CK0146 or TIP8311: Homework 01

Exercise 01.01 ✓(1.3 PRML). Suppose that we have three coloured boxes r (red), b (blue), and g (green)

- Box r contains 3 apples (a), 4 oranges (o), and 3 limes (l);
- Box b contains 1 apple (a), 1 orange (o), and 0 limes (l);
- Box g contains 3 apples (a), 3 oranges (o), and 4 limes (l).

If a box is chosen at random with probabilities

- $p(r) = 0.2$;
- $p(b) = 0.2$;
- $p(g) = 0.6$,

and a (piece of) fruit is removed from the box (with equal probability of selecting any of the items in the box), then

1) What is the probability of selecting an apple?

If we observe that the selected fruit is in fact an orange,

2) What is the (posterior) probability that it came from the green box?

Exercise 01.02 ✓. A six-sided die is loaded in a way that each even face is twice as likely as each odd face. All even faces are equally likely, as are all odd faces.

Calculate the probability that the outcome of a single roll of the die is less than 4.

Exercise 01.03 ✓. You enter a special kind of chess tournament, in which you play one game with each of three opponents, but you get to choose the order in which you play your opponents, knowing the probability of a win against each. You win the tournament if you win two games in a row, and you want to maximise the probability of winning.

Show that it is optimal to play the weakest opponent second, and that the order of playing the other two opponents does not matter.

Exercise 01.04 ✓. The release of two out of three prisoners has been announced, but their identity is kept secret. One of the prisoners considers asking a friendly guard to tell him who is the prisoner other than himself that will be released, but hesitates based on the rationale:

- at the prisoner's present state of knowledge, the probability of being released is $2/3$. After he knows the answer, the probability of being released will become $1/2$, since there will be two prisoners (including himself) whose fate is unknown and one of the two will be released.

What is wrong with this line of reasoning?

Exercise 01.05 ✓. The fraction of persons in a population who have a certain disease is 0.01. A diagnostic test is available to test for the disease. But for a healthy person the chance of being falsely diagnosed as having the disease is 0.05, while for someone with the disease the chance of being falsely diagnosed as healthy is 0.2.

Supposing that the test is performed on a person selected at random from the population, answer the following questions.

1) What is the probability that the test shows a positive result (meaning that the person is diagnosed as ill, either correctly or wrongly)?

2) What is the probability that the person selected at random is ill but he/she is diagnosed healthy?

3) What is the probability that diagnose is correct and he/she is healthy?

4) Suppose the test shows a positive result. What is the probability that the person actually has the disease?

5) Do the above probabilities admit a long-run frequency interpretation? Explain.

Exercise 01.06 ✓(1.5 PRML). Using the definition of the variance of variance $f(x)$, $\mathrm{var}[f(x)] = \mathbb{E}[(f(x) - \mathbb{E}[f(x)])^2]$, show that $\mathrm{var}[f(x)]$ satisfies $\mathrm{var}[f(x)] = \mathbb{E}[f(x)^2] - \mathbb{E}[f(x)]^2$.

Exercise 01.07 ✓(1.6 PRML). Show that if two variables x and y are independent, then their covariance is zero.

Exercise 01.08 ✓(1.11 PRML). By setting the derivative of the log likelihood function

$$\ln p(\mathtt{x}|\mu, \sigma^2) = -\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n - \mu)^2 - \frac{N}{2}\ln(\sigma^2) - \frac{N}{2}\ln(2\pi) \tag{1}$$

with respect to μ and σ^2 equal to zero, verify the following results

a)

$$\mu_{ML} = \frac{1}{N}\sum_{n=1}^{N} x_n; \tag{2}$$

b)

$$\sigma^2_{ML} = \frac{1}{N}\sum_{n=1}^{N}(x_n - \mu_{ML})^2. \tag{3}$$

Exercise 01.09 ✓(1.32 PRML). Evaluate the Kullback-Liebler divergence

$$\mathrm{KL}[p||q] = -\int p(x)\ln q(x)dx + \int p(x)\ln p(x)dx \tag{4}$$

between two Gaussians $p(x) = \mathcal{N}(x|\mu, \sigma^2)$ and $q(x) = \mathcal{N}(x|m, s^2)$.

