CSE 555 Spring 2009 Mid-Term Exam

Jason J. Corso
Computer Science and Engineering
University at Buffalo SUNY
jcorso@cse.buffalo.edu
Date 5 Mar 2009

The exam is worth 100 points total and each question is marked with its portion. The exam is closed book/notes. You have 70 minutes to complete the exam. Use the provided white paper, write your name on the top of each sheet and number them. Write legibly.

Problem 1: "Recall" Questions (25pts)

Answer each in one or two sentences.

1. (5pts) What is the fundamental difference between Maximum Likelihood parameter estimation and Bayesian parameter estimation?

2. (5pts) What quantity is PCA maximizing during dimension reduction?

3. (5pts) Describe Receiver Operating Characteristics or ROC-Curves (illustrate if necessary).

4. (5pts) What is the Curse of Dimensionality?

5. (5pts) What is the key idea of the No Free Lunch Theorem?

Solution:
Answers are all directly in the notes.

Problem 2: Bayesian Reasoning (25pts)

Monty Hall Formulate and solve this classifical problem using the Bayes rule. Imagine you are on a gameshow and you're given the choice of three doors: behind one door is a car and behind the other two doors are goats. You have the opportunity to select a door (say No. 1). Then the host, who knows exactly what is behind each door and will not reveal the car, will open a different door (i.e., one that has a goat). The host then asks you if you want to switch your selection to the last remaining door.

1. (3pts) Formulate the problem using the Bayes rule, i.e., what are the random variables and the input data. What are the meaning of the prior and the posterior probabilities in this problem (one sentence each).

   Solution:
   Without the loss of generality, suppose that we chose door 1 at the beginning.
   Random Variables: C_i represents the state that the car is behind door $i, i \in [1, 2, 3]$, and H_j represents the state that the host opens door $j, j \in [2, 3]$.
   Input Data: The door that the host opens.
   Prior Probability: The probability of winning the car before the host opens the door.
   Posterior Probability: The probability of winning the car after the host opens the door.

2. (3pts) What are the probability values for the prior?

   Solution:
   $P(C_1) = \frac{1}{3}, P(C_2) = \frac{1}{3}, P(C_3) = \frac{1}{3}$.

3. (3pts) What are the probability values for the likelihood?

   Solution:
   $P(H_2|C_1) = \frac{1}{2}, P(H_3|C_1) = \frac{1}{2},$
   $P(H_2|C_2) = 0, P(H_3|C_2) = 1,$
   $P(H_2|C_3) = 1, P(H_3|C_3) = 0.$

4. (3pts) Derive the posterior probability (include intermediate steps).

   Solution:
   (a) If we chose to stay after the host opened door 2, the probability of winning the car is:

   $$P(C_1|H_2) = \frac{P(H_2|C_1)P(C_1)}{P(H_2)} = \frac{P(H_2|C_1)P(C_1)}{P(H_2|C_1)P(C_1) + P(H_2|C_2)P(C_2) + P(H_2|C_3)P(C_3)}$$
   $$= \frac{1/2 \cdot 1/3}{1/2 \cdot 1/3 + 0 \cdot 1/3 + 1 \cdot 1/3} = \frac{1/6}{1/2} = \frac{1}{3}.$$

   (b) If we chose to stay after the host opened door 3, the probability of winning the car is:

   $$P(C_1|H_3) = \frac{P(H_3|C_1)P(C_1)}{P(H_3)} = \frac{P(H_3|C_1)P(C_1)}{P(H_3|C_1)P(C_1) + P(H_3|C_2)P(C_2) + P(H_3|C_3)P(C_3)}$$
   $$= \frac{1/2 \cdot 1/3}{1/2 \cdot 1/3 + 1 \cdot 1/3 + 0 \cdot 1/3} = \frac{1/6}{1/2} = \frac{1}{3}.$$

   $\Rightarrow$ Our posterior probability of winning the car if we chose not the switch is always $\frac{1}{3}$, no matter what door the host opened.

