COMP7015 Artificial Intelligence	**Semester 1, 2022/23**

Sample Solutions to In-class Quiz

Instructor: Dr. Kejing Yin — *updated on Nov. 20, 2022*

Remarks: There are three versions of the in-class quiz. They differ only in symbols, order of sub-questions, and exact numbers in data.

Bug fixed on Nov. 20, 2022:

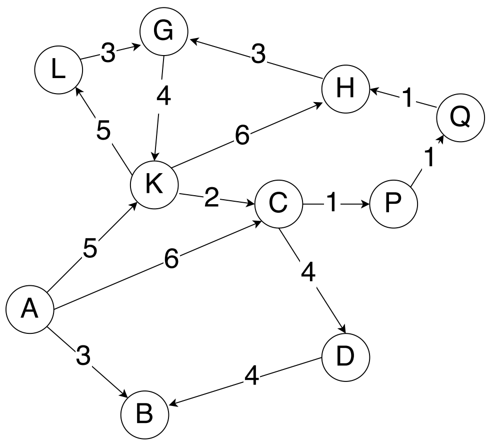
- Logistic loss are corrected.

Bug fixed on Oct. 27, 2022:

- Typo in answers to Q 3.b.1 is fixed.
- Answer to Q 3.c.3 is corrected.

Question 1 (20 marks)

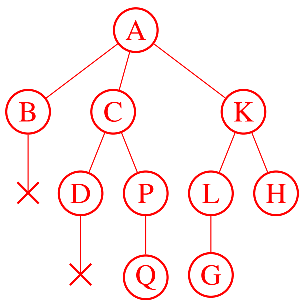
The following figure shows the flight connections between different cities. Each letter inside the circle represents a city, the arrow between two cities means that there is a flight between them and the direction of the arrow indicates the direction of the flight. The numbers above each arrow show the price of the flight tickets. Suppose you are running a travel agency and would like to use AI algorithms to find routes for your customers.



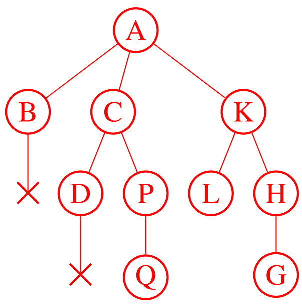
(a) A customer does not care about the price and would like to have the minimum number of flight transfer. What algorithm should be used? Plot the search tree constructed by using this algorithm and write down the order of nodes that you expanded. *(10 marks)*

(1) We use breadth-first search (BFS) as it finds the shallowest solution. In this problem, the flight with minimum transit corresponds to the shallowest solution. *(2 marks)*

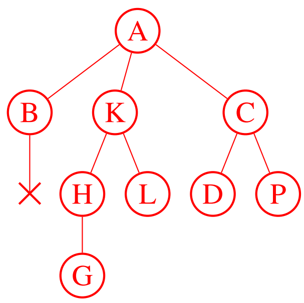
(2) Some examples of correct answers are: *(5 marks)*



or



or

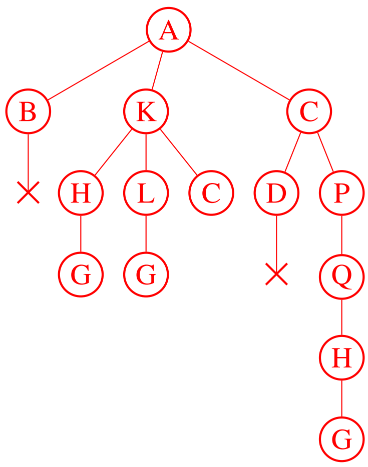


(3) Examples of correct answers (The answer must align with the search tree plotted): *(3 marks)*

- A, B, C, K, D, P, L (G) *(check goal state when generating the node)*
- A, B, C, K, D, P, L, H, Q, (G) *(check goal state when expanding the node)*

(b) Your customer would like to find the cheapest flight from A to G, no matter how many transits it needs. What algorithm should be used? Plot the search tree constructed by this algorithm and write down the order of nodes that you expanded. *(10 marks)*

