SVM classifier lab

In the Matlab statistics toolbox, there are functions svmtrain and svmclassify with fairly simple user interfaces and plotting tools that we will use to gain experience in how SVM works. The toolbox also contains a text on understanding SVM classifiers.

Since the basic formulation of SVM is restricted to 2 classes, we will start working with that.

Let us work with one of the common data sets in pattern recognition, the IRIS data set. This can be loaded by 'load fisheriris' and gives 150 samples from 3 classes with 4 features (see the documentation in matlab or on searching for 'iris data' on the web).

1. Partition the 150 samples into a training set and test set of the same size. Hint: use the function cvpartition(150,'holdout,'75)

2. Study selected scatter plots to find 2 classes and 2 features where the 2 classes are reasonable well separated. First select 2 classes that are reasonably well separated with a linear SVM.

3. Use the function svmtrain with a linear kernel and the option 'showplot' to plot the features, the support vectors, and the decision boundary.

4. Experiment with the cost of misclassification C (in this library this is called 'boxconstraint'). Which effect does changing the value have on the boundary and the support vectors?

5. Then use the function svmclassify to classify the test set. Compute the classification accuracy.

6. Now experiment with training a radial basis function kernel. Use the default value of C, but experiment with the parameter rbf_sigma. What effect does changing it have?

7. If you don't us the 'showplot'-option, you can use more that 2 features. Try that.

8. Generalize to multiclass by using the set of pairwise classifiers strategy and choosing the most frequenst class label.

You will find more data examples on using SVM in the matlab documentation (doc svmtrain). Scroll to the end of this help and select 'How to ->Support Vector Machines'. Study the example with two overlapping Gaussians.