INF5300 – spring 2015 | SVM classifier lab

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

Take as a starting point the script below:

```
load fisheriris  % Load the data
X = [meas(:,1), meas(:,2)]; % Select two features; here feature 1 and 2
Y = nominal(ismember(species,'setosa')); % Get an indicator vector for the
Setosa class
P = cvpartition(Y,'Holdout',0.30); % Randomly partitions observations into a
training set and a test set using stratified holdout

% Use a linear support vector machine classifier
svmStruct = svmtrain(X(P.training,:),Y(P.training),'showplot',true,
'boxconstraint', 1); % Note that 'boxconstraint' is our 'C'

% Run the trained SVM classifier on the test data
C = svmclassify(svmStruct,X(P.test,:),'showplot',true);
errRate = sum(Y(P.test)~= C)/P.TestSize  %mis-classification rate
conMat = confusionmat(Y(P.test),C) % the confusion matrix
```

It loads the «IRIS» dataset, selects two out of the four features, and builds and tests a linear SVM classifier.

Run the script and make yourself familiar with it.

Experiment with changing the cost of misclassification (or the cost of the slack variables), C (or 'boxconstraint' in the script). What is the effect of changing this variable?

Now, use a radial basis kernel (RBF kernel) instead to allow for non-linear decision boundaries. That is, include «'kernel_function', 'rbf'» in the svmtrain function. Experiment with changing the σ parameter ('rbf_sigma'). What is the effect of changing it?

Implement a grid-search of the C and σ parameters based on 10-fold crossvalidation.