UNIVERSITY OF CALIFORNIA, LOS ANGELES
DEPARTMENT OF COMPUTER SCIENCE

CS 276A assignment 1

Yang Pei
304434922

October 16, 2014

1 PROBLEM 1

1. Since we use a Bayes decision for $\alpha(x)$, according to the textbook, we can use $g_i(x) = -R(\alpha_i|x)$ as our discriminant function, which is given by:

$$R(\alpha_i|x) = \sum_{j=1}^{c} \lambda(\alpha_i|\omega_j)p(\omega_j|x) \tag{1.1}$$

The risk matrix is given and all we need to do is to calculate the posterior probability which is given by:

$$p(\omega_j|x) = p(x|\omega_j) * p(\omega_j) \tag{1.2}$$

Here we have omit the component $p(x)$ since it is only a scale factor. The $p(x|\omega_j)$ is given and all are mulit norm distribution with the covariance matrixs all the same, so we can omit the term $\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}$ and this would not effect our discriminant functions. Combine the information above, the final discriminant for each category are:

$$\begin{aligned} g_1 &= -1.2 * e^{-1/18*[(x_1-12)^2+(x_2-3)^2]} - 0.4 * e^{-1/18*[(x_1-3)^2+(x_2-5)^2]} \\ g_2 &= -0.8 * e^{-1/18*[(x_1-4)^2+(x_2-12)^2]} - 0.2 * e^{-1/18*[(x_1-3)^2+(x_2-5)^2]} \\ g_3 &= -1.2 * e^{-1/18*[(x_1-4)^2+(x_2-12)^2]} - 0.4 * e^{-1/18*[(x_1-12)^2+(x_2-3)^2]} \end{aligned} \tag{1.3}$$

2. The decision boundaries for each two categories are shown in figure 1.1.

3. Given the new prior distribution, the new decision boundaries for each two categories are shown in figure 1.2. We could see that the shape does not have too much change, but it is shfit from left to right due to a higher prior distribution for category 3.