(1) We use uniform-cost search as it fins the cost-optimal solution. In this problem, the cheapest flight corresponds to the cost-optimal solution when we consider the price as the cost. *(2 marks)*
(2) The search tree constructed is: *(5 marks)*



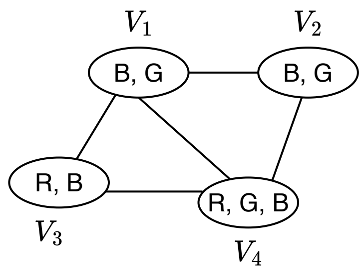
(3) The order of expanding nodes is: *(5 marks)*

- A, B, K, C, P, Q, H, D, L, H, G; *or*
- A, B, K, C, P, Q, H, L, D, H, G

Question 2 (10 marks)

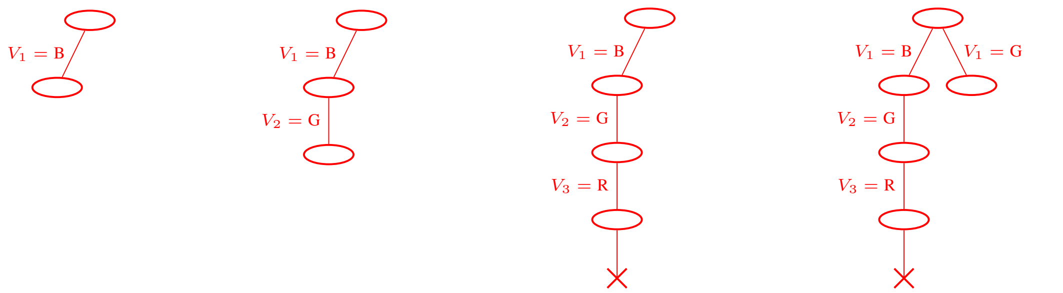
A constraint satisfaction problem is stated as follows:

- Variables: V_1, V_2, V_3, and V_4;
- Domains: $D_1 = \{\mathrm{B}, \mathrm{G}\}$; $D_2 = \{\mathrm{B}, \mathrm{G}\}$; $D_3 = \{\mathrm{R}, \mathrm{B}\}$; $D_4 = \{\mathrm{R}, \mathrm{G}, \mathrm{B}\}$
- Constraints: adjacent variables (variables connected with edges) cannot take the same value.



Show the results of using <u>backtrack with forward checking</u> to find assignments to all variables. Show the steps of your solution.

The solution is: $V_1 = \mathrm{G}$, $V_2 = \mathrm{B}$, $V_3 = \mathrm{B}$, and $V_4 = \mathrm{R}$. *(4 marks)*
The steps are shown as below. *(6 marks)*



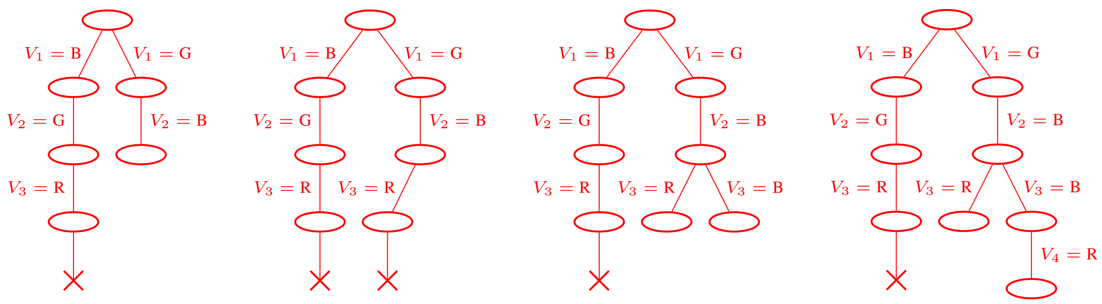
$V_1 \in \{\mathrm{B}, \cancel{\mathrm{G}}\}$, $V_2 \in \{\cancel{\mathrm{B}}, \mathrm{G}\}$, $V_3 \in \{\mathrm{R}, \cancel{\mathrm{B}}\}$, $V_4 \in \{\mathrm{R}, \cancel{\mathrm{B}}, \mathrm{G}\}$