Exercise 01.10 ✓(1.39 PRML). Consider two random variables x and y having joint distribution $p(x, y)$

		y = 0	y = 1
x	0	1/3	1/3
	1	0	1/3

a) Evaluate the following quantities:

a) $\mathrm{H}[x]$;
b) $\mathrm{H}[y]$;
c) $\mathrm{H}[y|x]$;
d) $\mathrm{H}[x|y]$;
e) $\mathrm{H}[x, y]$;
f) $\mathrm{I}[x; y]$.

b) Draw a diagram to show the relationship between these various quantities.

CK0146 or TIP8311: Extra 01

You are not requested to solve the following exercises, they are provided only for you to practice. Feel free to share your solutions.

Exercise 01.A. Two balls are placed in a box as follows. A fair coin is tossed and

- a white ball is placed in the box if a head occurs,
- otherwise a red ball is placed in the box

The coin is tossed again and

- a red ball is placed in the box if a tail occurs,
- otherwise a white ball is placed in the box

Balls are drawn from the box three times in succession (always with replacing the drawn ball back in the box), and it is found that on all three occasions a red ball is drawn.

What is the probability that both balls in the box are red?

Hint: Basically, we are asking

$$p(\text{2 red in box}|\text{3 red drawn}) = \frac{p(\text{3 red drawn}|\text{3 red in box})p(\text{2 red in box})}{p(\text{3 red drawn})}. \tag{5}$$

Exercise 01.B. A coin is tossed twice. Someone claims that the event of two heads is at least as likely if we know that the first toss is a head than if we know that at least one of the tosses is a head. Is he right?

Exercise 01.C. We roll two fair 6-sided dice. Each one of the 36 possible outcomes is assumed to be equally likely.

1. Find the probability that doubles are rolled.
2. Given that the roll results in a sum of 4 or less, find the conditional probability that doubles are rolled.
3. Find the probability that at least one die roll is a 6.
4. Given that the two dice land on different numbers, find the conditional probability that at least one die roll is a 6.

Exercise 01.D. Romeo e Juliet have a date at a given time, and each will arrive at meeting place with a delay between 0 and 1 hour, with all pairs of delays being equally likely. The first to arrive will wait for 15 minutes and will leave if the other has not yet arrived.

What is the probability they will meet?

Exercise 01.E (1.13 PRML). Suppose that the variance of a Gaussian is estimated using the result

$$\sigma^2_{ML} = \frac{1}{N}\sum_{n=1}^{N}(x_n - \mu_{ML})^2, \tag{6}$$

but with the maximum likelihood estimate μ_{ML} replaced with the true value μ of the mean.

Show that this estimator has the property that its expectation is given by the true variance σ^2.

Hint: First solve (or check the solution to) Exercise 1.12 of PRML, then advance in a similar way.

Exercise 01.F (1.23 PRML). Derive the criterion for minimising the expected loss when there is a general loss matrix and general prior probabilities for the classes.

Exercise 01.G (1.32 PRML). Consider a vector $\mathbf{x}$ of continuous variables with distribution $p(\mathbf{x})$ and corresponding entropy $\mathrm{H}[\mathbf{x}]$. Suppose we make a non-singular linear transformation of $\mathbf{x}$ to obtain a new variable $\mathbf{y} = \mathbf{A}\mathbf{x}$.

Show that the corresponding entropy is given by $\mathrm{H}[\mathbf{y}] = \mathrm{H}[\mathbf{x}] + \ln|\mathbf{A}|$, where $|\mathbf{A}|$ denotes the determinat of $\mathbf{A}$.

Exercise 01.H (1.33 PRML). Suppose that the conditional entropy $\mathrm{H}[y|x]$ between two discrete random variables x and y is zero.

Show that, for all values of x such that $p(x) > 0$, the variable y must be a function of x. (In other words, show that for each x there is only one value of y such that $p(y|x) \neq 0$.)