CSE 802 Spring 2008 Mid-Term Exam

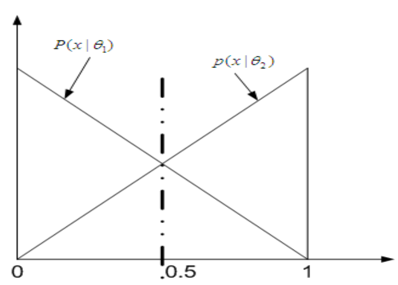
Name:

1. (20 pts) Consider the following two-class classification problem involving a single feature x. Assume equal priors and 0-1 loss function.

$$p(x|\omega_1) = \begin{cases} 2x & \text{for } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

$$p(x|\omega_2) = \begin{cases} 2 - 2x & \text{for } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

   (a) Sketch the two densities on the same plot.
   Solution:



(b) Derive the Bayes decision boundary and show the decision regions associated with classes ω_1 and ω_2 in the plot of (a).

Solution:
Bayes decision boundary is given by:

$$P(\omega_1|x) = P(\omega_2|x) \tag{1}$$
$$\Rightarrow p(x|\omega_1)P(\omega_1) = p(x|\omega_2)P(\omega_2) \tag{2}$$
$$\Rightarrow P(x|\omega_1) = P(x|\omega_2) \qquad \because \text{equal priors} \tag{3}$$
$$\Rightarrow 2x = 2(1-x) \tag{4}$$
$$\Rightarrow x = \frac{1}{2} \tag{5}$$
$$\tag{6}$$

(c) Derive the Bayes classification error for this two-class problem.

Solution:
Bayes classification error is given by

$$\int_0^{1/2} 2x(0.5)dx + \int_{1/2}^1 (2-2x)(0.5)dx = \frac{1}{4} \tag{7}$$

(d) How does the decision boundary move if $P(\omega_1) = 1/4$ and $P(\omega_2) = 3/4$?
Since the decision boundary is given by

$$p(x|\omega_1)P(\omega_1) = p(x|\omega_2)P(\omega_2) \tag{8}$$

$$\Rightarrow \frac{2x}{4} = 2(1-x)\frac{3}{4} \tag{9}$$

$$\Rightarrow x = \frac{3}{4} \tag{10}$$

$$\tag{11}$$

(e) How does the decision boundary move if the loss function is

$$\lambda(1,2) = 2$$
$$\lambda(2,1) = 3$$
$$\lambda(1,1) = \lambda(2,2) = 0$$

$\lambda(\omega_i, \omega_j)$ is the loss incurred in deciding ω_i when the true class is ω_j. Keep $P(\omega_1) = P(\omega_2) = 1/2$

Decision region for ω_1 is given by

$$(\lambda(2,1) - \lambda(1,1))P(\omega_1|x) \;>\; (\lambda(1,2) - \lambda(2,2))P(\omega_2|x) \tag{12}$$

$$\Rightarrow 3P(\omega_1|x) \;>\; 2P(\omega_2|x) \tag{13}$$

$$\Rightarrow 3(2x) \;>\; 2(2-2x) \tag{14}$$

$$\Rightarrow x \;>\; \frac{2}{5} \tag{15}$$

2. (20 pts) Consider a two class (ω_1, ω_2) and two feature (x_1, x_2) classification problem.

$$P(\mathbf{x}|\omega_1) \sim N(\mu_1, \Sigma_1),$$

$$P(\mathbf{x}|\omega_2) \sim N(\mu_2, \Sigma_2),$$

$$P(\omega_1) = P(\omega_2) = \frac{1}{2},$$

$$\mu_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \mu_2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$\Sigma_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \Sigma_2 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

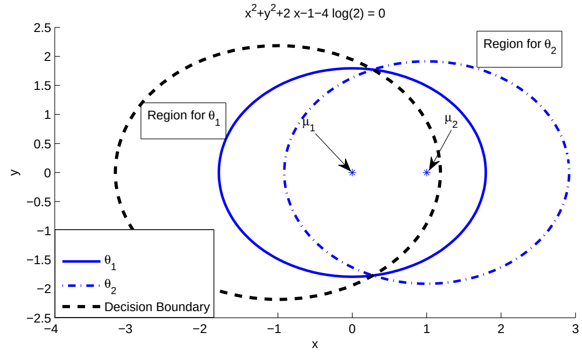
(a) Derive the Bayes decision boundary for the 0-1 loss function.
The descriminant function for the two classes are given by:

