Statistics 315a
Homework 4, due Wednesday March 14, 2012.

1. Write a function in R to implement the Naive Bayes classifier for categorical features (factors), using a barplot density estimate for each feature, and an arbitrary number of classes. Try to organize your code efficiently, and use vector/matrix operations wherever possible (and avoid loops). You will compute the training and test misclassification of your classifier on the spam data. Since the spam features are quantitative, write a function to convert each feature to a binary factor as a pre-processing step. A resonable choice for this is to threshold at the median for each feature (using the median on the training data).

2. *Semi-parametric Linear Model.* Consider an additive model $y_i = x_i^T\beta + f(z_i) + \epsilon_i,\ i = 1, \dots, n$. Here x_i is a vector of p predictors (including the element 1), and z_i is an additional confounding predictor. You plan to fit this model by penalized least squares, using a smoothing spline to control the roughness of f (β will be unpenalized).

   (a) Write down the penalized least-squares criterion for fitting this model.

   (b) Show that the minimizing $\hat\beta$ and $\hat f$ satisfy the following pair of estimating equations:

   $$\begin{aligned} \mathbf{X}\hat\beta &= \mathbf{H}(\mathbf{y} - \hat{\mathbf{f}}) \\ \hat{\mathbf{f}} &= \mathbf{S}_\lambda(\mathbf{y} - \mathbf{X}\hat\beta), \end{aligned}$$

   where $\mathbf{f}$ is the vector of n fitted values for the function f, and $\mathbf{S}_\lambda$ is the appropriate smoothing spline operator matrix. What is $\mathbf{H}$?

   (c) Assume the errors ϵ_i are iid with mean 0 and variance σ^2. Give expressions for the covariance matrices of $\hat{\mathbf{f}}$ and $\hat\beta$.

   (d) Show that you can solve these equations explicitly for $\hat\beta$, and hence for $\hat f$ as well.

   (e) The smoothing spline operation $\mathbf{S}_\lambda\mathbf{r}$ for any vector $\mathbf{r}$ can be computed in $O(n)$ operations. Discuss how you might organize the computations in the previous item in an efficient manner. What is the order of computations for your entire solution.

3. Read section 12.7 of ESL on mixture discriminant analysis. Then do problem ESL 12.11. (Note: in MDA, there is also the possibility of doing subspace reduction; here we will concentrate on the full-dimensional problem)

12.11(d) Suppose in addition to the N labeled data pairs $\{x_i, g_i\}_1^N$, you have an additional M observations on X: $\{x_{i'}\}_1^M$. Show how you can modify your EM algorithm to accomodate these unlabeled data. What log-likelihood are you maximizing?