

**IS 709/809: Computational Methods for IS Research**

Homework 4

(Handed Out: April 6, 2016 (Wednesday), Due: April 13, 2016 (Wednesday) Before Class)

General Instructions: Use printer paper for your answer sheets. Use blue or black ink. Number each page and write down the total number of pages on the upper right-hand corner of the first page. Thanks.

For this assignment you will familiarize yourself with the WEKA Machine Learning Software, to test various learning algorithms. You will also analyze a particular learning problem.

1. Download and install WEKA on your preferred platform. WEKA is available [here](#). Be sure to get the latest Developer version (3.7.13). WEKA is already installed on the machines in ITE 4XX.
2. First, we will analyze the contact-lenses dataset provided with WEKA (also available [here](#)).
  - a. Run WEKA and select the Explorer application.
  - b. Under the Preprocess tab, select "Open file..." and open the contact-lenses dataset. You should now see information about this dataset.
  - c. For each feature used to describe an instance, give the name of the feature, its type, and the number of possible values.
  - d. Give the size of the instance space. Justify your answer.
  - e. What is the class feature and its possible values?
  - f. Under the "Classify" tab, click on "Choose" and select the OneR classifier. OneR chooses one feature and attempts to use it for classification. Also, click on "OneR" to bring up its parameter list and change "minBucketSize" to 1.
  - g. Under "Test options" select "Use training set".
  - h. Click "Start" to run and retain the output.
  - i. Which feature did OneR select for its classifier?
  - j. How many errors did this classifier make on the training set?
3. Second, we will analyze the weather dataset provided with WEKA (also available [here](#)). Perform the same steps and answer the same questions/requests for this dataset as you did in the previous problem.
4. Finally, we will analyze a dataset based on the "family car" example in class.
  - a. Create a dataset in the ARFF format used by WEKA, with two numeric features ("price" and "engine-power") and one discrete class feature ("family-car") whose values can be "yes" or "no". The actual data is shown below.

Price	Engine Power	Family Car
7000	310	no
8000	180	no
14000	200	no

15000	280	yes
20000	250	yes
20000	340	no
21000	190	no
22000	300	yes
25000	260	yes
27000	285	yes
29000	340	no
30000	210	no
39000	160	no
40000	245	no
41000	285	no

- b. Perform the same WEKA test as before: Classifier "OneR -B 1", Test option "Use training set". Retain the output.
  - c. Which feature did OneR select for its classifier?
  - d. List the examples that this classifier incorrectly classified (if any).
5. Submit the hardcopy to me clearly addressing the following.
- o Nicely-formatted document (MSWord or PDF) showing the raw output from each of the three WEKA runs and your answers to all the above questions/requests.
  - o Your printed ARFF file for the family car dataset.