$$\begin{aligned}
g_1(x) &= -\frac{1}{2}(x-\mu_1)^t \Sigma_1^{-1}(x-\mu_1) - \frac{1}{2}\ln|\Sigma_1| + \ln P(\omega_1) && (16)\\
&= -\frac{1}{2}(x)^t(x) - \ln 2 && (17)\\
&= -\frac{1}{2}(x_1^2 + x_2^2) - \ln 2 && (18)\\
g_2(x) &= -\frac{1}{2}(x-\mu_2)^t \Sigma_2^{-1}(x-\mu_2) - \frac{1}{2}\ln|\Sigma_2| + \ln P(\omega_2) && (19)\\
&= -\frac{1}{2}(x-\mu_2)^t \Sigma_2^{-1}(x-\mu_2) - \ln 4 && (20)\\
&= -\frac{1}{4}((x_1-1)^2 + x_2^2) - \ln 4 && (21)
\end{aligned}$$

The decision boundary is given by

$$\begin{aligned}
g(x) =& g_2(x) - g_1(x) && (22)\\
=& -\frac{1}{4}((x_1-1)^2 + x_2^2) - \ln 4 + \frac{1}{2}(x_1^2 + x_2^2) + \ln 2 && (23)\\
=& \frac{1}{4}(x_1^2 + x_2^2 + 2x_1) - \frac{1}{4} - \ln 2 && (24)\\
=& 0 && (25)
\end{aligned}$$

(b) Plot the two mean vectors, the ellipse of concentration and the decision boundary in the two dimensional feature space. Also, mark the decision regions.



(c) The Bhattacharya bound is defined as

$$P(error) \leq \frac{1}{2}\exp\{-k(\frac{1}{2})\},$$

$$k(\frac{1}{2}) = \frac{1}{8}(\mu_2-\mu_1)^t\left[\frac{\Sigma_1+\Sigma_2}{2}\right]^{-1}(\mu_2-\mu_1) + \frac{1}{2}\ln\frac{|\frac{\Sigma_1+\Sigma_2}{2}|}{\sqrt{|\Sigma_1||\Sigma_2|}}$$

Compute the error bound for this problem. What is the interpretation of the bound?
The Bhattacharya bound is given by

$$k(1/2) = \frac{1}{8}[1 \quad 0]\frac{2}{3}I[1 \quad 0]^t + \frac{1}{2}\ln\frac{9}{8} \tag{26}$$

$$= \frac{1}{12} + \frac{1}{2}\ln\frac{9}{8} \tag{27}$$

Bhattacharya bound is the upper bound on the probability of error in case of Bayes classifier. However, the test error on a particular data sample can possibly be larger than this bound. As can be seen from the equation, the bound is inversely proportional to difference between the two means and directly proportional to the variance of the two classes.

3. (20 pts) Let θ represent the probability of heads in a coin-tossing problem. A total of n i.i.d observations are available to estimate θ. Note:

$$\int_0^1 \theta^p(1-\theta)^q d\theta = \frac{p!q!}{(p+q+1)!}.$$

(a) Let the prior density of θ be $p(\theta) = 3\theta^2, 0 \leq \theta \leq 1$.

i. Derive the posteriori density of θ.

Solution:
Let X be the set of all observations, posterior probability is given by

$$P(\theta|X) = \frac{p(X|\theta)P(\theta)}{\int p(X|\theta)P(\theta)d\theta} \tag{28}$$

since $p(X|\theta) = \binom{n}{n_H}\Pi_{i=1\ldots n}p(x_i|\theta) = \binom{n}{n_H}\theta^{n_H}(1-\theta)^{(n-n_H)}$ where there are n_H heads and $n - n_H$ tails in the coin tosses. The posterior is thus given as

$$P(\theta|X) = \frac{\binom{n}{n_H}\theta^{n_H}(1-\theta)^{(n-n_H)}\theta^2}{\binom{n}{n_H}\frac{!(n_H+2)!(n-n_H)}{!(n_H+n-n_H+3)}} \tag{29}$$

$$= \frac{(n+3)!}{(n_H+2)!(n-n_H)!}\theta^{(n_H+2)}(1-\theta)^{(n-n_H)} \tag{30}$$

ii. What is the Bayes estimate of θ under a squared-error loss function?