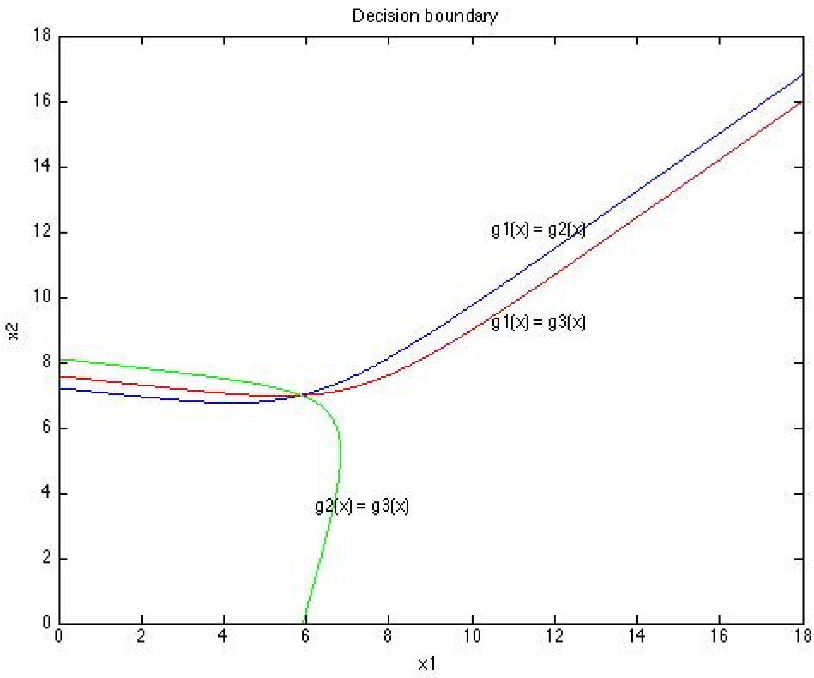

Figure 1.1: Decision boundaries

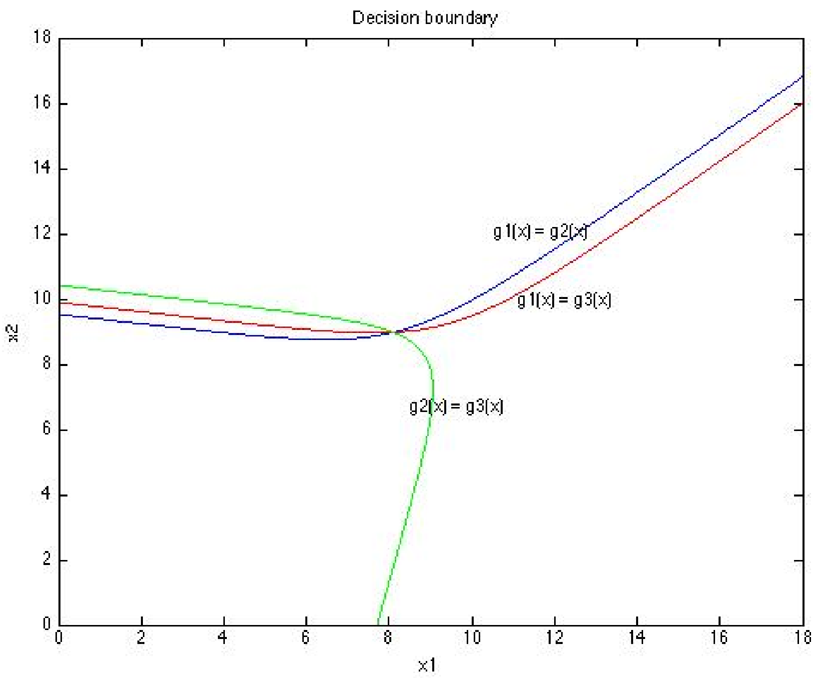

Figure 1.2: New decision boundaries

2 PROBLEM 2

1. The average risk is computed as

$$R = \int R(\alpha(x)|x)p(x)dx \tag{2.1}$$

where $\alpha(x)$ is the decision function which maps an input to an action. In Bayes decision, we always map to the action, or decision, with the largest posterior probability

$$\alpha_{Bayes}(x) = y^* = arg \max_{y \in \Omega^c} p(y|x)$$

and this is determined. In terms of randomized decision rule, the action is not determined, which is given by

$$\alpha_{ran}(x) = y \sim p(y|x)$$

In order to compute the average risk R_{ran}, we need to compute $R_{ran}(\alpha(x)|x)$ first, which is given by:

$$\begin{aligned} R_{ran}(\alpha(x)|x) &= \int R_{ran}(\alpha(x)|x,\beta)d\beta = \int \sum_{i=1}^{k} R(\alpha_i|x,\beta)d\beta \\ &= \int \sum_{i=1}^{k} R(\alpha_i|x) * p(\alpha_i|x,\beta)d\beta = \sum_{i=1}^{k} (R(\alpha_i|x) * \int p(\alpha_i|\beta)d\beta) \end{aligned} \tag{2.2}$$

The term $\int p(\alpha_i|\beta)d\beta$ in equation 2.2 stands for the chance that the random number generator β place the action to α_i and this is just given by the probability $p(y_i|x)$(since action is classification or decision in this problem). Repalce it in the equation 2.2 we get:

$$R_{ran}(\alpha(x)|x) = \sum_{i=1}^{k} R(\alpha_i|x) * p(y_i|x) \tag{2.3}$$

since we use 0-1 loss function, then we have $R(\alpha_i|x) = 1 - p(y_i|x)$, repalce it in the equation 2.3 we get the final representation:

$$R_{ran}(\alpha(x)|x) = \sum_{i=1}^{k}(1 - p(y_i|x)) * p(y_i|x) = \sum_{i=1}^{k} p(y_i|x) - \sum_{i=1}^{k} p(y_i|x)^2 = 1 - \sum_{i=1}^{k} p(y_i|x)^2 \tag{2.4}$$

Then, average risk for the randomized decision rule R_{ran} is

$$R_{ran} = \int (1 - \sum_{i=1}^{k} p(y_i|x)^2)p(x)dx \tag{2.5}$$

2. To prove R_{ran} is larger than or equal to Bayes risk R_{Bayes}, we prove for each x, $R_{ran}(\alpha(x)|x)$ is larger than $R_{Bayes}(\alpha(x)|x) = 1 - p(y_{max}|x)$, here y_{max} stands for y^* where max indi-

cates the class number to ease the discussion.

