Artificial Intelligence 2

Solutions

Main Summer Examinations 2023

Note

Answer ALL questions. Each question will be marked out of 20. The paper will be marked out of 60, which will be rescaled to a mark out of 100.

Question 1

A consultant doctor who specialised in 10 different lung diseases, denoted as θ_1 to θ_{10}, knows these diseases typically display 4 symptoms, denoted as x_1 to x_4. In the past year, the doctor saw 200 patients, and kept a record of joint observations of these patients in Table 1. The number $n(x_r, \theta_C)$ in each cell represents the number of patients with the symptoms x_r and disease θ_c. The doctor want to apply probability and Bayesian theorem to analyse this patient record.

	θ_1	θ_2	θ_3	θ_4	θ_5	Sum
x_1	8	9	9	5	6	37
x_2	3	5	8	9	30	55
x_3	0	1	1	10	59	71
x_4	0	0	1	0	36	37
Sum	11	15	19	24	131	200

Table 1: Joint observations of symptoms x_r, i.e., rows $r = \{1, 2, 3, 4\}$ and diseases θ_c, i.e., columns $c = \{1, 2, \ldots, 5\}$. The number $n(x_r, \theta_C)$ in each cell represents the number of patients with the symptoms x_r and disease θ_c.

(a) Let the discrete random variable X to represent the row number, i.e., one of the four symptoms, of which the range $R_X = \{x_1, x_2, x_3, x_4\}$ and Θ to represent the column number, i.e., one of the 10 diseases, of which the range $R_\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\}$. Derive the Joint Probability Mass Function (PMF) of two discrete random variables X and Θ and their marginal PMFs. The results should be presented in a table (Fractions are acceptable answers). **[8 marks]**

(b) What is the conditional probability $p(x_3|\theta_2)$ that a patient has the symptom x_3 given that he has the disease θ_2? **[2 marks]**

(c) What is the conditional probability $p(\theta_2|x_3)$ that a patient has the disease θ_2 given that this patient has the symptom x_3? **[4 marks]**

(d) What are the conditional probabilities of a patient having each of these diseases given that the patient has the symptom x_3. From your calculation, which disease this patient is more likely to have? **[6 marks]**

Model answer / LOs / Creativity:

(a) The students should first use random variables to model the patient record: let the discrete random variable X to represent the row number, i.e., one of the four symptoms, of which the range $R_X = \{x_1, x_2, x_3, x_4\}$ and Θ to represent the column number, i.e., one of the 10 diseases, of which the range $R_\Theta = \{\theta_1, \theta_2, \ldots, \theta_{10}\}$. We can now calculate the joint probability and obtain the joint PMF of X and Θ as the following table. Fractions are also acceptable.

	θ_1	θ_2	θ_3	θ_4	θ_5	$P(X)$
x_1	$\frac{8}{200}$	$\frac{9}{200}$	$\frac{9}{200}$	$\frac{5}{200}$	$\frac{6}{200}$	$\frac{37}{200}$
x_2	$\frac{3}{200}$	$\frac{5}{200}$	$\frac{8}{200}$	$\frac{9}{200}$	$\frac{30}{200}$	$\frac{55}{200}$
x_3	0	$\frac{1}{200}$	$\frac{1}{200}$	$\frac{10}{200}$	$\frac{59}{200}$	$\frac{71}{200}$
x_4	0	0	$\frac{1}{200}$	0	$\frac{36}{200}$	$\frac{37}{200}$
$P(\Theta)$	$\frac{11}{200}$	$\frac{15}{200}$	$\frac{19}{200}$	$\frac{24}{200}$	$\frac{131}{200}$	1

Table 2: Joint PMF of two discrete random variables X and Θ and their marginal PMFs. The marginal PMFs $P(X)$ and $P(\Theta)$ are also know as **the marginal likelihood** (of symptoms) and **the prior distribution** (of disease), respectively.

(b) The conditional probability $p(x_3|\theta_2)$ that a patient has the symptom x_3 given that he has the disease θ_2 is

$$p(x_3|\theta_2) = \frac{p(x_3, \theta_2)}{p(\theta_2)} = \frac{0.005}{0.075} = \frac{1}{15},$$

(c) The conditional probability $p(\theta_2|x_3)$ that a patient has the disease θ_2 give that he has the symptom x_3 is:

$$p(\theta_2|x_3) = \frac{p(x_3|\theta_2)p(\theta_2)}{p(x_3)} = \frac{\frac{1}{15} \times 0.075}{0.355} = \frac{1}{71}$$

(d)

$$\begin{aligned} p(\Theta|x_3) &= \frac{p(x_3|\Theta)p(\Theta)}{p(x_3)} \\ &= \{0, 1/71, 1/71, 10/71, 59/71\} \end{aligned}$$

Maximum a posteriori probability (MAP) estimate: the value of Θ that correspond to the maximum value of $p(\Theta|x_3)$, i.e.,

$$\hat{\theta}_{\text{MAP}} = \underset{\theta}{\operatorname{argmax}}\, p(\Theta|x_3) = \theta_5$$

Learning outcomes: The creative parts are (a) - (c).

