ENGN 2520 Spring 2024
Homework 3

Problem 1

We want to classify fish into two types: bass and salmon. We have 2 real valued features: length and weight.

Suppose we assume the length and weight of a fish are independent conditional on the fish type, and each is distributed according to a (different) gaussian distribution that depends on the fish type.

We estimate the parameters of

- $p(length|type = bass)$,
- $p(weight|type = bass)$,
- $p(length|type = salmon)$,
- $p(weight|type = salmon)$,
- $p(type = salmon)$,
- $p(type = bass)$,

using maximum likelihood estimation with a training set T.

(a) What is the classification rule defined by the Bayes optimal classifier in terms of the six distributions specified above. You should use a 0/1 loss to derive the classifier.

(b) Suppose the resulting classifier has poor performance on a test set. Give 3 different reasons why this might happen in practice.

Problem 2

Suppose we have an input space $X = \mathbb{R}$ and a label space $Y = \{0, 1\}$. Let $p(x, y)$ be a probability density defined as follows.

$$p(y = 0) = \frac{1}{3}$$

$$p(y = 1) = \frac{2}{3}$$

$p(x|y = 0)$ is uniform over $[5, 7]$ and zero outside the interval.
$p(x|y = 1)$ is uniform over $[1, 20]$ and zero outside the interval.

(a) Compute $p(y = 1|x = 6)$ and $p(y = 0|x = 6)$.

(b) What is the Bayes optimal classifier for this example? How does it partition the input space into decision regions?

Problem 3

Backround Reading: Section 2.1 in Bishop.

Let $x = (x_1, \ldots, x_M)$ be a vector of M binary random variables, $x_i \in \{0, 1\}$. Suppose the M random variables are independent, with x_i distributed according to a Bernoulli distribution with mean u_i. This leads to a joint distribution $p(x|u)$ parameterized by a vector of parameters $u = (u_1, \ldots, u_n)$.

(a) Give an expression for $p(x|u)$.

(b) Let T be a dataset with k independent samples from $p(x|u)$. Derive the maximum likelihood estimate for u.

Problem 4

In this assignment you will implement a naive Bayes classifier to recognize pictures of hand-written 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.

After loading the data in Matlab the training examples are available as 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 class of the example*. In this case $p(x|y)$ is a product of Bernulli distributions like in Problem 3. Let $u_{y,i}$ denote the mean of the Bernoulli distribution associated with the i-th feature (pixel) in the conditional distribution $p(x|y)$.

You can assume $p(y) = 1/10$ for each class y.

(a) 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 displaying a 28x28 image where the brightness of each pixel reflects the mean of the Bernoulli distribution associated with that pixel. Include the visualization of the models in your writeup.

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

(c) What are the typical errors made by the classifier? Is the naive Bayes assumption reasonable for this data?