$$\begin{aligned} R_{ran}(\alpha(x)|x) - R_{Bayes}(\alpha(x)|x) &= p(y_{max}|x) - \sum_{i=1}^{k} p(y_i|x)^2 \\ &= p(y_{max}|x)(1 - p(y_{max}|x)) - \sum_{i \neq max} p(y_i|x)^2 \\ &= p(y_{max}|x) * \sum_{i \neq max} (p(y_i|x)) - \sum_{i \neq max} p(y_i|x)^2 \\ &= \sum_{i \neq max} (p(y_{max}|x) - p(y_i|x)) * p(y_i|x) \end{aligned} \tag{2.6}$$

we know that $p(y_{max}|x)$ is the largest among all posterior probability, so for each $p(y_{max}|x) - p(y_i|x) \geq 0$ always valid, then we have $R_{ran}(\alpha(x)|x) - R_{Bayes}(\alpha(x)|x) \geq 0$ which is $R_{ran}(\alpha(x)|x) \geq R_{Bayes}(\alpha(x)|x)$. So we have reached the conclusion that R_{ran} is alwasy larger than or equal to R_{Bayes}.

3. From 2.6, we know the only situation where $R_{ran} = R_{Bayes}$ is

$$\forall i \quad p(y_{max}|x) - p(y_i|x) = 0$$

which means that all $p(y_i|x)$ are the same. So, when each category has the same posteriro probability, the two decision rules become the same.

3 Problem 3

1. First, we could calculate the posterior distribution of positive sample $p(y = -1|x)$ and negitive $p(y = 1|x)$, ignoring $p(x)$, which is given by:

$$\begin{aligned} p(y = -1|x) = p(x|y = -1) * p(y = -1) = 0.44 * G(\mu_1 = 2, 5^2) \\ p(y = +1|x) = p(x|y = +1) * p(y = +1) = 0.56 * G(\mu_2 = 6, 4^2) \end{aligned} \tag{3.1}$$

and their distributions are shown in figure 3.1.

To draw the ROC curve, we need hit rate which is the positive input that has been classifed as positive, given by the cdf of $p(y = -1|x)$ at T, and false alarm, which is the negitive that has been classified as positive, given by the cdf of $p(y = +1|x)$ at T. Then we can draw the picture at each T having the hit rate and false alarm, shown in figure 3.2.

To draw the PR curve, we need TP, which is the decision that classify positive to positive, given by the cdf of $p(y = -1|x)$ at T; FP, which is the decision that classify negitive to positive, givne by the cdf of $p(y = +1|x)$ at T; FN, which is the decision that classify positive to negitive, given by integrate from T to $+\infty$. Then precision is given by $TP/(TP + FP)$ and recall is given by $TP/(TP + FN)$. Then we could draw at each T, shown in figure 3.3.

2. When we change μ_2 to 4, both ROC and PR would be worse, this is due to the fact that the two distribution become closer and there are more overlaps, and according to textbook, discriminability is defined as $d' = \frac{|\mu_2 - \mu_1|}{\sigma}$, and a higher d' is desirable.