   (c) If we chose to switch doors (from 1 to 3) after the host opened door 2, the probability of winning is:

   $$P(C_3|H_2) = \frac{P(H_2|C_3)P(C_3)}{P(H_2)} = \frac{P(H_2|C_3)P(C_3)}{P(H_2|C_1)P(C_1) + P(H_2|C_2)P(C_2) + P(H_2|C_3)P(C_3)}$$
   $$= \frac{1 \cdot 1/3}{1/2 \cdot 1/3 + 0 \cdot 1/3 + 1 \cdot 1/3} = \frac{1/3}{1/2} = \frac{2}{3}.$$

   (d) If we chose to switch doors (from 1 to 2) after the host opened door 3, the probability of winning is:

   $$P(C_2|H_3) = \frac{P(H_3|C_2)P(C_2)}{P(H_3)} = \frac{P(H_3|C_2)P(C_2)}{P(H_3|C_1)P(C_1) + P(H_3|C_2)P(C_2) + P(H_3|C_3)P(C_3)}$$
   $$= \frac{1 \cdot 1/3}{1/2 \cdot 1/3 + 1 \cdot 1/3 + 0 \cdot 1/3} = \frac{1/3}{1/2} = \frac{2}{3}.$$

   $\Rightarrow$ Our posterior probability of winning the car if we chose not the switch is always $\frac{2}{3}$, no matter what door the host opened.

5. (3pts) Is it in the contestant's advantage to switch his/her selection? Why?

   Solution:
   Yes, due to the posterior probabilities we calculated above.

6. (10pts) Now, consider the following twist. The host is having trouble remembering what is behind each of the doors. So, we cannot guarantee that he will not accidentally open the door for the car. We only know that he will not open the door the contestant has opened. Indeed, if he accidentally opens the door for the car, the contestant

wins. How does this change the situation? Is it now in the contestant's advantage to switch? Rederive your probabilities to justify your answer.

Solution:

(a) Again, let's assume that we chose door 1 without the loss of generality. The probability values for the likelihoods will change to:
$P(H_2|C_1) = \frac{1}{2}, P(H_3|C_1) = \frac{1}{2},$
$P(H_2|C_2) = \frac{1}{2}, P(H_3|C_2) = \frac{1}{2},$
$P(H_2|C_3) = \frac{1}{2}, P(H_3|C_3) = \frac{1}{2},$
thus causing the posterior probabilities to change.

(b) It doesn't matter, because the posterior probabilities of switching or not switching is equivalent.

(c)   i. If we chose to stay after the host opened door 2, the probability of winning the car is:

$$P(\text{winning}|H_2) = P(C_1|H_2) + P(C_2|H_2) = \frac{P(H_2|C_1)P(C_1)}{P(H_2)} + \frac{P(H_2|C_2)P(C_2)}{P(H_2)}$$
$$= \frac{1/2 \cdot 1/3}{1/2} + \frac{1/2 \cdot 1/3}{1/2} = \frac{2}{3}.$$

ii. If we chose to stay after the host opened door 3, the probability of winning the car is:

$$P(\text{winning}|H_3) = P(C_1|H_3) + P(C_3|H_3) = \frac{P(H_3|C_1)P(C_1)}{P(H_3)} + \frac{P(H_3|C_3)P(C_3)}{P(H_3)}$$
$$= \frac{1/2 \cdot 1/3}{1/2} + \frac{1/2 \cdot 1/3}{1/2} = \frac{2}{3}.$$

$\Rightarrow$ Our posterior probability of winning the car if we chose not the switch is always $\frac{2}{3}$, no matter what door the host opened.

iii. If we chose to switch doors (from 1 to 3) after the host opened door 2, the probability of winning is:

$$P(\text{winning}|H_2) = P(C_3|H_2) + P(C_2|H_2) = \frac{P(H_2|C_3)P(C_3)}{P(H_2)} + \frac{P(H_2|C_2)P(C_2)}{P(H_2)}$$
$$= \frac{1/2 \cdot 1/3}{1/2} + \frac{1/2 \cdot 1/3}{1/2} = \frac{2}{3}.$$

iv. If we chose to switch doors (from 1 to 2) after the host opened door 3, the probability of winning is:

$$P(\text{winning}|H_3) = P(C_2|H_3) + P(C_3|H_3) = \frac{P(H_3|C_2)P(C_2)}{P(H_3)} + \frac{P(H_3|C_3)P(C_3)}{P(H_2)}$$
$$= \frac{1/2 \cdot 1/3}{1/2} + \frac{1/2 \cdot 1/3}{1/2} = \frac{2}{3}.$$

$\Rightarrow$ Our posterior probability of winning the car if we chose not the switch is always $\frac{2}{3}$, no matter what door the host opened.

