CSE 555 Pattern Recognition Homework 1
Reference Solution

Albert Y. C. Chen

Computer Science and Engineering
University at Buffalo, The State University of New York

Assigned: 1/18/2009
Due: 02/03/2009

Problem 1 (25%)

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:

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

Let's abbreviate the above equation to just

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

1. 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 $\sum_j \big(1 - P(\omega_j|x)\big)P(\omega_j|x)$ on average. Thus,

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

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

Problem 2 (25%)

Note that discriminant functions are not unique. The following is the most straightforward one:

1. The general multivariate normal density in d dimensions is written as:

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}|\mathbf{\Sigma}|^{1/2}}\exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^t\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right].$$

By applying the the typical discriminant function (eq. 28, to remove the "exp" in the gaussian), the general discriminant function for multivariate normal distribution is:

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^t\mathbf{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i) - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|\mathbf{\Sigma}_i| + \ln P(\omega_i)$$

In the given situation, where $\mathbf{\Sigma} = 4\mathbf{I}$ (*i.e.* $\sigma = 2$), we have $|\mathbf{\Sigma}_i| = \sigma^{2d}$ and $\mathbf{\Sigma}_i^{-1} = (1/\sigma^2)\mathbf{I}$. Because both $|\mathbf{\Sigma_i}|$ and the $(d/2)$ term in the above equation are independent of i, they can be ignored, and we get the discriminant function of:

$$g_i(\mathbf{x}) = -\frac{||\mathbf{x}-\boldsymbol{\mu}_i||^2}{2\sigma^2} + \ln P(\omega_i).$$

Thus, the discriminant function for the two classes are:

$$g_1(\mathbf{x}) = -\frac{||\mathbf{x}-[4,16]^t||^2}{2(2)^2} + \ln(0.6) = -\frac{||\mathbf{x}-[4,16]^t||^2}{8} - 0.5108$$
$$g_2(\mathbf{x}) = -\frac{||\mathbf{x}-[16,4]^t||^2}{2(2)^2} + \ln(0.4) = -\frac{||\mathbf{x}-[16,4]^t||^2}{8} - 0.9163$$

2. The decision boundary $g_1(\mathbf{x}) = g_2(\mathbf{x})$ is equivalent to $g_1(\mathbf{x}) - g_2(\mathbf{x}) = 0$. Let $\mathbf{x} = [x_1, x_2]$ in this two dimension case:

$$\left(-\frac{||\mathbf{x}-[4,16]^t||^2}{8} - 0.5108\right) - \left(-\frac{||\mathbf{x}-[16,4]^t||^2}{8} - 0.9163\right) = 0$$
$$\Rightarrow \frac{-[(x_1-4)^2+(x_2-16)^2]+(x_1-16)^2+(x_2-4)^2}{8} + 0.4055 = 0$$
$$\Rightarrow \frac{(-x_1^2+8x_1-16+x_1^2-32x_1+196)+(-x_2^2+32x_2-196+x_2^2-8x_2+16)}{8} + 0.4055 = 0$$
$$\Rightarrow \frac{-24x_1+24x_2}{8} + 0.4055 = 0$$
$$\Rightarrow x_1 - x_2 = 0.1351$$

we have $x_1 - x_2 = 0.1351$ as our decision boundary.

Drawing is omitted. The decision boundary is the line that goes through $(0.1351, 0)$ and $(0, -0.1351)$, which is biased towards class two since class one has a slightly higher prior.

Problem 3 (10%)

1. Can be proved by convoluting the two distributions. Please refer to:
   `http://en.wikipedia.org/wiki/Sum_of_normally_distributed_random_variables`
2. $\mu_3 = \mu_1 + \mu_2$.
3. $\sigma_3^2 = \sigma_1^2 + \sigma_2^2$

Problem 4 (40%)

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:

1. The random variable x is all the possible answers from the janitor; the input data is the janitor's answer; the prior probability is the chance of A being executed before asking the janitor; the posterior probability is the chance of A being executed after asking the janitor.
2. 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. Let L_Y, where Y = {B, C}, denote the event that the janitor tells A that Y will live. The likelihoods are: $P(L_B|E_A) = P(L_C|E_A) = \frac{1}{2}$.
4. 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. As shown in the posterior probability, it's still $\frac{1}{3}$.
6. Nope!
7. Helps decompose complicated problems into tractable parts, especially to determine the probability of an event A after observing the happening of event B.