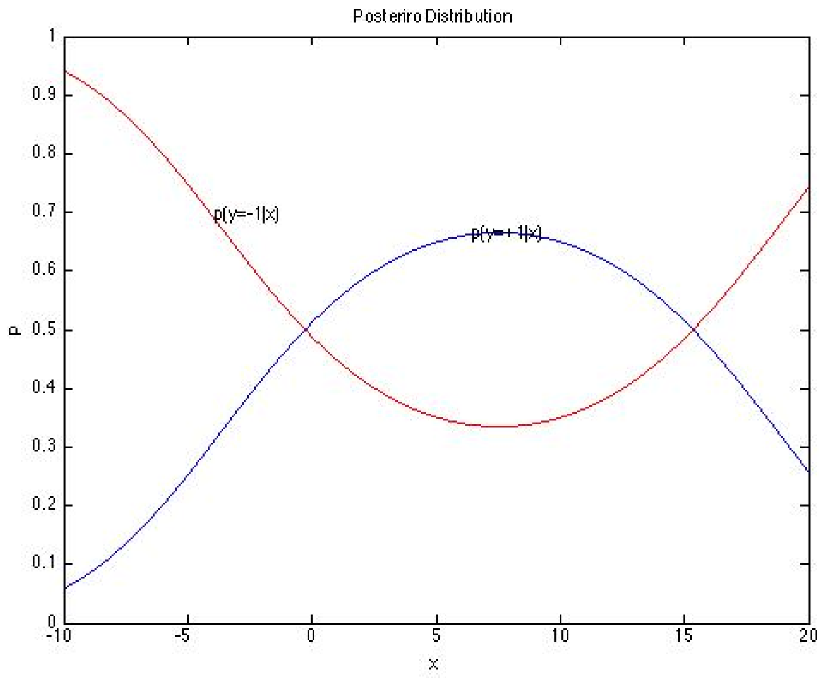

Figure 3.1: Posteriro distribution

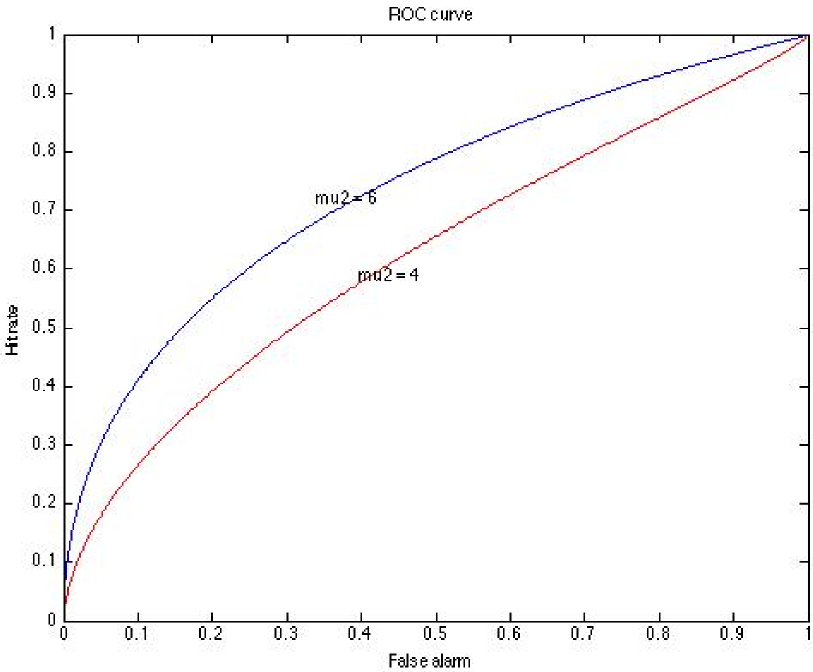

Figure 3.2: ROC curve

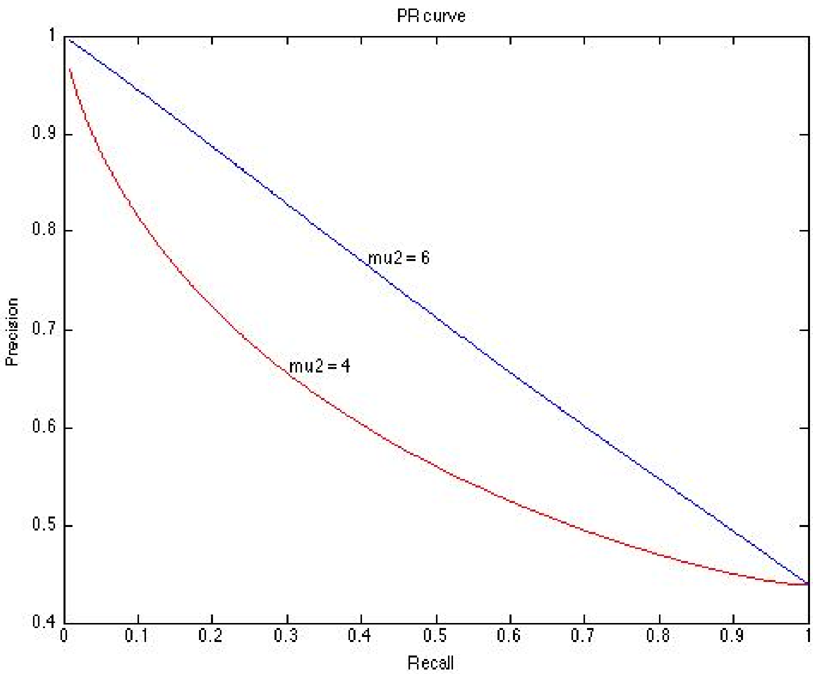

Figure 3.3: PR curve