stratification
  - holdout method, repeated holdout, n-fold cross-validation, leave-one-out cross-validation, bootstrap method
  - confidence interval
  - confusion matrix
6. building a good decision tree
  - the issue of overfitting
  - pruning
    - prepruning
    - postpruning

**Possible Problems**

1. Build a decision tree
  - (a) information gain
  - (b) gain ratio
  - (c) Gini index
2. Prune a decision tree
  - (a) with a pruning set (reduced error pruning)
  - (b) without a pruning set (pessimistic error pruning)
3. Discretize an attribute with numerical values
  - (a) equal interval binning
  - (b) equal frequency binning
  - (c) global entropy based discretization
  - (d) local entropy based discretization
  - (e) chi-square method

**Homework Assignments:**

Homework-2, Homework-3, Homework-4