Homework #7 – Due April 27

Use the iris dataset from the UCI repository to answer the questions below using the Weka Explorer.

1. Start by removing all attributes except sepal length, petal length, and class, and create a training dataset and a test dataset as described in step 2.

2. Using the filter Remove (unsupervised), remove 30% of the instances to generate your training dataset. Using the full iris dataset generated at step 1, apply the filter Resample (30%) with randomSeed=7 to generate your test dataset.

3. Use Instance Based learning scheme IBk for k=1 and k=3 (both with no distance weighting and with weight =1/distance). Report on the confusion matrix obtained in each case and explain the role of k and the difference between no distance weighting and weight =1/distance.