April 11, 1:00pm PDT
TO BE PROOFREAD

STAT 161/261: Homework 1 Solutions
Bayes Decision Theory.

Prof. Alyson Fletcher

1. Recommended problems: Problem 1.1 and 1.5, Alpayin. **Do not submit**.

2. (a) This problem is a classic application of total probability. Let F, DH and DT be the events that the selected coin is fair, double-headed or double-tailed. We have

$$P(F) = 3/5, \quad P(DH) = P(DT) = 1/5.$$

Let Top $= H$ or T indicate whether the top of the coin is a head or tail. We know,

$$P(\text{Top} = H|F) = \frac{1}{2}, \quad P(\text{Top} = H|DH) = 1, \quad P(\text{Top} = H|DT) = 0.$$

Now using total probability,

$$\begin{aligned} &P(\text{Top} = H) \\ &\quad = P(\text{Top} = H|F)P(F) + P(\text{Top} = H|DH)P(DH) + P(\text{Top} = H|DT)P(DT) \\ &\quad = \frac{1}{2}\frac{3}{5} + (1)\frac{1}{5} + (0)\frac{1}{5} = \frac{1}{2}. \end{aligned}$$

(b) For this part, we use the conditional probability formula

$$P(\text{Bot} = \text{H}|\text{Top} = \text{H}) = \frac{P(\text{Bot} = \text{H}, \text{Top} = \text{H})}{P(\text{Top} = \text{H})}.$$

Now, the top and the bottom of the coin is a head if and only if you select the double-headed coin. So, $P(\text{Bot} = \text{H}, \text{Top} = \text{H}) = P(DH) = 1/5$. Also, from part (a), $P(\text{Top} = \text{H}) = 1/2$. Hence,

$$P(\text{Bot} = \text{H}|\text{Top} = \text{H}) = \frac{1/5)}{1/2} = \frac{2}{5}.$$

3. (a) The likelihood ratio is

$$T(x) = \frac{p(x|y=1)}{p(x|y=0)} = \frac{3x^2/2}{1/2} = 3x^2.$$

The ML classifier takes $\widehat{y} = 1$ if and only if $T(x) \geq 1$,

$$\widehat{y} = \begin{cases} 1 & \text{if } |x| \geq 1/\sqrt{3}, \\ 0 & \text{if } |x| < 1/\sqrt{3}. \end{cases}$$

(b) Let $\gamma = 1/\sqrt{3}$ be the above threshold level. The probability of mis-detection is

$$\begin{aligned} P_{MD} &= P(\widehat{y} = 0|y = 1) = P(|x| \leq \gamma | y = 1) \\ &= \int_{-\gamma}^{\gamma} p(x|y=1)dx = 3\int_0^{\gamma} x^2 dx = \gamma^3 = (3)^{-3/2}. \end{aligned}$$

The probability of false alarm is

$$\begin{aligned} P_{FA} &= P(\widehat{y} = 1|y = 0) = P(|x| \geq \gamma | y = 0) = P(|x| \geq \gamma | y = 0) \\ &= 2\int_{\gamma}^{1} p(x|y=0)dx = \int_{\gamma}^{1} dx = 1 - \gamma = 1 - \frac{1}{\sqrt{3}}. \end{aligned}$$

(c) The expected Bayes risk is

$$\begin{aligned} &\mathbb{E}\left[L(\widehat{y}, y)\right] = c_1 \Pr(\widehat{y} = 1, y = 0) + c_2 \Pr(\widehat{y} = 0, y = 1) \\ &\quad = c_1 \Pr(\widehat{y} = 1|y = 0)P(y = 0) + c_2 \Pr(\widehat{y} = 0|y = 1)P(y = 1) \\ &\quad = c_1 P_{FA}(1-q) + c_2 P_{MD} q, \end{aligned}$$

where P_{FA} and P_{MD} are given above.

4. (a) In this problem, it is easier to use the log-likelihood ratio given by,

$$T(x) = \ln p(x|y=1) - \ln p(x|y=0).$$

Now, the density of a scalar Gaussian is

$$p(x|y=i) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-(x-\mu_i)^2/(2\sigma_i^2)},$$

so the log likelihood ratio is

$$T(x) = \frac{1}{2}\ln\left[\frac{\sigma_0^2}{\sigma_1^2}\right] + \frac{(x-\mu_0)^2}{2\sigma_0^2} - \frac{(x-\mu_1)^2}{2\sigma_1^2}. \tag{1}$$

The ML classifier takes $\widehat{y} = 1$ if and only if $T(x) \geq 0$.

