Class Exercises. Set 2

1. Load the height/weight data using rawdata = dlmread('heightWeightData.txt'). The first column is the class label (1=male, 2=female), the second column is height, the third weight.

a) Write a Matlab/Python script to fit a Gaussian model to each class using all the data for training. What's the training error?

b) Repeat a) imposing the same covariance matrix for both classes.

c) Repeat a) imposing the diagonal covariance matrices.

2. Assume that X is distributed according to the Gaussian density with mean $\mu_X = 0$ and variance $\sigma_X^2 = 1$; Y is distributed according to the Gaussian density with mean $\mu_Y = 1$ and variance $\sigma_Y^2 = 1$.

a) What is the probability density function f_X at X = 0?

b) What is the probability density function f_Y at Y = 0?

c) Assume the distribution P(Z=z) = 0.5P(X = z)+0.5P(Y = z) known as mixture (i.e., 1/2 of the time points are generated by the X process and 1/2 of the time by the Y process). If Z = 0, what is the probability that the X process generated this data point?

3. Derive a Bayes discriminant function for a negative loss function such as $\lambda_{ik} = \begin{cases} -h & if\;\; k = 1 \\ 0 & otherwise \end{cases}$

4. Fitting a naïve bayes spam filter by hand

Consider a Naïve Bayes model (multivariate Bernoulli version) for spam classification with the vocabulary V = "secret", "offer", "low", "price", "valued", "customer", "today", "dollar", "million", "sports", "is", "for", "play", "healthy", "pizza".
We have the following example spam messages "million dollar offer", "secret offer today", "secret is secret" and normal messages, "low price for valued customer", "play secret sports today", "sports is healthy", "low price pizza". Give the MLEs for the following parameters: θ_{spam}, $\theta_{secret|spam}$, $\theta_{secret|non\text{-}spam}$, $\theta_{sports|non\text{-}spam}$, $\theta_{dollar|spam}$.

5. Let the features $X=(x_1, x_2, \ldots, x_d)$ be binary valued (1 or 0). Let p_{ij} denote the probability that the feature x_i takes on the value 1 given class j. Assume that there are only two classes and that they are equally probable. Let the features be conditionally independent for both classes. Finally, assume that d is odd and the $p_{i1}=p>1/2$ and $p_{i2}=1-p$, for all i. Show that the optimal Bayes decision rule becomes: decide class one if $x_1+x_2+\ldots+x_d>d/2$, and class two otherwise.