$V_1 \in \{\mathrm{B}, \cancel{\mathrm{G}}\}$, $V_2 \in \{\cancel{\mathrm{B}}, \mathrm{G}\}$, $V_3 \in \{\mathrm{R}, \cancel{\mathrm{B}}\}$, $V_4 \in \{\mathrm{R}, \cancel{\mathrm{B}}, \cancel{\mathrm{G}}\}$

$V_1 \in \{\mathrm{B}, \cancel{\mathrm{G}}\}$, $V_2 \in \{\cancel{\mathrm{B}}, \mathrm{G}\}$, $V_3 \in \{\mathrm{R}, \cancel{\mathrm{B}}\}$, $V_4 \in \{\cancel{\mathrm{R}}, \cancel{\mathrm{B}}, \cancel{\mathrm{G}}\}$

$V_4 = \emptyset$, backtrack

$V_1 \in \{\cancel{\mathrm{B}}, \mathrm{G}\}$, $V_2 \in \{\mathrm{B}, \cancel{\mathrm{G}}\}$, $V_3 \in \{\mathrm{R}, \mathrm{B}\}$, $V_4 \in \{\mathrm{R}, \mathrm{B}, \cancel{\mathrm{G}}\}$



$V_1 \in \{\cancel{\mathrm{B}}, \mathrm{G}\}$, $V_2 \in \{\mathrm{B}, \cancel{\mathrm{G}}\}$, $V_3 \in \{\mathrm{R}, \mathrm{B}\}$, $V_4 \in \{\mathrm{R}, \cancel{\mathrm{B}}, \cancel{\mathrm{G}}\}$

$V_4 = \emptyset$

$V_1 \in \{\cancel{\mathrm{B}}, \mathrm{G}\}$, $V_2 \in \{\mathrm{B}, \cancel{\mathrm{G}}\}$, $V_3 \in \{\mathrm{R}, \cancel{\mathrm{B}}\}$, $V_4 \in \{\cancel{\mathrm{R}}, \cancel{\mathrm{B}}, \cancel{\mathrm{G}}\}$

$V_1 \in \{\cancel{\mathrm{B}}, \mathrm{G}\}$, $V_2 \in \{\mathrm{B}, \cancel{\mathrm{G}}\}$, $V_3 \in \{\cancel{\mathrm{R}}, \mathrm{B}\}$, $V_4 \in \{\mathrm{R}, \cancel{\mathrm{B}}, \cancel{\mathrm{G}}\}$

$V_1 \in \{\cancel{\mathrm{B}}, \mathrm{G}\}$, $V_2 \in \{\mathrm{B}, \cancel{\mathrm{G}}\}$, $V_3 \in \{\cancel{\mathrm{R}}, \mathrm{B}\}$, $V_4 \in \{\mathrm{R}, \cancel{\mathrm{B}}, \cancel{\mathrm{G}}\}$

solution found

Question 3 (25 marks)

(a) Given P is true and Q is false, determine the truth value for the following propositions. *(8 marks)*

1. $\neg Q \wedge P$ true *(1 mark)*
2. $Q \vee \neg\neg P$ true *(1 mark)*
3. $(P \wedge Q) \vee Q \Rightarrow \neg P \vee Q$ true *(2 mark)*
4. $\neg P \vee (P \wedge Q) \Leftrightarrow P \vee \neg Q$ false *(2 mark)*
5. $\neg(\neg Q \Rightarrow \neg P) \wedge (P \Leftrightarrow Q)$ false *(2 mark)*

(b) Are the following correct? Justify your answer. *(8 marks)*

1. $(A \vee B) \wedge \neg(A \Rightarrow B)$ is satisfiable.

   The statement is correct.

