CSE 569 Homework #1 Solutions

Q4. Depending on how you factor the joint density via conditional densities, there are different (and equally correct) ways of finding the solutions. Following solutions are just for your reference.

(a)

$$\begin{aligned}
P(w_0|x_1) &= P(w_0, x_1)/P(x_1) \\
&= \sum_{Y,Z} P(x_1, Y, Z, w_0)/P(x_1) \\
&= \sum_{Y,Z} P(x_1)P(Y|x_1)P(Z|Y)P(w_0|Z)/P(x_1) \\
&= P(y_0|x_1)P(z_0|y_0)P(w_0|z_0) \\
&\quad + P(y_1|x_1)P(z_0|y_1)P(w_0|z_0) \\
&\quad + P(y_0|x_1)P(z_1|y_0)P(w_0|z_1) \\
&\quad + P(y_1|x_1)P(z_1|y_1)P(w_0|z_1) \\
&= (1-0.4)(1-0.6)(1-0.3) \\
&\quad + (0.4)(1-0.25)(1-0.3) \\
&\quad + (1-0.4)(0.6)(1-0.45) \\
&\quad + (0.4)(0.25)(1-0.45) \\
&= \boxed{0.631}
\end{aligned}$$

(b)

$$\begin{aligned}
P(x_0|w_1) &= P(x_0, w_1)/P(w_1) \\
&= \frac{\sum_{Y,Z} P(x_0)P(Y|x_0)P(Z|Y)P(w_1|Z)}{\sum_{X,Y,Z} P(X)P(Y|X)P(Z|Y)P(w_1|Z)} \\
&= \frac{\sum_{Y,Z} P(x_0)P(Y|x_0)P(Z|Y)P(w_1|Z)}{\sum_{Y,Z} P(x_0)P(Y|x_0)P(Z|Y)P(w_1|Z) + \sum_{Y,Z} P(x_1)P(Y|x_1)P(Z|Y)P(w_1|Z)} \\
&= \boxed{0.403}
\end{aligned}$$

Q5. False.

Brief explanation: We considered two special cases of Gaussian densities where the decision boundaries are linear (hyperplanes). In general, if the covariance matrices for all classes are arbitrary and not the same, the decision boundaries will be hyperquadrics (and in general non-linear).

Q7. Part A. Search on-line for "Monty Hall Problem" for solutions to this problem or its variants.

Part B. This is an example of Monte Carlo simulation, simulating random experiments for solving a problem. You can ask questions in our office hours if you have no clue how to do the simulation.

Q1 (a) $P(\text{error}) = P(w_1$ is the true class but the decision is w_2 (i.e., $x \le \theta$))
$+ P(w_2$ is the true class but the decision is w_1 (i.e., $x > \theta$))

$$= P(w_1) P(x \le \theta | w_1) + P(w_2) P(x > \theta | w_2)$$

$$= P(w_1) \int_{-\infty}^{\theta} p(x|w_1) dx + P(w_2) \int_{\theta}^{\infty} p(x|w_2) dx$$

(b) Taking the derivative, and setting it to zero,

$$\frac{\partial P(\text{error})}{\partial \theta} = P(w_1) p(\theta|w_1) - P(w_2) p(\theta|w_2) = 0$$

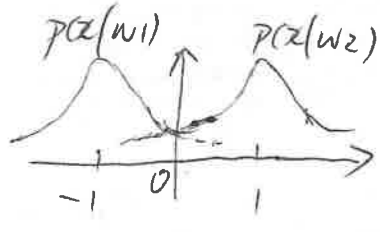
(For this step, you need to review "derivatives of integrals", and note that the second term has θ in the lower limit, hence the minus sign after taking the derivative.)

$\therefore$ We have $P(w_1) p(\theta|w_1) = P(w_2) p(\theta|w_2)$

(c). No, the condition does not uniquely define θ.

E.g., if $P(w_1) = P(w_2)$, and $p(x|w_1) = p(x|w_2)$ $\forall x \in [a, b]$, then θ can take any value in $[a, b]$.

(d) Let $P(w_1) = P(w_2) = \frac{1}{2}$, $p(x|w_1) \sim N(-1, 1)$, $p(x|w_2) \sim N(1, 1)$



Then $\theta = 0$ satisfies the condition.

But the rule "Decide w_1 if $x > 0$" will give the worst case result.