Solution:
The Bayes estimate of theta is obtained as

$$\hat{\theta} = \int \theta P(\theta|X)d\theta \tag{31}$$

$$= \frac{(n+3)!}{(n_H+2)!(n-n_H)!}\int \theta^{n_H+3}(1-\theta)^{(n-n_H)}d\theta \tag{32}$$

$$= \frac{(n+3)!}{(n_H+2)!(n-n_H)!}\frac{(n_H+3)!(n-n_H)!}{(n+4)!} \tag{33}$$

$$= \frac{(n_H+3)}{(n+4)} \tag{34}$$

iii. What is the Bayes estimate of θ under a 0-1 loss function?

Solution:
The Bayes estimate of theta is obtained as

$$\begin{aligned}
\hat{\theta} = & \quad argmax_\theta P(\theta|X) & (35)\\
= & \quad argmax_\theta \ln P(\theta|X) & (36)\\
= & \quad argmax_\theta \ln \frac{(n+3)!}{(n_H+2)!(n-n_H)!}\theta^{(n_H+2)}(1-\theta)^{(n-n_H)} & (37)\\
= & \quad argmax_\theta (n_H+2)\ln\theta + (n-n_H)\ln(1-\theta) & (38)
\end{aligned}$$

taking derivative and equating to zero
$\hat{\theta} = \frac{n_H+2}{n+2}$

(b) How does the Bayes estimate of θ in part (a) compare to the estimates derived using the prior $p(\theta) = \theta$ in HW3 and $p(\theta) = 1$ worked out in the class, where $0 \leq \theta \leq 1$. Why are these estimates different? When will these estimates of θ become (almost) identical?

Solution:
The prior $p(\theta) = 1$ leads to the Bayes estimate of $\hat{\theta_{(1)}} = \frac{n_H+1}{n+2}$, $p(\theta) = \theta$ leads to the Bayes estimate of $\hat{\theta_{(1)}} = \frac{n_H+2}{n+3}$, and $p(\theta) = \theta^2$ leads to the Bayes estimate of $\hat{\theta_{(1)}} = \frac{n_H+3}{n+4}$. Each offsetting $\hat{\theta}$ to a value closer to 1 increasingly as the power of θ in the prior increases. Note that $p(\theta) = \theta^k$ is an increasing function with slope proportional to k. However, the estimates will be identical if $n \to \infty$.

4. (20 pts) Consider a c-class classification problem with rejection option. Let

$$\lambda(\alpha_i|\omega_j) = \begin{cases} 0 & i = j \quad i,j = 1,...,c \\ \lambda_r & i = c+1 \\ \lambda_s & \text{otherwise,} \end{cases}$$

where $\lambda(\alpha_i|\omega_j)$ is the loss incurred in taking action α_i given the true class is ω_j; λ_r is the loss incurred for choosing the (c+1)th action, i.e. rejection, and λ_s is the loss incurred for making any substitution error. In HW 2 it was shown that the minimum risk is obtained if we decide ω_i if $P(\omega_i|\mathbf{x}) \geq P(\omega_j|\mathbf{x})$ for all j and $P(\omega_i|\mathbf{x}) \geq 1 - \lambda_r/\lambda_s$, and reject otherwise.

(a) Verify if the following is an optimal discriminant function:

$$g_i(\mathbf{x}) = \begin{cases} p(\mathbf{x}|\omega_i)P(\omega_i) & i = 1,...,c \\ \\ \frac{\lambda_s - \lambda_r}{\lambda_s} \sum_{j=1}^{c} p(\mathbf{x}|\omega_j)P(\omega_j) & i = c+1 \end{cases}$$

Solution:
Using the above discriminant function, class ω_i is selected if $g_i(\mathbf{x}) \geq g_j(\mathbf{x})$ for all $j = 1,...,c+1$ i.e. if
$p(x|\omega_i)P(\omega_i) \geq p(\mathbf{x}|\omega_k)P(\omega_k)$ for all $k = 1,...,c$ and
$p(\mathbf{x}|\omega_i)P(\omega_i) \geq \frac{\lambda_s - \lambda_r}{\lambda_s} \sum_{j=1}^{c} p(\mathbf{x}|\omega_j)P(\omega_j)$.

Since
$p(\mathbf{x}|\omega_i)P(\omega_i) = P(\omega_i|\mathbf{x})p(x)$ and
$\sum_{j=1}^{c} p(\mathbf{x}|\omega_j)P(\omega_j) = p(x)$,

the conditions become $P(\omega_i|\mathbf{x}) \geq P(\omega_k|\mathbf{x})$ for all $k = 1,...,c$ and $P(\omega_i|\mathbf{x}) \geq 1 - \lambda_r/\lambda_s$ which is the required condition.

