STAT 675 – Homework 3
Due: Oct. 23

1. Let's look at the digits data with sparse logistic regression, LDA, and (kernel) support vector machines.

   (a) Download the digits data (`http://yann.lecun.com/exdb/mnist/`) and read the website for relevant information about the dataset.

   (b) We wish to do two-class classification in two cases. Apply the above mentioned techniques to these data sets.

      i. Comparing 3's to 5's
      ii. Comparing 4's to 9's.

      Be sure to plot the results of LDA on the plane $\mathcal{H}_1$. Try different kernels for the SVM.

   (c) By looking at $\mathcal{H}_1$, do you feel a kernelization using these coordinates as data would improve the classification rate.

   (d) Which of the methods has a better test error rate?

   (e) State how you could kernelize LDA into a nonlinear method. Note: It is clear how to kernelize it through a feature map. This might be very expensive, however. Can you kernelize it through inner products instead?

2. Show the following (using notation from lecture)

   (a)
   $$\mathbb{P}(Y \neq \hat{g}(X)|X = x) = 1 - (\mathbf{1}m(x) + (1 - \mathbf{1})(1 - m(x))),$$
   where $\mathbf{1} \equiv \mathbf{1}_{g(x)=1}(x)$.

   (b)
   $$\mathbb{P}(Y \neq \hat{g}(X)|X = x) - \mathbb{P}(Y \neq g^*(X)|X = x) = 2|m(x) - 1/2|\mathbf{1}_{g^*(x)\neq\hat{g}(x)}(x)$$

   (c) Use these results to show that g^* is the Bayes' rule.