American University of Beirut

CMPS 396

Advanced Algorithms and Data Structures

Fall 2003-2004

Take Home Final Exam

Instructor: Chiraz Benabdelkader

Date Assigned: Jan. 23th, 2004.
Date Due: Jan. 28th, 2004 at 5PM sharp.

ID #: ------------------------------

Your exam should have 9 pages, and there are 7 questions totaling 100 points. Your answers should be concise and clearly written. Feel free to use extra white pages to write your solutions.

In case some of you forgot, your exam should be based on your absolute own individual effort. This is a take-home exam, yet you should treat it as an in-class exam in that you are not to discuss any of the questions with **anybody**, be it your colleagues or any other human being (discussions with your favourite pet is allowed however). Also, you are not to copy answers from the Internet or a book. The slightest evidence of cheating will be treated according to the University's Dishonesty Policy. No exceptions.

Finally, please check your email periodically, in case I need to mail you any typo corrections for the exam. Also, be prepared to answer questions and clarifications about your solutions after you submit them.

	Prob. 1	Prob. 2	Prob. 3	Prob. 4	Prob. 5	Prob. 6	Prob. 7	Total
Max Grade	10	15	10	20	15	10	15	**100**
Your Grade								

Problem 1 (10 points)

You are a witness to a night-time hit-and-run accident involving a taxi in Beirut. All taxis in Beirut are either yellow or beige. In the court trial, you swear under oath that the taxi was yellow.

Now, the taxi-driver's smart lawyer presents results based on extensive scientific experiments, showing that under the dim lighting conditions typical of Beirut, witnesses observing yellow and beige cars generally report the *correct* color only 85% of the time, regardless of the actual color of the car. Fortunately, the judge has taken a course on Bayes decision theory before. She first determines the most likely color of the car (C=yellow or C=beige), then gives a verdict for the accused taxi-driver: "guilty" if C matches his taxi's color, and "innocent" otherwise.

a) Given that 8 out of 10 taxis in Beirut are beige, what is the judge's verdict, assuming she uses optimal Bayes rule? What is the probability that she made a correct decision in this case? Justify your answer.

b) Suppose now that the judge takes into account the following costs associated with each verdict:

	Judge's Verdict	
True state	Guilty	Innocent
Guilty	0	2
Innocent	10	0

(i) Give a plausible explanation for what these costs might represent? (be creative)

(ii) What is then the judge's verdict, assuming she wants to minimize risk, and what is the risk that she made the wrong decision?

SOLUTION

Problem 2 (15 points)

Every student will tell you that exam grades are variable and that their exam performance will vary depending on mood, level of preparation, and maybe even luck. Let's suppose that a student's "intrinsic" ability is represented by an unknown α that is measured as a percentage (i.e. anywhere between 0 and 100). However, the grade g (also measured as a percentage) that he gets in an exam varies as a Gaussian distribution with mean α and standard deviation $\sigma=10$:

$$p(g \mid \alpha) \sim N(\alpha, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2} \frac{(g-\alpha)^2}{\sigma^2} \right]$$

(a) What is the probability that a student of intrinsic ability α will get a grade below 60% on an exam?

(b) Normally, we try to get a better idea of a student's ability by having them take several exams. Suppose a student writes N exams, getting grades $g_1, g_2, \ldots, g_N$, and that his grades are statistically independent. What is the maximum-likelihood estimate of the student's intrinsic ability, $\hat{\alpha}_{ML}(g_1, g_2, \ldots, g_N)$?

(c) What is the probability that $\hat{\alpha}_{ML}(g_1, g_2, \ldots, g_N)$ is below 60%? Express your answer as a function of the true α.

(d) It is interesting to assess how more and more exams (i.e. larger N) affect the precision of the estimated value, $\hat{\alpha}_{ML}$. Determine:

$$\lim_{N \to \infty} \Pr(\hat{\alpha}_{ML}(g_1, \ldots, g_N) < 60)$$

for the following three cases: when the true intrinsic ability is $\alpha=59$, $\alpha=61$, and $\alpha=70$. Show all work!

SOLUTION

Problem 3 (10 points)

This problem is an extension of Problem 2. It turns out that maybe a Gaussian distribution is not such a good model after all for how an exam grade g is distributed; It seems more likely that one would have a bad day, or accidentally mess up a question, than to accidentally get a whole lot of questions right. Maybe it is more reasonable then that the actual grade can never be higher than α. Based on these observations, we decide to instead model g as:

$$p(g \mid \alpha) = \begin{cases} 0 & g > \alpha \\ \frac{1}{\sigma}\exp\left[\frac{g-\alpha}{\sigma}\right] & g \leq \alpha \end{cases}$$

(a) Sketch p(g|α) for α=70 and σ=10.