(b) Plot these discriminant functions and the decision regions for the two-category one-dimensional case having

- $p(\mathbf{x}|\omega_1) \sim N(1,1)$,
- $p(\mathbf{x}|\omega_2) \sim N(-1,1)$,
- $P(\omega_1) = P(\omega_2) = 1/2$, and
- $\lambda_r/\lambda_s = 1/4$.

Solution:
Note that

$$p(x) = \sum_{j=1}^{2} p(\mathbf{x}|\omega_j)P(\omega_j) \tag{39}$$

$$= \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2\sigma_1^2}(x-\mu_1)^2}\frac{1}{2} + \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{1}{2\sigma_2^2}(x-\mu_2)^2}\frac{1}{2} \tag{40}$$

$$= \frac{1}{2\sqrt{2\pi}}\left\{e^{-\frac{(x-1)^2}{2}} + e^{-\frac{(x+1)^2}{2}}\right\} \tag{41}$$

The detection region for the class 1 is obtained when $P(\omega_1|\mathbf{x}) \geq 1-\lambda_r/\lambda_s$ and $P(\omega_1|\mathbf{x}) \geq P(\omega_2|\mathbf{x})$.

$$\frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2\sigma_1^2}(x-\mu_1)^2}\frac{1}{2} \geq \{1-\lambda_r/\lambda_s\}\frac{1}{2\sqrt{2\pi}}\left\{e^{-\frac{(x-1)^2}{2}} + e^{-\frac{(x+1)^2}{2}}\right\} \text{ AND } \quad |x-1| \leq |x+1| \tag{42}$$

$$\text{i.e. } e^{-\frac{(x-1)^2}{2}} \geq \frac{3}{4}\left\{e^{-\frac{(x-1)^2}{2}} + e^{-\frac{(x+1)^2}{2}}\right\} \text{ AND } \quad x \geq 0 \tag{43}$$

$$\text{i.e. } e^{-\frac{(x-1)^2}{2}} \geq 3e^{-\frac{(x+1)^2}{2}} \text{ AND } \quad x \geq 0 \tag{44}$$

$$\text{i.e. } x \geq 1/2\ln(3) \text{ AND } \quad x \geq 0 \tag{45}$$

similarly, region for class 2 is obtained when $P(\omega_2|\mathbf{x}) \geq 1-\lambda_r/\lambda_s$ and $P(\omega_2|\mathbf{x}) \geq P(\omega_1|\mathbf{x})$.

$$\frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{1}{2\sigma_2^2}(x-\mu_2)^2}\frac{1}{2} \geq \{1-\lambda_r/\lambda_s\}\frac{1}{2\sqrt{2\pi}}\left\{e^{-\frac{(x-1)^2}{2}} + e^{-\frac{(x+1)^2}{2}}\right\} \text{ AND } \quad |x+1| \leq |x-1| \tag{46}$$

$$\text{i.e. } e^{-\frac{(x+1)^2}{2}} \geq \frac{3}{4}\left\{e^{-\frac{(x-1)^2}{2}} + e^{-\frac{(x+1)^2}{2}}\right\} \text{ AND } \quad x \leq 0 \tag{47}$$

$$\text{i.e. } e^{-\frac{(x+1)^2}{2}} \geq 3e^{-\frac{(x-1)^2}{2}} \text{ AND } \quad x \leq 0 \tag{48}$$

$$\text{i.e. } x \leq -1/2\ln(3) \text{ AND } \quad x \leq 0 \tag{49}$$

(c) Describe quantitatively what happens as λ_r/λ_s is increased from 0 to 1.

Solution:
When $\lambda_r/\lambda_s = 0$, then the condition for deciding a particular class i.e. $P(\omega_i|\mathbf{x}) \geq 1 - \lambda_r/\lambda_s$ is never satisfied. Thus all the samples are rejected. But when $\lambda_r/\lambda_s = 1$, $g_{c+1}(x) = 0$ thus no sample is ever rejected.

5. (20 pts)

   (a) State and explain the difference between the following terms. Provide relevant figures in (a), (c) and (e).

      i. PCA vs LDA

      ii. MLE vs MAP parameter estimates

      iii. Inter-class variability and intra-class variability

      iv. Neyman-Pearson vs Bayes classifier

(b) State the classification problem used by Trunk in his paper to demonstrate the existence of the curse of dimensionality. Summarize his main results.