ENGN 2520 / CSCI 1950-F Homework 2
Due Friday February 15 by 4pm

Problem 1

In this problem we consider binary classification under a non-uniform loss function.

Let X be a finite input space and $Y = \{0, 1\}$ be two possible labels that can be associated with an example. Let $p(x, y)$ be a distribution over labeled examples.

Let $L(a, b)$ be a loss function specifying a cost for assigning label a to an example whose true label is b. We have $L(0, 0) = L(1, 1) = 0$, $L(0, 1) = 1$ and $L(1, 0) = 10$. That is, the cost of assigning label 1 when the true label is 0 is much higher than the cost of assigning label 0 when the true label is 1.

Derive the classification rule $c : X \to Y$ that minimizes the expected value of the loss function $E[L(c(x), y)]$. Simplify your answer as much as possible. The optimal classifier should be defined in terms of $p(x, y)$ or quantities that are defined in terms of $p(x, y)$, such as $p(y)$ and $p(x|y)$. Justify your answer.

Problem 2

(a) Let $x = (x_1, \ldots, x_n)$ be a random vector where each x_i is a binary random variable. The n random variables are independent, with x_i distributed according to a Bernoulli distribution with mean u_i. This leads to a distribution $p(x|u)$ over the random vectors that is parameterized by a vector of parameters $u = (u_1, \ldots, u_n)$.

Suppose we have a training set T with k independent samples from $p(x|u)$. Derive the maximum likelihood estimate for u. Justify your answer.

Problem 3

Let x be a real valued random variable with a uniform distribution $p(x)$ on some unknown interval $[a, b]$. Suppose we have a training set T with k independent samples from $p(x)$. What is the maximum likelihood estimator for $[a, b]$? Justify your answer.

Problem 4

In this assignment you will implement a naive Bayes classifier to recognize handwritten digits. There are 10 classes y corresponding to digits 0 through 9. Each example x is a 28x28 binary image represented as a 784 dimensional binary vector.

The data for this problem is available on the course website. The training examples are loaded into matrices "train?" where "?" is a digit and the test examples are similarly loaded into matrices "test?". Each matrix has one example per row and the examples can be reshaped into a 28x28 matrix for visualization as follows.

> load('digits');

> A = reshape(train3(43,:),28,28)';

> image(A);

Under the naive Bayes model we assume the features (pixel values) are independent conditional on the image class. We will use a different Bernoulli distribution to model each feature (pixel) of each class. Note that in this case $p(x|y)$ is a product of Bernulli distributions like in Problem 2.

Let $u_{y,i}$ denote the mean of the Bernoulli distribution associated with the i-th feature (pixel) of class y. You can assume that the $p(y) = 1/10$ for each class y.

(a) What is the maximum likelihood estimate for the parameters $u_{y,i}$ given training data $\{(x_1, y_1), \ldots, (x_n, y_n)\}$?

(b) Use ML estimation to train a model for each digit using the training data from the course website. Make a visualization of the model for each digit by drawing a 28x28 image where the brightness of each pixel specifies the mean of the Bernoulli distribution associated with that pixel.

(c) Use the resulting models to classify the test data. What fraction of the test digits were correctly classified? Compute a 10x10 confusion matrix where entry (i,j) specifies how often digit i was classified as digit j.

You should turn in a writeup that includes the answers to the questions above (including the visualization of the models you learned) and a printout of your source code.