   For variables A and B, there are four possible models: $m_1 = \{A = 1, B = 1\}$, $m_2 = \{A = 1, B = 0\}$, $m_3 = \{A = 0, B = 1\}$, and $m_4 = \{A = 0, B = 0\}$, where we use 1 to represent true and 0 to represent false. We can examine the four models one by one.

   i) $A = 1$ and $B = 1$, $(A \vee B) \wedge \neg(A \Rightarrow B)$ is false. Thus, m_1 does not satisfy this formula.
   ii) $A = 1$ and $B = 0$, $(A \vee B) \wedge \neg(A \Rightarrow B)$ is true. Thus, m_2 satisfies this formula.
   iii) $A = 0$ and $B = 1$, $(A \vee B) \wedge \neg(A \Rightarrow B)$ is false. Thus, m_3 does not satisfy this formula.
   iv) $A = 0$ and $B = 0$, $(A \vee B) \wedge \neg(A \Rightarrow B)$ is false. Thus, m_4 does not satisfy this formula.

   The set of all models of the given formula is $M\{(A \vee B) \wedge \neg(A \Rightarrow B)\} \neq \emptyset$. Therefore, the formula is satisfiable.

2. $(A \vee B) \vDash (A \Leftrightarrow B)$.

   The statement is not correct.

   For variables A and B, there are four possible models: $m_1 = \{A = 1, B = 1\}$, $m_2 = \{A = 1, B = 0\}$, $m_3 = \{A = 0, B = 1\}$, and $m_4 = \{A = 0, B = 0\}$, where we use 1 to represent true and 0 to represent false. Similar to the steps above, we can obtain that:

   i) $M(A \vee B) = \{m_1, m_2, m_3\}$
   ii) $M(A \Leftrightarrow B) = \{m_1, m_4\}$

   Obviously, $M(A \vee B) \not\subseteq M(A \Leftrightarrow B)$. Therefore, $(A \vee B)$ does not entail $(A \Leftrightarrow B)$.

(c) Write a first-order logic formula for each of the following English sentence. *(9 marks)*

1. There exists a student that can play Tennis.

   $\exists x \;\; \text{IsStudent}(x) \wedge \text{CanPlayTennis}(x)$ *(3 marks)*

2. All students need to take a course that covers AI.

   $\forall x \;\; \text{IsStudent}(x) \Rightarrow (\exists y \;\; \text{IsCourse}(y) \wedge \text{TakeCourse}(x, y) \wedge \text{CoversAI}(y))$ *(3 marks)*

3. There are some students who take all courses that cover AI.

   $\exists x \;\; \text{IsStudent}(x) \wedge (\forall y \;\; \text{IsCourse}(y) \wedge \text{CoversAI}(y) \Rightarrow \text{TakeCourse}(x, y))$ *(3 marks)*

Question 4 (20 marks)

Given the following dataset D with three features about watermelon, *i.e.*, their surface, texture, and density. We would like to build a decision tree model using the three features to predict if the watermelon is ripe.

Features			Label
surface	texture	density	ripe?
hard	clear	0.77	yes
hard	clear	0.56	yes
hard	slightly blurry	0.64	no
hard	clear	0.61	yes
hard	clear	0.63	yes
hard	slightly blurry	0.66	no
hard	slightly blurry	0.67	no
hard	clear	0.44	yes
soft	clear	0.24	no
soft	clear	0.40	yes
hard	blurry	0.25	no
soft	blurry	0.34	no
soft	slightly blurry	0.48	yes

(a) Compute the entropy of the dataset $\mathrm{Ent}(D)$. Show the steps. *(3 marks)*

$$\begin{aligned}\mathrm{Ent}(D) &= -\sum_{k=1}^{K} p_k \log_2 p_k \\ &= -\frac{7}{13}\log_2\frac{7}{13} - \frac{6}{13}\log_2\frac{6}{13} \\ &= -\frac{7}{13}(2.81-3.7) - \frac{6}{13}(2.58-3.7) \\ &= 0.996\end{aligned}$$

(b) Compute the Information Gain Ratio of the feature "texture". Show the steps. *(5 marks)*

