Data Mining, CSCI 347, Fall 2019
Outputs of Data Mining, August 30

Supervised learning – classifications (class values) are known and used in the learning
Unsupervised learning – classifications are not used in the learning (they may or may not be known)

Output: Representing Structural Descriptions

- Many different ways of representing patterns
- All can be considered "knowledge" representation
    - Tables (decision tables or Bayesian conditional probability tables)
    - Linear Models (linear regression or logistic regression)
    - Trees
    - Rules (classification or association)
    - Instance-based
    - Clusters

Decision Tables

- Tabular form that presents a set of conditions and their corresponding actions
- Simplest way of representing output (looks very similar to the input) but time consuming to create and maintain large decision tables

Bayesian conditional probability tables

- Simple probabilistic modeling (also called statistical learning)
- Naïve Bayes
    - Assumes that attributes are independent and contribute equally to the prediction
    - Bayes Theorem: Pr[H|E] = Pr[E|H]*Pr[H] / Pr[E]
        - H is the hypothesis
        - E is the evidence
    - Read Pr[X] - the probability that X occurs
    - Read Pr[X|Y] – the probability that X occurs given that the condition Y occurred

Linear models,

- Typically used with numeric attributes. Class value can be numeric or nominal.
- When class value is nominal, call this "logistic regression", Section 4.6
- When class value is numeric, call this "linear regression".
    - Attributes, including class attribute, is numeric
    - Weighted sum of the attribute values
    - For two dimensions, this is the straight line which best fits the points
    - Tricky part is choosing weights

Trees

- "Divide-and-conquer" approach
- Tree built by splitting on attributes, where the number of children is equal to the number of values for the attributes
- The leaves tell the classification
- Attributes typically only get tested once

Rules

- Can be seen as "nuggets" of knowledge

Classification Rules

- Used to predict class, also called "covering rules"
- Antecedent (pre-condition) – IF portion
- Consequent (conclusion) – THEN portion

Association Rules

- Can predict any attribute and combinations of attributes
- Discovering regularities between products in point-of-sale systems, for example {onions, potatoes} => {burger}
- Immense number of possible associations so typically specify the minimum "support" (number of instances predicted correctly) and "accuracy" (number of correct predictions as proportion of all instances to which that rule applies)

Instance-Based Learning

- No structure is learned, given an instance to predict, simply find its nearest neighbor and predict the class of the nearest neighbor, Section 4.7
- Alternatively, find the nearest k-neighbors and predict the class which appears most frequently for those k neighbors

Clusters

- Divide instances into natural groups
- Non-supervised learning as classes aren't being predicted