Assignment 4 – COMP 135 – Spring 2016

Due: Mar 15 **Turn in hardcopy in class**

1. Compare and contrast filter and wrapper methods, focusing on wrapper methods. What are the pros and cons of each? Include at least 1 advantage and at least 1 disadvantage of wrapper methods.

2. The Iris dataset is a common example dataset in machine learning. Features encode measurements of flower geometry, and classes are species of Iris. Use both correlation and mutual information to rank the features (as one would do for feature selection via filter methods). Show the rankings for each heuristic and describe/show how they were calculated.

The data is available here: https://archive.ics.uci.edu/ml/datasets/Iris

3. Make an *Iris versicolor* detector! Using the Iris dataset linked above, convert the class labels to binary (versicolor = true; both setosa and virginica = false) and use Z-normalization on the features to ensure mean=0 and standard deviation=1.

Now define and train a linear threshold unit to classify these labels. Use the delta-rule with a single linear threshold unit (LTU's threshold=0), a learning rate of 0.1, and initialize all weights to 0.5. Introduce a constant input, such that there are 5 inputs to the LTU, where the last input is 1 for all 150 of the instances in the Iris dataset (i.e. the first row should read: 5.1, 3.5, 1.4, 0.2, 1)

Show the progression of weights for 10 iterations.