Statistics 503 Exam 2 Guide 2005

Open notes

1. The following code fits a neural network model to the data:

   It results in these missclassification tables:

   giving training and test error of zero. Why is the test error so small?

2. The following missclassification table for the test data show the results of running SVM to classify cancer patients.

   Compute the overall test error. Compute the test error separately for normal and cancerous tissue. Whats wrong with the classifier?

3. Match up the clusters as best possible for the two clustering methods in this confusion table.

4. Circle the results of k-means clustering on the SOM map.

5. We generate data from two concentric circles which form two clusters, and fit a radial kernel SVM classifer. Cases 1-15 are in class 1, the inner circle, and cases 16-40 are in class 2, the outer circle. Identify which points are the unbounded support vectors from this model fit.

6. In the music data, and the movies data that we've looked at this semester, several of the variables were highly skewed. In the plot below, we have conducted cluster analysis on both raw data where the variables have skewed distributions and the logged data. Why do the results differ? What is the impact of skewed data on the results of cluster analysis? Which result do you prefer in this example?

7. We constructed a simulated data set to test the effect of the number of variables relative to the number of cases and model complexity. The data contains separated classes in the first variable (X_1) but all other variables ($X_2 - X_{50}$) are samples from Gaussian (normal) distributions, all similar to X_2 plotted below. There are 20 observations.

   (a) We fit the SVM model to variables X_1, X_2. Three cases are chosen as support vectors. Write down the SVM model for this data.

   (b) Following this we fit the SVM model for increasing numbers of variables. This table shows the number of support vectors as number of variables increases. Why do you think the number of support vectors change?