Question 2

Breast cancer is the 2nd most common cause of female cancer deaths in the UK, which causes around 11,500 deaths in women every year. To diagnose breast cancer, University of Wisconsin collected a set of breast tissue images from 569 patients of which 212 were diagnosed as 'Malignant' and 357 as 'Benign'. They further processed the images to generate a dataset with 30 independent variables that describe characteristics of the cell nuclei present in the images. We listed 10 main real-value independent variables in Table 3.

θ_i	Features
θ_1	Radius
θ_2	Texture
θ_3	Perimeter
θ_4	Area
θ_5	Smoothness
θ_6	Compactness
θ_7	Concavity
θ_8	Concave points
θ_9	Symmetry
θ_{10}	Fractal dimension

Table 3: The Wisconsin Breast Cancer Diagnostic dataset main features

You will use the dataset to train a logistic regression model to predict if a new patient is likely to develop malignancy or not.

(a) After data exploratory analysis, you discover some pairs of the 10 independent variables are highly correlated. What problems would these highly-correlated independent variables cause? Describe in your own words why mutual information based feature selection method could solve this problem. **[8 marks]**

(b) You use two independent variables: Radius (θ_1) and Texture (θ_2) to build a logistic regression model to classify patients into being 'Malignant' or 'Benign'. The fitted model is

$$\log\left(\frac{p}{1-p}\right) = -7.69 + 0.035\theta_1 + 0.447\theta_2$$

(i) Based on the fitted logistic regression model, interpret how Radius (θ_1) and Texture (θ_2) affect the risk of malignancy, respectively. **[5 marks]**

(ii) The logistic regression model uses a threshold, i.e., $p = 0.5$ to map this probability prediction to two discrete classes: when $p > 0.5$ it will classify the patient as '1' ('Malignancy') and $p < 0.5$ it will classify the patient as '0' ('Benign').

The patients that have probability of exactly 0.5 is called decision boundary. Given our fitted logistic regression model, calculate its decision boundary. Show your working. **[4 marks]**

(iii) A new patient has the the following measures: Radius $\theta_1 = 10$ and Texture $\theta_2 = 20$. What is the probability that this patient is Malignancy? Show your working. **[3 marks]**

Model answer / LOs / Creativity:

Learning outcomes: The creative part is (b).

(a) The highly correlated features will cause the so-called feature redundancy problem, which will:

- Slow Down the Training Process: The more features you have, the slower the calculations are.
- Reduced generalisation performance: The optimisation problem will have more dimensions because of highly correlated features. More importantly, having correlated features in the training set makes the loss landscape ill-conditioned which is more difficult to obtain optimal coefficients.
- The fitted model is hard to interpret since there are too many correlated features.

The method to solve this problem is to apply mutual information based feature selection method. The mutual information is to measure the information that two random variables share. In other words, it measures how much knowing one of these variables reduces uncertainty about the other. We can then use a forward selection, in which we also find the feature set which achieves the maximum mutual information among all the remaining independent variables in set and also make sure there must be minimal redundancy between the candidate feature and the set of selected features.

(b) Logistic regression interpretation

(i) The Radius coefficient $\theta_1 = 0.035$ means that after holding the other factors fixed i.e., Texture, for every extra unit of Radius, the log odds of a patient's risk of malignancy goes up by 0.035. For the Texture coefficient $\theta_2 = 0.447$ means that all else equal, for every extra unit increase of Texture, the log odds of a patient's risk of malignancy goes up by 0.447.

(ii)

$$p = \frac{1}{2} \Rightarrow \log\left(\frac{p}{1-p}\right) = 0 \Rightarrow 0.035\theta_1 + 0.447\theta_2 = 7.69 \Rightarrow \theta_2 = 17.20 - 0.078\theta_1$$

, which means the decision boundary is a straight line with the interception of 17.2 and the slope of 0.078.

