CSE 455/555 Spring 2011 Homework 1: Bayesian Decision Theory

Jason J. Corso
Computer Science and Engineering
SUNY at Buffalo
jcorso@buffalo.edu

Date Assigned 24 Jan 2011
Date Due 14 Feb 2011
Homework must be submitted in class. No late work will be accepted.

Remember, you are permitted to discuss this assignment with other students in the class (and not in the class), but you must write up your own work from scratch.
*I am sure the answers to some or all of these questions can be found on the internet. Copying from **any** another source is indeed cheating.*
This class has a zero tolerance policy toward cheaters and cheating. Don't do it.

Problem 1: Bayesian Decision Rule (30%)

Suppose the task is to classify the input signal x into one of K classes $\omega \in \{1, 2, \ldots, K\}$ such that the action $\alpha(x) = i$ means classifying x into class i. The Bayesian decision rule is to maximize the posterior probability

$$\alpha_{\text{Bayes}}(x) = \omega^* = \arg\max_{\omega} p(\omega|x) \ .$$

Suppose we replace it by a *randomized decision rule*, which classifies x to class i following the posterior probability $p(\omega = i|x)$, i.e.,

$$\alpha_{\text{rand}}(x) = \omega \sim p(\omega|x) \ .$$

Solution:

Maximizing the posterior probability is equivalent to minimizing the overall risk.
Using the zero-one loss function, the overall risk for the Bayes Decision Rule is:

$$\begin{aligned} R_{Bayes} &= \oint R(\alpha_{Bayes}(x)|x)p(x)dx \\ &= \oint \Big\{1 - \max\big[P(\omega_j|x) \mid j = 1, ..., k\big]\Big\} p(x)dx \end{aligned}$$

For simplicity, the class with max posterior probability is abbreviated as ω_{max}, and we get:

$$R_{Bayes} = \oint (1 - P(\omega_{max}|x))p(x)dx.$$

1. What is the overall risk R_{rand} for this decision rule? Derive it in terms of the posterior probability using the zero-one loss function.

Solution:

For any given x, the probability of each class $j = 1, ..., k$ being the correct class is $P(\omega_j|k)$. With the randomized algorithm, it will select the correct class with probability $P(\omega_j|k)$, which means that it will select the wrong class with probability $1 - P(\omega_j|k)$. Thus, the zero-one conditional risk will become $\sum_j P(\omega_j|x)\big[1 - P(\omega_j|x)\big]$ on average. Thus,

$$\begin{aligned} R_{rand} &= \oint \Big\{ \sum_j P(\omega_j|x)\big[1 - P(\omega_j|x)\big] \Big\} p(x)dx \\ &= \oint \Big\{ \sum_j \big[P(\omega_j|x) - P(\omega_j|x)^2\big] \Big\} p(x)dx \\ &= \oint \Big[1 - \sum_j P(\omega_j|x)^2\Big] p(x)dx \end{aligned}$$

2. Show that this risk R_{rand} is always no smaller than the Bayes risk R_{Bayes}. Thus, we cannot benefit from the randomized decision.

Solution:

Proving $R_{rand} \geq R_{Bayes}$ is equivalent to proving $\sum_j P(\omega_j|x)^2 \leq P(\omega_{max}|x)$:

$$\sum_j P(\omega_j|x)^2 \leq \sum_j P(\omega_j|x) P(\omega_{max}|x) = P(\omega_{max}|x),$$

thus proved. R_{rand} is always no smaller than R_{Bayes}.

3. Under what conditions on the posterior are the two decision rules the same?

Solution:

When the posterior probabilities of all classes are uniform distributions with equivalent value.

Problem 2: Bayesian Classification Boundaries for the Normal Distribution (30%) Suppose we have a two-class recognition problem with salmon ($\omega = 1$) and sea bass ($\omega = 2$).

