Homework 1

Yuan Li, N19728558

Part I: Written Exercises

1. (a) Answer:

Set "p_i" as the probability of finding the correct $Label_i$ in a cave, $i = 1, 2, ..., 5$; "p_a" as the probability of incorrectly labeling a crab if the scientist predicts the species uniformly at random.

$$\begin{aligned} p_a &= \sum_{i=1}^{5} 0.2 * (1 - p_i) \\ &= 0.2 * (1 - 0.25) + 0.2 * (1 - 0.05) + 0.2 * (1 - 0.2) + 0.2 * (1 - 0.35) + 0.2 * (1 - 0.15) \\ &= 0.15 + 0.19 + 0.16 + 0.13 + 0.17 \\ &= \mathbf{0.8} \end{aligned}$$

1. (b) Answer:

Yes, the scientist can improve his error by picking the same label every time. The label which the scientist choose is **Label 4**. We can calculate the Bayes error rate. Set the probability as "p_b"

$$\begin{aligned} p_b &= 1 - E(maxp(p_i|p_b)) \\ &= 1 - p_4 \\ &= 1 - 0.35 \\ &= \mathbf{0.65} \end{aligned}$$

2. Answer:

We can bring the value into the following formula:

$$p(x|\mu, \Sigma) = \frac{1}{\sqrt{2\pi}|\Sigma|^{1/2}} exp(-\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu))$$

if $log(p(x|\mu_{blue}, \Sigma_{blue})) > log(p(x|\mu_{orange}, \Sigma_{orange}))$, then the crab's color is blue, otherwise the color is orange.

$$\begin{aligned} log(p(x|\mu_{blue}, \Sigma_{blue})) &= log[\frac{1}{\sqrt{2\pi}|\Sigma_{blue}|^{1/2}} exp(-\frac{1}{2}(x - \mu_{blue})^T {\Sigma_{blue}}^{-1} (x - \mu_{blue}))] \\ &= -637.0431475 \end{aligned}$$

$$\begin{aligned} log(p(x|\mu_{orange}, \Sigma_{orange})) &= log[\frac{1}{\sqrt{2\pi}|\Sigma_{orange}|^{1/2}} exp(-\frac{1}{2}(x - \mu_{orange})^T {\Sigma_{orange}}^{-1} (x - \mu_{orange}))] \\ &= -323.8980596 \end{aligned}$$

Because $log(p(x|\mu_{blue}, \Sigma_{blue})) < log(p(x|\mu_{orange}, \Sigma_{orange}))$, so it's more likely to be **orange**.

3. (a) Answer:

Because θ is the probability of Heads, So:

$$p(Heads) = \theta$$
$$p(Tails) = 1 - \theta$$

Then we can get $p(HHTHH)$ as:

$$\begin{aligned} p(HHTHH) &= \theta * \theta * (1-\theta) * \theta * \theta \\ &= \theta^4 * (1-\theta) \end{aligned}$$

3. (b) Answer:

$$\begin{aligned} logp(HHTHH) &= log(\theta^4 * (1-\theta)) \\ &= log\theta^4 + log(1-\theta) \\ &= 4log\theta + log(1-\theta) \end{aligned}$$

3. (c) Answer:

Because the log function is a monotonically increasing function. so we can calculate the the maximum likelihood estimate of θ as follow:

$$\begin{aligned} 0 &= \frac{\partial logp(HHTHH)}{\partial \theta} \\ &= \frac{\partial}{\partial \theta}[4log\theta + log(1-\theta)] \\ &= \frac{4}{\theta} - \frac{1}{1-\theta} \end{aligned}$$

So we can get the θ should be:

$$\theta = \mathbf{0.8}$$