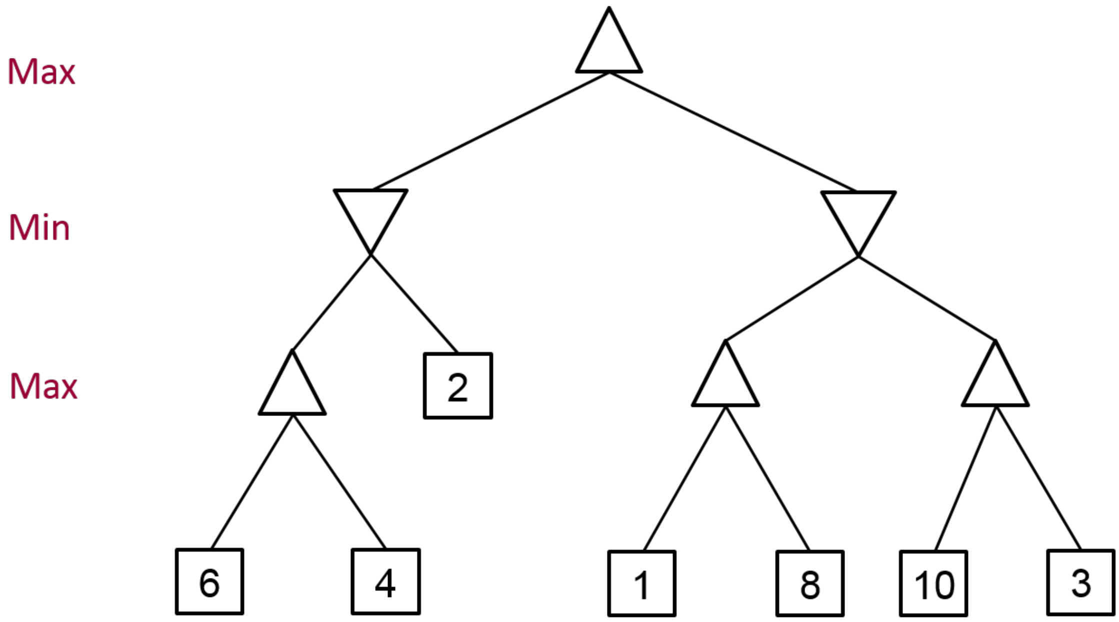
(iii)

$$\log\left(\frac{p}{1-p}\right) = -7.69 + 0.035 \times 10 + 0.447 \times 20 = 1.6$$

$$p = \frac{1}{1+e^{-1.6}} = 0.83$$

Question 3 Games

(a) Consider the following minimax game tree. There are two players Max and Min; the player Max wants to maximise the utility and the player Min wants to minimise the utility. The tree has four layers and we can use Lm-n to denote the nth node from left to right in the layer m; for example, the root node can be denoted by L1-1, and the first node at the bottom layer (with value 6) can be denoted by L4-1. Give the minimax value of the root node (L1-1). **[4 marks]**



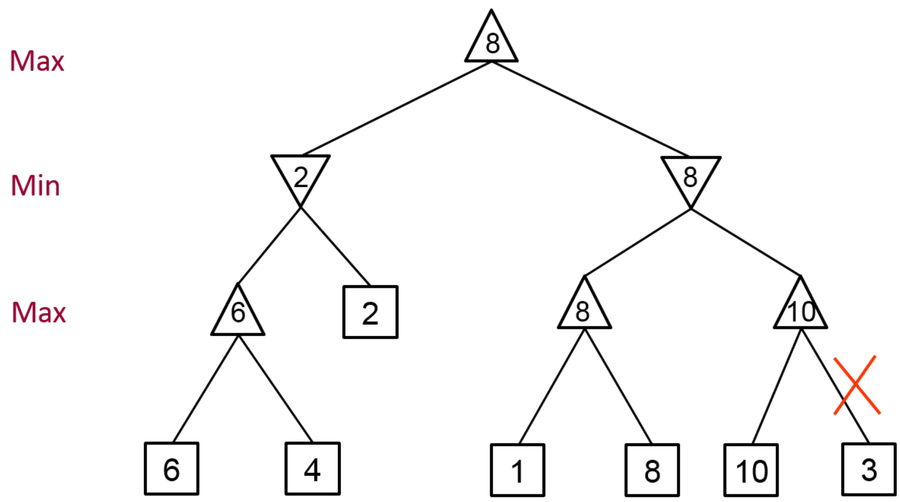
(b) We use the alpha-beta pruning algorithm to prune the above tree. List all the pruned nodes. Assume that child nodes are visited from left to right. **[8 marks]**

(c) Now suppose that when doing the tree search, child nodes are not necessarily visited from left to right. In other words, for any node you can explore its children in any order (e.g., the left child first for one node and the right child first for another node). Then in the best-case scenario, how many nodes in the above tree can be pruned by alpha-beta pruning algorithm? What are they? **[8 marks]**

Model answer / LOs / Creativity:

The creative parts are (c).

(a) The value of each node of the tree is shown in the following figure. So the minimax value of the root node is 8 (i.e., v(L1-1) = 8).



(b) The pruned node is L4-6, as shown in the figure.

(c) To prune the most nodes in the tree, the order of the nodes to be visited is L1-1 => L2-2 => L3-3 => L4-3 => L4-4 => L3-4 => L4-5 => L2-1 => L3-2. There are four nodes to be pruned. They are L4-6, L3-1, L4-1, L4-2, as shown in the following figure.