CAP 5610: Introduction to Machine Learning
Homework 3

HW 3 is due Sunday November 24, at Midnight. Check the policy on Homeworks.

Chapter 10. Logistic Regression and gradient descent

Alpaydin, Exercises: 3,7

Chapter 13. Kernel Machines

1)[From Marsland '09] Supposte that the following are a set of points in two classes:

class 1: $\begin{pmatrix}1\\1\end{pmatrix}$, $\begin{pmatrix}1\\2\end{pmatrix}$, $\begin{pmatrix}2\\1\end{pmatrix}$

class 2: $\begin{pmatrix}0\\0\end{pmatrix}$, $\begin{pmatrix}1\\0\end{pmatrix}$, $\begin{pmatrix}0\\1\end{pmatrix}$

Plot them and find the optimal separating line. What are the support vectors, and what is the margin?

2) [Adapted from Marsland '09] Use Support Vector Machines (For example `http://scikit-learn.org/stable/modules/svm.html`) and apply it to the *wine* dataset `http://archive.ics.uci.edu/ml/datasets/Wine`. Compare the results to using Logistic Regression (`http://scikit-learn.org/stable/auto_examples/linear_model/plot_iris_logistic.html`). Do the same for the yeast dataset.

3) [Adapted from Melanie Mitchel '10] Using the optdigits dataset `http://archive.ics.uci.edu/ml/machine-learning-databases/optdigits/` do the following:

a) For each digit, 0-9, train an SVM, using a linear kernel, and test each SVM on its corresponding test data. Take the instances of digit i from optdigits.train as your positive training examples, and instances of digits $j \neq i$ in the same file as your negative training. Make sure that your training and test data have approximately balanced numbers of positive and negative examples. For each model and corresponding test set, report:

- How many support vectors were used;
- How many test examples were missed;
- Choose one of the digits, for which a small number of test examples were missed, and include the images of the test examples that were missed for that digit, along with two or three test examples that were classified correctly, and speculate, if you can, on why the missed ones might have been classified incorrectly.

b) Choose the digit that seems the hardest for the linear SVM to correctly classify, and experiment with SVMs using polynomial kernels. Report the number of support vectors and performance on the test data for SVMs using a polynomial kernel with d = 1, 2, 3, 4, 5. You will (likely) see the accuracy improve with increasing d, but when d becomes too high, the accuracy will drop.

c) Repeat b) using radial basis function kernel and report number of support vectors and accuracy. Follow
`http://scikit-learn.org/stable/auto_examples/svm/plot_rbf_parameters.html`
to find good values for the C and γ parameters.

Chapter 15. Hidden Markov Models

1) Alpaydin, Exercises: 1,2.

2) The following is a 4 state HMM that with balls of three different colors, Green, Blue, and Red. These are the parameters:

State	Initial Probabilities	P(R)	P(G)	P(B)
1	.4	.5	.3	.2
2	.2	.3	.4	.3
3	.2	.2	.5	.3
4	.2	.2	.2	.6

Assuming the following state transition matrix

.3	.3	.2	.2
.3	.3	.3	.1
.1	.3	.3	.3
.1	.2	.3	.4

For the above HMM, solve:

a) Consider the following observation *rbggb*. Find the probability of observing this sequence.

b) For the observation *rbggb*, find the optimal state sequence corresponding to this observation.