The information gain is *(2 marks)*

$$\begin{aligned}\mathrm{Gain}(D, \text{“texture”}) &= \mathrm{Ent}(D) - \sum_{v=1}^{V}\frac{|D^v|}{|D|}\mathrm{Ent}(D^v) \\ &= 0.996 - \frac{7}{13}\mathrm{Ent}(D^{\text{clear}}) - \frac{4}{13}\mathrm{Ent}(D^{\text{slightly blurry}}) - \frac{2}{13}\mathrm{Ent}(D^{\text{blurry}}) \\ &= 0.996 - \frac{7}{13}\times 0.60 - \frac{4}{13}\times 0.82 - \frac{2}{13}\times 0 \\ &= 0.427\end{aligned}$$

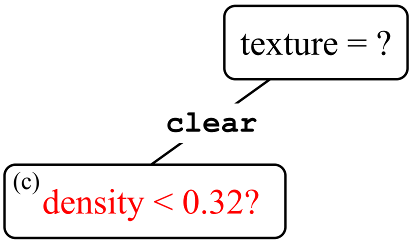
The intrinsic value is *(2 marks)*

$$\mathrm{IV}(\text{“texture”}) = -\sum_{v=1}^{V}\frac{|D^v|}{|D|}\log_2\frac{|D^v|}{|D|} = -\frac{7}{13}\log_2\frac{7}{13} - \frac{4}{13}\log_2\frac{4}{13} - \frac{2}{13}\log_2\frac{2}{13} = 1.42$$

Therefore, the information gain ratio can be computed by *(1 mark)*

$$\mathrm{Gain\_ratio}(D, \text{“texture”}) = \frac{\mathrm{Gain}(D, , \text{“texture”})}{\mathrm{IV}(\text{“texture”})} = \frac{0.47}{1.42} = 0.33$$

(c) Use the Information Gain as the criterion for selecting the splitting features. The first splitting feature is "texture", as shown in the figure below. What feature should be the next splitting feature if "texture=clear"? Fill in the box in the figure and show the steps to justify your answer. *(12 marks)*



The splitting feature to be filled in is "density $<$ 0.32?". *(2 marks)*

The detailed steps are shown as follows. First we extract out the data points with "texture"="clear", which gives the following dataset. We denote it by D in this sub question. There are two candidate features: "surface" and "density".

Features			Label
surface	texture	density	ripe?
hard	clear	0.77	yes
hard	clear	0.56	yes
hard	clear	0.61	yes
hard	clear	0.63	yes
hard	clear	0.44	yes
soft	clear	0.24	no
soft	clear	0.40	yes

(1) The entropy of this dataset is: *(2 marks)*

$$\mathrm{Ent}(D) = -\sum_{k=1}^{K} p_k \log_2 p_k = -\frac{6}{7}\log_2\frac{6}{7} - \frac{1}{7}\log_2\frac{1}{7} = 0.60$$

(2) The information gain of the feature "surface" is: *(2 marks)*

$$\begin{aligned}
\mathrm{Gain}(D, \text{"surface"}) =& \mathrm{Ent}(D) - \sum_{v=1}^{V} \frac{|D^v|}{|D|}\mathrm{Ent}(D^v) \\
=& 0.60 - \frac{5}{7}\mathrm{Ent}(D^{\text{hard}}) - \frac{2}{7}\mathrm{Ent}(D^{\text{soft}}) \\
=& 0.60 - \frac{5}{7}\times 0 - \frac{2}{7}\times 1 \\
=& 0.31
\end{aligned}$$

(3) To compute the information gain of the feature "density" using bi-partitioning, we first list out the candidates of the partition thresholds: $T = [0.32, 0.42, 0.5, 0.585, 0.62, 0.7]$. *(2 marks)*

(4) We can compute the information gain of bi-partitioning D by a threshold $t \in T$ as follows:

$$\mathrm{Gain}(D, a, t) = \mathrm{Ent}(D) - \sum_{\lambda\in\{+,-\}} \frac{|D_t^\lambda|}{|D|}\mathrm{Ent}(D_t^\lambda)$$

$$\mathrm{Gain}(D, \text{"density"}, 0.32) = 0.59 \quad \mathrm{Gain}(D, \text{"density"}, 0.42) = 0.31$$
$$\mathrm{Gain}(D, \text{"density"}, 0.5) = 0.20 \quad \mathrm{Gain}(D, \text{"density"}, 0.585) = 0.13$$
$$\mathrm{Gain}(D, \text{"density"}, 0.62) = 0.08 \quad \mathrm{Gain}(D, \text{"density"}, 0.7) = 0.03$$

