ML 310 Advanced Machine Learning

Homework Assignment#1 Assigned: 1/27/17 Due: 2/9/17 [100 points]

- Problem Set 1 **Naïve Bayes, ID3, Covering Rules and 1R**

Table 1. Data on customers who purchased a PC at a computer store, showing age group, income, education and whether they purchased a PC earlier.

Age	Income	Education	Prior Purchase	Buys PC?
Young adult	High	College	Yes	Yes
Middle Age	Low	College	No	No
Middle Age	High	High School	Yes	No
Senior	Low	College	No	No
Young adult	Medium	High School	Yes	Yes
Senior	High	College	yes	Yes
Middle Age	Medium	College	No	No
Young adult	Low	High School	No	No
Middle Age	Low	College	Yes	Yes
Young adult	High	College	No	Yes
Senior	Medium	High School	No	No

1. For the above PC purchase dataset given in Table 4 for Problem #7 above, develop a Naïve Bayes Classification method, and use the same to classify the new example given below. [5 Points]

Senior	High	High School	Yes	?

Clearly show all of your calculations, step by step.

2. Using the **ID3 Algorithm** method, construct the full decision tree for the the same PC purchase dataset given in Table 1 above where only Purchase = YES or NO are the 2 possible class targets (classification). Show all the steps of the calculations with Information and Information Gain etc. and of the construction of your decision tree. [10 Points]

3. For the same dataset given above, construct **Covering Rules** using the PRISM algorithm method of rules generation. Use it to classify the new instance below. [10 Points]

Young adult	Low	College	No	?

4. Construct a 1 Rule for classifying the above data for whether a customer will purchase a PC or not. Use the 1 Rule you constructed, to find out the class (Purchase YES or NO?) for the last 2 rows in the Table. (Pretend you do not know the Class (Y/N) answer for those 2 rows and answer this.) [5 Points]

Problem Set 2 **Decision Trees, Association and Classification Rules**

Study Chapter 1 **and Chapters 3 and 4 of** Data Mining: Practical Machine Learning Tools and Techniques (2013) by Witten, I., Frank, E. and Hall, M. A.

Understand how you inspect a data set and build a Decision Tree, look for Classification Rules and Association Rules. Then, answer the below questions. Submit your answers typed as a Word document (or PDF) report. Do not use any data mining software or tool, but use your own work. No programming or implementation is necessary.

5. For the same data set, construct about 10 to 15 Classification Rules which classify the data into a mammal, bird, reptile, fish, and amphibian, insect or invertebrate. [20 Points]
6. For the same data set, construct about 10 to 15 Association Rules which associate the key features with the animal type or other features together. [20 Points]

- Problem Set 3 **KNN and Apriori**

7. Study the KNN Algorithm, and how it is used to do missing value imputation, here. [10 points]

Then, Using the KNN Algorithm, estimate for the missing value (?) for the Lease Prices XLS dataset.

See http://www.si-journal.org/index.php/JSI/article/viewFile/178/134

8. Study the Apriori algorithm data mining and how discovers items that are frequently associated together, for Market Basket Analysis. [20 points]
   http://software.ucv.ro/~cmihaescu/ro/teaching/AIR/docs/Lab8-Apriori.pdf
   page 1 -2 (small dataset from Supermarket: Transaction ID/ milk /Bread /butter /beer/
   Conduct Market Basket Analysis on the same market data set using R. Compare your notes.

   As with HW #1, please expalin the data context and results adequately.