CS105 Lab 11: Association Learning

Brian Borucki – bborucki@bu.edu

Apriori Algorithm

- Build Item Sets
  - Start with “one item sets”, then construct “two item sets” from these, then “three items sets,” and so on
  - We use 'support' to threshold our item sets
- Form Rules
  - From all of the item sets that we have created, we then form rules
  - We use 'confidence' to select rules which will work well

Apriori Algorithm - Notes

- When combining item sets of length n, we only try to combine sets which have the same n-1 attributes
  - Why?
    - Rules generated any other way either:
      - would be generated this way
      - would not have enough support

- At each step in combining item sets, we always need to check support
  - Why?
    - We've thrown out all item sets we know don't have enough support, but need to check the ones we didn't throw out

Apriori Algorithm - Notes

- incomeRange = 20-30k, creditCardIns = No
  - If incomeRange = 20-30k
    then creditCardIns = No
  - If creditCardIns = No
    then incomeRange = 20-30k

- How many rules do I get from a three item set?

Final Projects

- This project is about creating a model, and using that model to predict some behavior
- You need to be able to say "I am modeling _____"

- If your data set looks like this:
  - <attribute1, attribute2, attribute3, …, attributeN>
- Then the goal of your project is
  - <attribute1, attribute3, …, attributeN> → attribute2

This Lab

- You'll be using the Apriori Algorithm on the dataset from last week
- Once you're done with that feel free to work on your projects
- Please submit answers.txt under lab 11