Problem 3: Discriminants (25pts)

Consider two univariate Gaussian distributions with means μ_1 and μ_2, respectively, and an arbitrary, but known and

equivalent variance σ^2. Assume you know the priors $P(\omega_1)$ and $P(\omega_2)$. The class-conditional densities are given by

$$P(x|\mu_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\frac{(x-\mu_i)^2}{\sigma^2}\right] \quad , i = 1, 2.$$

Consider only the the case of the zero-one loss function.

1. (3pts) Write down the Bayes discriminants, $g_1(x)$ and $g_2(x)$.

   Solution:

   $$g_i(x) = P(\omega_i|x) \quad \text{or} \quad = P(\mu_i|x) \quad , i = 1, 2$$

2. (9pts) The dichotomizer is a single discriminant function for two-class situations, $g(x) = g_1(x) - g_2(x)$. And, we decide ω_1 if $g(x) > 0$. Derive the log-likelihood ratio form of the dichotomizer (the rule for selecting class ω_1).

   Solution:

   $$P(\omega_1|x) > P(\omega_2|x)$$
   $$\log P(\omega_1|x) > \log P(\omega_2|x)$$
   $$\log P(x|\omega_1) - \log P(x|\omega_2) > \log P(\omega_2) - \log P(\omega_1)$$
   $$\log \frac{P(x|\omega_1)}{P(x|\omega_2)} > \log \frac{P(\omega_2)}{P(\omega_1)}$$

3. (3pts) What role do the priors play in this discriminant-based classification ($<$ two sentences)?

   Solution:

   They shift the decision boundary.

4. (10pts) Derive the Fisher linear discriminant (for the distributions in question) directly from the negative log-likelihood ratio.

   Solution:
   WRONG For two Gaussian distributions with equal covariance, the optimal decision boundary has the form:

   $$\mathbf{w}^t\mathbf{x} + w_0 = 0, \qquad \mathbf{w} = \frac{1}{\sigma^2}(\mu_1 - \mu_2),$$

   where $\mathbf{w}$ is the Fisher's linear discriminant (the linear function yielding the maximum ratio of $\mathbf{S}_B/\mathbf{S}_W$, and w_0 is the point where the posteriors of the two classes are equal, i.e.,
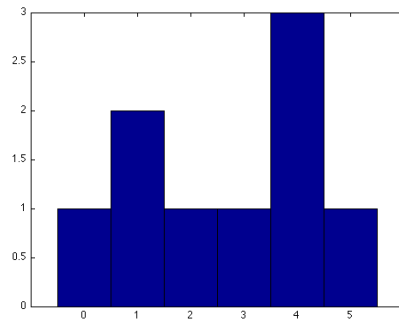
   $$\log \frac{P(x|w_1)}{P(x|w_2)} - \log \frac{P(w_2)}{P(w_1)} = 0$$

Problem 4: Non-Parametric Methods (25pts)

You are given a dataset $\mathcal{D} = \{0, 1, 1, 2, 3, 4, 4, 4, 5\}$. Using techniques from non-parametric density estimation, answer the following questions:

1. (2pts) Draw a histogram of $\mathcal{D}$ with a bin-width of 1 and bins centered at $\{0, 1, 2, 3, 4, 5\}$.

Solution:

2. (3pts) Write the formula for the kernel density estimate given an arbitrary kernel K.

   Solution:
   Straight from the notes (not defining the notation):

   $$p_n(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{V_n}\varphi\left(\frac{\mathbf{x}-\mathbf{x}_i}{h_n}\right) \ .$$

3. (5pts) Select a triangle kernel as your window function:

   $$K(u) = (1-|u|)\delta(|u| \le 1).$$

   where u is a function of the distance of sample x_i to the value in question x divided by the bandwidth. Compute the kernel density estimates for the following values of $x = \{2,4\}$ bandwidths of 2. Compare the kernel density estimate of these values to the histogram estimate.

   Solution:
   0.1389 and 0.2222 respectively.

4. (5pts) Given the triangle kernel and our integer dataset, specify a bound on the bandwidth such that the kernel density will equal the histogram density.

   Solution:
   $0 < w \le 1$

5. (10pts) Describe a method for estimating the bias and the variance of the kernel density estimate with the triangle kernel, if you think it is possible. If not, state why you it is not possible to compute estimates of the bias and variance. If they are possible, then compute them.

   Solution: