Question 2

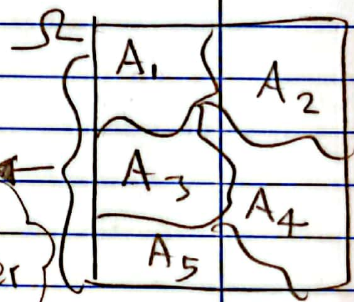
Red box → 3 apples 4 oranges 3 limes (r)
Green box → 3 apples, 3 oranges 4 limes (g)
Blue box → 1 apple, 1 orange, 0 limes (b)

Probability of choosing each box: $p(r) = 0.2$, $p(g) = 0.6$
$p(b) = 0.2$

a) Probability of selecting an apple?

If Ω is the sample space (i.e. all possible outcomes) and $A_1, A_2, \dots, A_n$ are events that partition it, then:

$$P(X) = \sum_{i=1}^{n} P(X \cap A_i)$$



exactly one of A_i occurs whenever X occurs

Also:

$$P(X \cap A_i) = P(X|A_i)P(A_i)$$

(definition of conditional probability)

$$\Rightarrow P(X) = \sum_{i=1}^{n} P(X|A_i)P(A_i)$$ (total probability theorem)

Let A = "We select an apple". Then, based on the above:

$$P(A) = P(A|r) \cdot P(r) + P(A|g)P(g) + P(A|b)P(b)$$

$P(A|r)$ → given box r what is the probability of picking an apple?

$$P(A|r) = \frac{\text{number of apples in r}}{\text{number of fruit pieces in r}}$$

$$= \frac{3}{10}$$

Similarly: $P(A|g) = \frac{3}{10}$, $P(A|b) = \frac{1}{2}$

Then: $P(A) = \frac{3}{10} \cdot \frac{2}{10} + \frac{3}{10} \cdot \frac{6}{10} + \frac{1}{2} \cdot \frac{2}{10}$

$$= \frac{6}{100} + \frac{18}{100} + \frac{1}{10} = \frac{34}{100} = 0.34$$

b) If we observe that the selected fruit is an orange, what is the probability that it came from the green box?

Let O = "The selected fruit is an orange"

This is again <u>conditional</u> probability:

$P(g|O) \rightarrow$ given that we picked an orange, what is the probability it came from g?

<u>Bayes' rule</u>: $P(g|O) = \frac{P(O|g) \cdot P(g)}{P(O)}$

$$P(O|g) = \frac{\text{number of oranges in } g}{\text{pieces of fruit in } g} = \frac{3}{10}$$

$$P(O) = P(O|g)P(g) + P(O|r)P(r) + P(O|b)P(b) =$$

$$= \frac{3}{10} \cdot \frac{6}{10} + \frac{4}{10} \cdot \frac{2}{10} + \frac{1}{2} \cdot \frac{2}{10} = \frac{18}{100} + \frac{8}{100} + \frac{1}{10}$$

$$= \frac{36}{100}$$

$$P(O|g) = \frac{\frac{3}{10} \cdot \frac{6}{10}}{\frac{36}{100}} = \frac{18}{36} = \frac{1}{2}$$

Question 3

$C = \{C_1, \dots, C_N\} \rightarrow$ classes of the problem

$$L_{kj} = \begin{cases} 0, & \text{if } k=j \text{ (correct prediction)} \\ l_r, & \text{if } j=N+1 \text{ (rejection choice)} \\ l_s, & \text{otherwise (missclassification)} \end{cases}$$

a) Expected loss: $E[L] = \sum_k \sum_j \int_{R_j} L_{kj}\, p(x, C_k)\, dx$

$\rightarrow R_j$ is the region of the input space where our classifier outputs <<class j>>

$\Rightarrow$ Sum over j examines each region, sum over k examines each ground truth class

Rewriting: $E[L] = \sum_j \int_{R_j} \sum_k L_{k,j}\, p(x, C_k)\, dx$

$= \sum_j \int_{R_j} \sum_k L_{k,j}\, p(x) p(C_k|x)\, dx$

$\therefore$ Summation terms don't affect each other $\Rightarrow$ to minimize the sum minimize each term $\Rightarrow$ minimize "pointwise" over $x \Rightarrow$ for each x, choose j so that:

$$\arg\min_j \Big\{ \sum_k L_{k,j}\, p(C_k|x)\, \underbrace{p(x)}_{\text{doesn't change with } j} \Big\} = \arg\min_j \Big\{ \sum_k L_{kj}\, p(C_k|x)$$

If $j \neq N+1$:

$$\sum_k L_{kj}\, p(C_k|x) = \sum_{k \neq j} L_{k,j}\, p(C_k|x) + \overset{0}{\cancel{L_{jj}}}\, p(C_j|x)$$

$$= \sum_{k \neq j} l_s\, p(C_k|x) = l_s \sum_{k \neq j} p(C_k|x) = l_s(1 - p(C_j|x$$

If $j = N+1$ (reject), the cost is l_r

$\Rightarrow$ Cost l_r when choosing reject, cost $l_s(1-p(C_j|x))$ when choosing j

$\Rightarrow$ To minimize, we should choose «reject» iff:

$$\ell_r < \ell_s(1-p(c_j|x)) \quad \forall j \quad \overset{\ell_s \neq 0}{\Longleftrightarrow}$$

$$\frac{\ell_r}{\ell_s} < 1-p(c_j|x) \Rightarrow p(c_j|x) < 1 - \frac{\ell_r}{\ell_s}$$

If $p(c_j|x) \geq 1 - \frac{\ell_r}{\ell_s}$ then choose j so that $\ell_s(1-p(c_j|x))$ is minimum $\Rightarrow$ $p(c_j|x)$ is maximum

Therefore the optimal classifier is:

$$\text{classify}(x) = \begin{cases} c_j & \text{iff } p(c_i|x) \leq p(c_j|x) \ \forall i \text{ and } p(c_j|x) \geq 1 - \frac{\ell_r}{\ell_s} \\ c_{rej} & \text{otherwise} \end{cases}$$

b) If $\ell_r = 0$ then we always reject unless $p(c_j|x) = 1$ («absolute certainty»)

c) If $\ell_r > \ell_s$ then $1 - \frac{\ell_r}{\ell_s} < 0 \Rightarrow$ we never reject

Question 4

$$p(x|\theta) = \theta^2 x e^{-\theta x} g(x), \quad g(x) = \begin{cases} 1, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

N measurements $x_1, x_2, \dots, x_N > 0$

What is $\tilde{\theta}$ of MLE?

$\tilde{\theta}$ is such that $p(x_1, x_2, \dots, x_N|\tilde{\theta})$ is maximum

$$p(x_1, x_2, \dots, x_N|\theta) = p(x_1|\theta)\, p(x_2|\theta) \dots p(x_N|\theta)$$

$$= \theta^2 x_1 e^{-\theta x_1} g(x_1)\, \theta^2 x_2 e^{-\theta x_2} g(x_2) \dots \theta^2 x_N e^{-\theta x_N} g(x_N)$$

$\Rightarrow g(x_i) = 1$ since $x_i > 0$. Then:

$$p(x_1, \dots, x_N|\theta) = \theta^{2N} \prod_{i=1}^{N} x_i\, e^{-\theta \sum_{i=1}^{N} x_i}$$

To maximize wrt. θ we need the derivative wrt. θ

$$\frac{\partial}{\partial\theta}\left(p(x_1,\dots,x_N|\theta)\right) = \frac{\partial}{\partial\theta}\left(\theta^{2N}\prod_{i=1}^{N} x_i\, e^{-\theta\sum_{i=1}^{N} x_i}\right)$$

$$= \prod_{i=1}^{N} x_i \frac{\partial}{\partial\theta}\left(\theta^{2N} e^{-\theta\sum_{i=1}^{N} x_i}\right) = \left(\theta^{2N-1} e^{-\theta\sum_{i=1}^{N} x_i}\, 2N\right.$$

$$\left. + \theta^{2N} e^{-\theta\sum_{i=1}^{N} x_i}\left(-\sum_{i=1}^{N} x_i\right)\right)\cdot\prod_{i=1}^{N} x_i$$

Set to 0: $\prod_{i=1}^{N} x_i\left(2N\theta^{2N-1} e^{-\theta\sum_{i=1}^{N} x_i} + \theta^{2N} e^{-\theta\sum_{i=1}^{N} x_i}\left(-\sum_{i=1}^{N} x_i\right)\right) = 0$

$\Rightarrow$ (Divide with $\prod_{i=1}^{N} x_i \neq 0$) $\left(2N + \theta\left(-\sum_{i=1}^{N} x_i\right)\right) e^{-\theta\sum_{i=1}^{N} x_i}\, \theta^{2N-1} = 0$

$\Rightarrow \theta = 0$ or $2N + \theta\left(-\sum_{i=1}^{N} x_i\right) = 0 \Rightarrow \theta = \dfrac{2N}{\sum_{i=1}^{N} x_i}$

$\theta = 0$ is rejected as a possible maximum position since $p(x|0) = 0\ \forall x$ (definitely not maximum likelihood)

Therefore $\boxed{\hat{\theta} = \dfrac{2N}{\sum_{i=1}^{N} x_i}}$

Note: You can also do this by maximizing $\log p(x_1,\dots,x_N|\theta)$, as is common in many cases. In that case just note that this is not defined for $\theta = 0$ and observe that the maximum is not achieved there, so it's not an issue