1. First, assume we have one feature, the pdfs are the Gaussians $\mathcal{N}(0, \sigma^2)$ and $\mathcal{N}(1, \sigma^2)$ for the two classes, respectively. Show that the threshold τ minimizing the average risk is equal to

$$\tau = \frac{1}{2} - \sigma^2 \ln \frac{\lambda_{12} P(\omega_2)}{\lambda_{21} P(\omega 1)} \tag{1}$$

where we have assumed $\lambda_{11} = \lambda_{22} = 0$.

Solution:

Define the $R(\tau)$ is the average risk for the threshold τ:

$$R(\tau) = \int_0^\tau \lambda_{12} P(\omega_2) p(x|\omega = 2)\, dx + \int_\tau^{+\infty} \lambda_{21} P(\omega_1) p(x|\omega = 1)\, dx$$

Get derivative about τ for $R(\tau)$, then obtain the minimization when the derivative equals to 0, so make it equals to 0:

$$\lambda_{12} P(\omega_2) \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\tau^2}{2\sigma^2}} - \lambda_{21} P(\omega_1) \cdot \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(\tau-1)^2}{2\sigma^2}} = 0$$

Therefore,

$$\tau = \frac{1}{2} - \sigma^2 \ln \frac{\lambda_{12} P(\omega_2)}{\lambda_{21} P(\omega_1)}$$

2. Next, suppose we have two features $\mathbf{x} = (x_1, x_2)$ and the two class-conditional densities, $p(x|\omega = 1)$ and $p(x|\omega = 2)$, are 2D Gaussian distributions centered at points $(4, 11)$ and $(10, 3)$ respectively with the same covariance matrix $\mathbf{\Sigma} = 3\mathbf{I}$ (with $\mathbf{I}$ is the identity matrix). Suppose the priors are $P(\omega = 1) = 0.6$ and $P(\omega = 2) = 0.4$.

   (a) Suppose we use a Bayes decision rule, write the two discriminant functions $g_1(\mathbf{x})$ and $g_2(\mathbf{x})$. *Solution:*
   According to bayes decision rule:

$$g_1(x) = p(x|\omega = 1) \cdot P(\omega = 1) = \frac{1}{2\pi \cdot 3} e^{-\frac{(x_1-4)^2+(x_2-11)^2}{2*3}} \cdot 0.6$$

$$g_2(x) = p(x|\omega = 2) \cdot P(\omega = 2) = \frac{1}{2\pi \cdot 3} e^{-\frac{(x_1-10)^2+(x_2-3)^2}{2*3}} \cdot 0.4$$

   (b) Derive the equation for the decision boundary $g_1(\mathbf{x}) = g_2(\mathbf{x})$. *Solution:*
   Derive the equation for the decision boundary $g_1(x) = g_2(x)$, we get:

$$-6x_1 + 8x_2 - 14 + 3 \ln \frac{3}{2} = 0$$

   (c) How would the decision boundary change if we changed the priors? the covariances?
   Solution:

When all the covariance matrices are the same, the decision boundary will be a straight line in two dimensional case, and plane or hyper-plane in three or higher dimensional space. In particular if the covariance matrix are in a special diagonal form $\Sigma = \sigma^2 \cdot \mathrm{I}$, the direction of the decision surface will be perpendicular to the direction between two means. If the covariance matrices are different for each class, then the quadratic term in the function of decision surface will not be canceled, and thus it will not be a straight plane. Let's restrict us to a simpler case that all the covariance matrices are the same and analyze the influence of class priors to the position of the decision boundary. Now we can write the decision boundary using the following equation:

$$\mathbf{w}^T(\mathbf{x} - \mathbf{x_0}) = 0$$

where

$$\mathbf{w} = \Sigma^{-1}(\mu_i - \mu_j)$$

and

$$\mathbf{x_0} = \frac{1}{2}(\mu_i + \mu_j) - \frac{ln[P(\omega_i)/P(\omega_j)]}{(\mu_i - \mu_j)^T\Sigma^{-1}(\mu_i - \mu_j)}(\mu_i - \mu_j)$$