In this problem, we assume $\sigma_1 > \sigma_0$ and $\mu_1 > \mu_0$. In this case, $T(x)$ is an upward facing (i.e. convex) parabola. This parabola will always have two roots, which we will call x_1^* and x_2^* with $x_2^* > x_1^*$. Since $T(x)$ is upward facing, $T(x) \leq 0$ for $x \in [x_1^*, x_2^*]$ and $T(x) > 0$ otherwise. Hence, the classifier is

$$\widehat{y} = \begin{cases} 0 & \text{if } x \in [x_1^*, x_2^*], \\ 1 & \text{else.} \end{cases}$$

For this problem, if you just said that $T(x)$ has two roots, you will get full marks. But, you can show that $T(x)$ will always have two roots as follows: Suppose that $T(x)$ doesn't have two roots. Since it is a quadratic, it must have either one or zero roots. Moreover, since $T(x)$ is an upward facing parabola it would then never cross the x-axis and hence would always be positive. But, $T(x) \geq 0$ for all x would imply that $p(x|y=1) \geq p(x|y=0)$ for all x which is impossible since $\int p(x|y)dx = 1$ for any y.

(b) We first compute P_{MD}:

$$P_{MD} = P(\widehat{y} = 0|y = 1) = P(x \in [x_1^*, x_2^*]|y = 1).$$

To compute this probability in terms of the Q-function, first note that

$$Q(\alpha) = \Pr(z \geq \alpha),$$

for any standard normal variable $z \sim \mathcal{N}(0,1)$. Now, we use a standard trick and let $z = (x - \mu_1)/\sigma_1$. Conditional on $y = 1$, $x \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and hence $z \sim \mathcal{N}(0,1)$. Thus,

$$\begin{aligned} P_{MD} &= P(x \in [x_1^*, x_2^*]|y = 1) \\ &= P(z \geq \frac{x_1^* - \mu_1}{\sigma_1}) - P(z \geq \frac{x_2^* - \mu_1}{\sigma_1}) \\ &= Q\left(\frac{x_1^* - \mu_1}{\sigma_1}\right) - Q\left(\frac{x_2^* - \mu_1}{\sigma_1}\right). \end{aligned}$$

Similarly, for the false alarm rate,

$$P_{FA} = P(x \notin [x_1^*, x_2^*]|y = 0) = 1 - P(x \in [x_1^*, x_2^*]|y = 0) 1 - Q\left(\frac{x_1^* - \mu_0}{\sigma_0}\right) + Q\left(\frac{x_2^* - \mu_0}{\sigma_0}\right).$$

(c) This is identical to the previous problem: The expected Bayes risk is

$$\mathbb{E}\left[L(\widehat{y}, y)\right] = c_1 P_{FA}(1 - q) + c_2 P_{MD} q,$$

where P_{FA} and P_{MD} are given above.

5. (a) This is a special case of the previous problem, but with $\sigma_1 = \sigma_2 = \sigma$. In this case, $T(x)$ in (1) is a linear function given by

$$T(x) = \frac{1}{2\sigma^2}\left[2(\mu_1 - \mu_0)x + (\mu_0^2 - \mu_1^2)\right].$$

We select $\widehat{y} = 1$ when $T(x) \geq 0$ which will occur when

$$\widehat{y} = \begin{cases} 1 & \text{if } x \geq x^*, \\ 0 & \text{if } x < x^*, \end{cases} \quad x^* = \frac{\mu_0 + \mu_1}{2}.$$

In this case $x^* = (0 + 1)/2 = 0.5$.

(b) For the remaining parts, see the published MATLAB script.

6. Problem 2.7 from Alpaydin, 3rd edition. We have the loss function,

$$E := \frac{1}{N}\sum_{t=1}^{N}\left[r^t - (w_1 x^t - w_0)\right]^2.$$

To minimize the function we take the partial derivatives,

$$\frac{\partial E}{\partial w_0} = 0 \Rightarrow \frac{2}{N}\sum_{t=1}^{N}(r^t - (w_1 x^t - w_0)) = 0 \Rightarrow \bar{r} = w_1\bar{x} + w_0$$
$$\frac{\partial E}{\partial w_1} = 0 \Rightarrow \frac{2}{N}\sum_{t=1}^{N}(r^t - (w_1 x^t - w_0))x^t = 0 \Rightarrow \overline{rx} = w_1\overline{x^2} + w_0\bar{x}.$$

Solving the two equations and two unknowns,

$$w_0 = \bar{r} - w_1\bar{x}, \quad w_1 = \frac{\overline{xr} - \bar{x}\bar{r}}{\overline{x^2} - \bar{x}^2}.$$