Homework 3

CS 591Q/791V - Pattern Recognition

Instructor: Dr. Arun Ross

Due Date: March 25, 2010

Note: You are permitted to discuss the following questions with others in the class. However, you *must* write up your *own* solutions to these questions. Any indication to the contrary will be considered an act of academic dishonesty. Code developed as part of this assignment must be placed in a zip file and sent to arun.ross at mail.wvu.edu with the subject line "CS 591Q/791V : Homework 3". Also, include a hard-copy of the code when you submit the homework.

---

1. [15 points] Consider a set of n i.i.d. samples (one-dimensional training patterns), $D = \{x_1, x_2, \ldots x_n\}$, that are drawn from the following distribution:

$$p(x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$$

   (a) Derive the maximum likelihood estimate of λ, i.e., $\widehat{\lambda}_{mle}$.

   (b) Assume that $D = \{12, 17, 20, 25, 30\}$. Plot the distribution after computing $\widehat{\lambda}_{mle}$ using D.

2. [15 points] Consider a set of n i.i.d. samples (one-dimensional training patterns), $D = \{x_1, x_2, \ldots x_n\}$, that are drawn from the following distribution:

$$p(x|\theta) = \begin{cases} 2\theta x e^{-\theta x^2}, & x \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

   (a) Derive the maximum likelihood estimate of θ, i.e., $\widehat{\theta}_{mle}$.

   (b) Assume that $D = \{12, 17, 20, 25, 30\}$. Plot the distribution after computing $\widehat{\theta}_{mle}$ using D.

3. [10 points] Let $\boldsymbol{x}$ be a d-dimensional binary (0 or 1) vector with a multivariate Bernoulli distribution

$$P(\boldsymbol{x}|\boldsymbol{\theta}) = \prod_{i=1}^{d} \theta_i^{x_i} (1 - \theta_i)^{1-x_i},$$

where $\boldsymbol{\theta} = (\theta_1, \ldots \theta_d)^t$ is an unknown parameter vector, θ_i being the probability that $x_i = 1$. Let $D = \{\boldsymbol{x_1}, \ldots \boldsymbol{x_n}\}$ be a set of n i.i.d. training samples. Show that the maximum likelihood estimate for $\boldsymbol{\theta}$ is

$$\widehat{\boldsymbol{\theta}} = \frac{1}{n} \sum_{k=1}^{n} \boldsymbol{x_k}.$$

4. Consider a two-category (ω_1 and ω_2) classification problem with equal priors. Each feature is a two-dimensional vector $\boldsymbol{x} = (x_1, x_2)^t$. The class-conditional densities are:
   $p(\boldsymbol{x}|\omega_1) \sim N(\boldsymbol{\mu_1} = [0, 0]^t, \Sigma_1 = 2I)$,
   $p(\boldsymbol{x}|\omega_2) \sim N(\boldsymbol{\mu_2} = [1, 2]^t, \Sigma_2 = I)$.

   Generate 50 bivariate *random* training samples from each of the two densities.

   (a) [10 points] Find the values for the maximum likelihood estimates of $\boldsymbol{\mu_1}$, $\boldsymbol{\mu_2}$, Σ_1, and Σ_2 using these training samples (see page 89, equations (18) and (19)).

   (b) [10 points] Compute the Bayes decision boundary using the *estimated* parameters and plot it along with the training points.

   (c) [10 points] Compute the Bayes decision boundary using the *true* parameters and plot it on the same graph.

   (d) [10 points] Repeat (a) - (c) after generating 1,000 and 10,000 random training samples from each of the two densities. Comment on your result.

5. The IMOX dataset consists of 192 8-dimensional patterns pertaining to four classes (digital characters 'I', 'M', 'O' and 'X'). There are 48 patterns per class. The 8 features correspond to the distance of a character to the (a) upper left boundary, (b) lower right boundary, (c) upper right boundary, (d) lower left boundary, (e) middle left boundary, (f) middle right boundary, (g) middle upper boundary, and (h) middle lower boundary. Note that the class labels (1,2,3, or 4) are indicated at the end of every pattern.

   Assume that each class can be modeled by a multivariate Gaussian density with unknown mean and covariance, i.e., $p(\boldsymbol{x}|\omega_i) \sim \mathrm{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, $i = 1, 2, 3, 4$. Design a Bayes classifier and test it as follows:

   (a) [10 points] Train the classifier: Using the first 24 patterns of each class (training data), compute the maximum likelihood estimate of the model parameters. Report these estimates, i.e., $\hat{\boldsymbol{\mu}}_i$ and $\widehat{\boldsymbol{\Sigma}}_i$, $i = 1, 2, 3, 4$.

   (b) [10 points] Design the classifier: Assuming that the four classes are equally probable, write a program that inputs an 8-dimensional pattern $\boldsymbol{x}$ and assigns it to one of the four classes based on the maximum posterior rule, i.e., assign $\boldsymbol{x}$ to ω_j if,

   $$j = \arg \max_{i=1,2,3,4} \{P(\omega_i|\boldsymbol{x})\}.$$

   (c) [10 points] Test the classifier: Classify the remaining 24 patterns of each class (test data) using the Bayes classifier constructed above and report the confusion matrix for this four-class problem.

CS 791V students will have to solve the following problem in addition to the ones above

1. [10 points] Let x have a uniform density

$$p(x|\theta) \sim U(0,\theta) = \begin{cases} 1/\theta, & 0 \leq x \leq \theta \\ 0, & \text{otherwise.} \end{cases}$$

   (a) Suppose that n samples $D = \{x_1, \ldots x_n\}$ are drawn independently according to $p(x|\theta)$. Show that the MLE for θ is $\max[D]$ - that is, the value of the maximum element in D.

   (b) Suppose that $n = 5$ points are drawn from the distribution and the maximum value of which happens to be 0.6. Plot the likelihood $p(D|\theta)$ in the range $0 \leq \theta \leq 1$. Explain in words why you do not need to know the values of the other 4 points.