In general, while the ratio of the distance between the means of the classes and the covariance are relatively close to one, as the prior of a class increases, the decision boundary will move toward the other class. But if the variance is relatively small as compared to the distance between the two means, i.e., the denominator is relatively large in above equation, the influence of class priors will be relatively small to the position of the decision boundary, e.g., in case of two well separated Gaussians and very peaked at the corresponding mean. On the other hand, if the variance is relatively large as compared to the distance between the two means, the position of the decision boundary is mainly determined by the class priors (e.g., it should be intuitive while two classes are heavily overlapped, the decision are mainly based on the prior knowledge we have).

(d) Using computer software, sample 100 points from each of the two densities. Draw them and draw the boundary on the feature space (the 2D plane). *Solution:*

Problem 4: Bayesian Reasoning (40%)

Formulate and solve this classical problem using the Bayes rule. There are three criminals A, B, and C waiting in three separate jail cells. One of them will be executed in the next morning when the sun rises. A is very nervous, as he has 1/3 chance to be the one. He tries to get some information from the janitor: "I know you cannot tell me whether I will be executed in the next morning, but can you tell me which of my inmates B and C will not be executed? Because one of them will not be executed anyway, by pointing out who will not be executed, you are not telling me any information." This sounds quite logical. So the Janitor tells A that C won't be executed. At a second thought, A gets much more worried. Before he asked the janitor, he thought he had 1/3 chance, but with C excluded, he seems to have 1/2 chance. A says to himself: "What did I do wrong? Why did I ask the janitor?"

1. 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 posterior probabilities in this problem?

Solution:

Since who's going to be executed tomorrow has already been decided before A asked the janitor - otherwise the janitor won't be able to tell A which one of A's inmates will live - we have:

The random variable is all the possible answers from the janitor; the input data (observation) is the janitor's answer; the prior probability is the chance of A being executed before observing the janitor's answer; the posterior probability is the chance of A being executed after observing the janitor's answer.

2. What are the probability values for the prior?

   Solution:

   Let E_X, where X = $\{$A, B, C$\}$, denote the event that X is going to be executed. The prior probability of A being executed tomorrow is $P(E_A) = P(E_B) = P(E_C) = \frac{1}{3}$ (suppose they kill at random).

3. What are the probability values for the likelihood?

   Solution:

   Let L_Y, where Y = $\{$B, C$\}$, denote the event that: knowing that X will be executed tomorrow, the likelihood of the janitor telling A that Y will live. The likelihoods are: $P(L_B|E_A) = P(L_C|E_A) = \frac{1}{2}$; $P(L_B|E_B) = 0, P(L_C|E_B) = 1$; $P(L_B|E_C) = 1, P(L_C|E_C) = 0$.

4. Calculate the posterior probability (you need to derive the probability values with intermediate steps, not simply showing the final values).

   Solution:

   The posterior probability of A being executed is

$$
\begin{aligned}
P(E_A|L_C) &= \frac{P(L_C|E_A) \cdot P(E_A)}{P(L_C)} \\
&= \frac{P(L_C|E_A) \cdot P(E_A)}{P(L_C|E_A) \cdot P(E_A) + P(L_C|E_B) \cdot P(E_B) + P(L_C|E_C) \cdot P(E_C)} \\
&= \frac{1/2 \cdot 1/3}{1/2 \cdot 1/3 + 1 \cdot 1/3 + 0 \cdot 1/3} \\
&= \frac{1}{3}.
\end{aligned}
$$

5. What is the probability of A being executed after he knows that C is excluded?

   Solution:

   As shown in the posterior probability, it's still $\frac{1}{3}$.

6. Did the janitor tell us any information about A's fate?

   Solution:

   NO

7. Explain how the Bayes rule helps you. *Solution:*

   Helps decompose complicated problems into tractable parts, especially to determine the probability of an event A after observing the happening of event B.