(b) What is the probability that a student of ability α will get a grade below 60%?

(c) Suppose a student writes N tests, getting grades $g_1, g_2, \ldots, g_N$, and that these grades are statistically independent. What is $\hat{\alpha}_{ML}(g_1, g_2, \ldots, g_N)$ for this problem?

SOLUTION

Problem 4 (20 points)

Suppose we have a one-dimensional non-parametric estimation problem. We are given n samples $x_1, x_2, \ldots, x_n$ of a random variable x, and we want to estimate its probability density function, $p(x)$, from these samples.

(a) Write down the general form for $p(x)$ estimated using the Parzen windows method. Show that $\int \hat{p}_{parz}(x)dx = 1$. Note: you do not need to make any assumptions about the type of window shape or its width.

Suppose now that we have two *equally likely* classes and a total of 4 data points:

Class A: $x_1 = -3$, $x_2 = -1$

Class B: $x_3 = +1$, $x_4 = +2$

We want to estimate $p(x|A)$ and $p(x|B)$ from these 4 data points using the Parzen method.

(b) Draw a **neat** sketch of $\hat{p}(x \mid A)$ and $\hat{p}(x \mid B)$ based on the Parzen method: (i) using a Gaussian window with standard deviation 1, (ii) using a rectangular window with half-width 0.5, and (iii) using a triangular window with half-width 0.5. Compare your results. Which window do you recommend using for this problem.

(c) Now suppose we use the $\hat{p}(x \mid A)$ and $\hat{p}(x \mid B)$ from part **(c)** estimated with the Gaussian window to develop a Bayes classifier. Assuming equal priors, how many classification boundaries will there be? Why?

(d) Suppose I propose the following decision rule for the classifier:

Choose class A if $x < 0$

Choose class B if $x \geq 0$

Based on $\hat{p}(x \mid A)$ and $\hat{p}(x \mid B)$ as in part **(d)**, what is the probability of classification error for this classifier?

Problem 5:

For each of the following two-class cases, give an equation for the boundary and <u>sketch</u> it along with the <u>unit</u> standard deviation contours. Make your plots approximately to scale. For all the cases, assume $\mu_1=[0\ 0]^T$ and $[\mu_2= [0\ 4]^T$. You may also assume equal priors.

$$\text{Case 1: } \Sigma_1 = \begin{bmatrix} 25 & 0 \\ 0 & 5 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\text{Case 2: } \Sigma_1 = \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\text{Case 3: } \Sigma_1 = \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix}$$

Note: The purpose of this problem is to test your analytical skills, and not to test your ability to use Matlab, Mathematica or any other mathematical software. But since there is no way for me to prevent you from doing this (which I will by the way consider as a form of **cheating**), you need to show ALL your work (including any intermediate steps), both for the boundary equations and the contours.

SOLUTION

Problem 6:

Consider the neural network in Figure 6.1 of the textbook. In fact this is not the only solution to the XOR problem. Design a different neural network using the <u>same architecture</u> to solve the XOR problem. In other words, you need to provide a different set of weight values for the network of Figure 6.1 that still solves the problem. Show ALL your work.

Hint: Do the *reverse* analysis of what you did in HW3 for this problem.

SOLUTION

Problem 7:

Suppose we have the following two-dimensional training data from three categories:

Class A: (3,1), (3,-2), (5,0)

Class B: (1,1), (2,2), (2,0)

Class C: (0,0), (1,0), (0,1)

(a) Plot these training points, and construct by inspection the weight vector for the optimal separating hyperplane for every two categories. Sketch your answers on the same plot.

(b) We consider separating categories B and C. Assuming the initial estimate of the decision boundary is: x-2y-1=0, use the **batch-relaxation** learning rule with b=0 and η=1 to determine the first three estimates of the decision boundary. Sketch these three boundaries in (x,y)-space and clearly label them. Again, show all work.

(c) We now consider the problem of separating all three categories:

(i) Derive Kesler's construction for solving this three-category problem. Your answer should consist of all the η_{ij}^{k} vectors described in the textbook.

(ii) Explain how you would use the MSE procedure to solve this problem, and derive all the relevant Y and B matrices (as described in the textbook).

SOLUTION