

**IS 698/800: Smart Home Health Analytics**

Homework 2

(Handed Out: February 20, 2017 (Monday), Due: February 27, 2017 (Monday) Before Class)

General Instructions: Use printer paper for your answer sheets. Number each page and write down the total number of pages on the upper right-hand corner of the first page. Turn in the hard-copy before the start of the class. Thanks.

For this assignment you will learn about the NaiveBayes classifier and compare it to the OneR classifier.

1. Consider the simple [loan.arff](#) dataset that provides 18 training examples of whether or not a loan is approved for an applicant based on their income, debt and education. Use WEKA to run the NaiveBayes classifier on the loan dataset. Use the default parameter settings for NaiveBayes, and use the training set as the test option.
  - a. Include in your report the printed results from WEKA. Note that the classifier is output as counts of attribute values by class. Also note that the counts are one higher than in the dataset to avoid any zero probabilities for  $P(\text{attribute}=\text{value}|\text{class})$ .
  - b. What is the percent accuracy of NaiveBayes on the training set?
  - c. How would this NaiveBayes classifier classify the instance: Income=Medium, Debt=Medium, Education=MS? Show your work.
  
2. Use WEKA to run the OneR classifier on the loan dataset. Use the default parameter settings for OneR, and use the training set as the test option.
  - a. Include in your report the printed results from WEKA.
  - b. What is the percent accuracy of OneR on the training set?
  - c. How would this OneR classifier classify the instance: Income=Medium, Debt=Medium, Education=MS? Justify your answer.
  
3. Which learner, NaiveBayes or OneR, has more inductive bias? Justify your answer.
  
4. Perform the following experiment comparing the OneR and NaiveBayes classifiers using default parameter settings on the datasets: loan (from above), contact-lenses, iris, labor, segment-challenge, soybean, weather, and weather.nominal (these 7 come with WEKA).
  - a. In WEKA, choose the "Experimenter" application.
  - b. Under "Experiment Configuration Mode" choose "New".

- c. Under "Experiment Type" choose "Train/Test Percentage Split (data randomized)".
- d. Under "Datasets", choose "Add new..." for each of the 8 datasets. Note that the segment-test dataset has been removed from the list.
- e. Under "Algorithms", choose "Add new..." for OneR and again for NaiveBayes.
- f. Now click on "Run" at the top, and then click "Start".
- g. Now click on "Analyse" at the top.
- h. Under "Source", choose "Experiment".
- i. Near the bottom, click on "Perform test".

Include in your report a table giving the percent correctly classified instances in the test split for both classifiers on each dataset.

5. Compare the performance of the OneR and NaiveBayes classifiers based on the results from the previous problem. Specifically, which classifier performs better on which datasets and why. The "why" part should consider the characteristics of the data, the hypothesis space, and the learning algorithm.
6. Turn in your nicely-formatted printed report (PDF preferred) containing your responses to the above problems in class.