Q2. The optimal decision is given by

Decide ω_1 if $(\lambda_{21}-\lambda_{11})P(\omega_1)p(x|\omega_1) > (\lambda_{12}-\lambda_{22})P(\omega_2)p(x|\omega_2)$

otherwise decide ω_2 ★

Plug in all the given values, we have the above ★ simplified to

$$\lambda_{21}p(x|\omega_1) > 2\lambda_{12}p(x|\omega_2)$$

$$\Leftrightarrow \quad \lambda_{21}\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu_1)^2}{2\sigma^2}\right) > \lambda_{12}\frac{2}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu_2)^2}{2\sigma^2}\right)$$

$$\Leftrightarrow \quad \exp\left(\frac{(x-\mu_2)^2-(x-\mu_1)^2}{2\sigma^2}\right) > \frac{2\lambda_{12}}{\lambda_{21}}$$

taking log

$$\Leftrightarrow \quad \frac{2x(\mu_1-\mu_2)+(\mu_2^2-\mu_1^2)}{2\sigma^2} > \log\frac{2\lambda_{12}}{\lambda_{21}}$$

This can be further simplified to (if we assume $\mu_1 > \mu_2$)

$$x > \frac{\sigma^2}{\mu_1-\mu_2}\log\frac{2\lambda_{12}}{\lambda_{21}} + \frac{\mu_1+\mu_2}{2}$$

Therefore, the decision rule is,

Decide ω_1 if $x \in R_1$; otherwise decide ω_2,

with R_1 given by $R_1 = \left\{x \,\middle|\, x > \frac{\sigma^2}{\mu_1-\mu_2}\log\frac{2\lambda_{12}}{\lambda_{21}} + \frac{\mu_1+\mu_2}{2}\right\}$

Q3. The optimal decision rule is given by

Decide w_1 if $\underline{P(w_1)p(x|w_1) > P(w_2)p(x|w_2)}$ ☆

Otherwise decide w_2

The condition ☆ becomes $2\,p(x|w_1) > p(x|w_2)$, since $P(w_1)=\frac{2}{3}$.

This is always true for $x<0$ or $x>3$, since $p(x|w_2)=0$ if $\begin{cases} x<0 \\ \text{or} \\ x>3 \end{cases}$.

On $[0,3]$, the condition is $\frac{2}{\sqrt{2\pi}}\exp\left[-\frac{(x-1)^2}{2}\right] > \frac{1}{3}$.

Then we can figure out $x < 2.32$ (roughly)

In summary, we decide w_1 for $x<2.32$ or $x>3$

decide w_2 for $x \in [2.32, 3]$.

The Bayes error is given by

$$\underbrace{P(w_1)\int_{2.32}^{3} p(x|w_1)\,dx}_{\substack{\text{error for classifying} \\ \text{an } X \text{ from } w_1 \text{ into } w_2}} + \underbrace{P(w_2)\int_{0}^{2.32} p(x|w_2)\,dx}_{\substack{\text{error for classifying} \\ \text{an } X \text{ from } w_2 \text{ into } w_1}}$$

The second term is easy, $= \frac{1}{3} \times 2.32 \times \frac{1}{3}$

For the first term, you will need to refer to the tables for CDF of a standard normal density.

Q6. (a) The minimum error is given by the Bayes rule for decision, i.e., decide w_1 if $P(w_1)p(x|w_1) > P(w_2)p(x|w_2)$

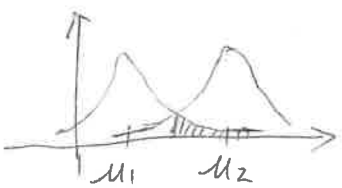
Plug in the give $P(w_i)$ & density, we have

$$\frac{1}{2}\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-\mu_1)^2}{2\sigma^2}\right] > \frac{1}{2}\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-\mu_2)^2}{2\sigma^2}\right]$$

This can be simplified to $(x-\mu_1)^2 < (x-\mu_2)^2$

i.e., we decide w_1 if $|x-\mu_1| < |x-\mu_2|$.

Without loss of generality, assume $\mu_1 < \mu_2$, then the error is given by



$$p[\text{error}] = P(w_1)P\left[|x-\mu_1| > |x-\mu_2| \text{ while true class is } w_1\right]$$
$$+ P[w_2]P\left[|x-\mu_1| < |x-\mu_2| \text{ while true class is } w_2\right]$$

The two terms are equal, see the figure illustrated on the left for the probability in the first term

$$\therefore p[\text{error}] = 2\times\frac{1}{2}\frac{1}{\sqrt{2\pi}\sigma}\int_{\frac{\mu_1+\mu_2}{2}}^{\infty} e^{-\frac{(x-\mu)}{2\sigma^2}}dx,$$

do a change of variable by $y = \frac{x-\mu}{\sigma}$

$$= \frac{1}{\sqrt{2\pi}}\int_{\frac{\mu_2-\mu_1}{2\sigma}}^{\infty} e^{-y^2/2}dy$$

(b) From (a) $p[\text{error}] \leq \frac{1}{\sqrt{2\pi}a}e^{-\frac{a^2}{2}}$, using the given inequality

where $a = \frac{\mu_2-\mu_1}{2\sigma}$

But $e^{-\frac{a^2}{2}} \to 0$ when $a \to \infty$, or $\frac{\mu_2-\mu_1}{2\sigma} \to \infty$.

$\therefore p\{\text{error}\} \to 0$, when $\frac{\mu_2-\mu_1}{2\sigma} \to \infty$.

This is to say, for finite σ, if the means go far away from each other, the error diminishes, very intuitive.