Therefore, $\mathrm{Gain}(D, \text{"density"}) = \max_{t\in T} \mathrm{Gain}(D, \text{"density"}, t) = 0.59$ and the threshold is $t = 0.32$. *(2 marks)*

(5) We split the dataset by the feature with the largest information gain. Therefore, we split by "density $<$ 0.32?". *(2 marks)*

Question 5 (25 marks)

Given the following dataset:

Features x_1	x_2	x_3	x_4	Label
6.3	2.5	4.9	1.5	+1
5.4	3.4	1.7	0.2	-1
4.6	3.6	1.0	0.2	+1
6.5	2.8	4.6	1.5	+1
6.0	2.7	5.1	1.6	-1
5.6	3.0	2.6	1.5	+1
5.4	3.7	1.5	0.2	+1
5.0	3.6	1.4	0.2	-1

Consider a linear model with the weight vector and bias given by:

$$\mathbf{w} = [0.4, \; -0.7, \; 2.0, \; 0.8]^\top \qquad b = -6$$

(a) We first compute the scores for each data point. Fill in the missing cells in the "Scores" column using the table below. Then, compute the margin, the one-zero loss, hinge loss, and the logistic loss for each data point and fill them in the table below. *(10 marks)*

Features x_1	x_2	x_3	x_4	Label	Score	Margin	One-zero loss	Hinge loss	Logistic loss
6.3	2.5	4.9	1.5	+1	5.77	5.77	0	0	0
5.4	3.4	1.7	0.2	-1	-2.66	2.66	0	0	0.03
4.6	3.6	1.0	0.2	+1	-4.52	-4.52	1	5.52	1.97
6.5	2.8	4.6	1.5	+1	5.04	5.04	0	0	0
6.0	2.7	5.1	1.6	-1	5.99	-5.99	1	6.99	2.6
5.6	3.0	2.6	1.5	+1	0.54	0.54	0	0.46	0.2
5.4	3.7	1.5	0.2	+1	-3.27	-3.27	1	4.27	1.44
5.0	3.6	1.4	0.2	-1	-3.56	3.56	0	0	0.01

(b) Can we use zero-one loss function under the loss minimization framework with gradient descent algorithm? Briefly explain the reasons. *(2 marks)*

No, since the gradient of zero-one loss function is zero almost everywhere. Therefore, the parameters cannot be updated using gradient descent with zero-one loss function.

(c) Briefly discuss the merits of the hinge loss and logistic loss. *(3 marks)*

(1) They both have non-zero gradient when the model make wrong predictions, so the gradient descent algorithm can be used.
(2) The hinge loss will try to increase the margin even when a sample is correctly classified. It tries to make the model classify all points correctly and confidently.
(3) The logistic loss always have non-zero loss and tries to increase margin even when it already exceeds 1.

(d) Suppose we use $\hat{y} = \text{sign}(\mathbf{w}^\top \mathbf{x} + b)$ as predicted labels. Plot the confusion matrix, compute the accuracy, precision, recall, and F1 scores. *(10 marks)*

(1) The confusion matrix is as follows: *(2 marks)*

	Predicted Positive	Predicted Negative
Ground-truth Positive	TP=3	FN=2
Ground-truth Negative	FP=1	TN=2

(2) The accuracy is given by: *(2 marks)*

$$\text{Accuracy} = \frac{TP + TN}{N} = \frac{5}{8} = 0.625$$

(3) The precision is given by: *(2 marks)*

$$\text{Precision} = \frac{TP}{TP + FP} = \frac{3}{4} = 0.75$$

(4) The recall is given by: *(2 marks)*

$$\text{Recall} = \frac{TP}{TP + FN} = \frac{3}{5} = 0.6$$

(5) The F1 score is given by: *(2 marks)*

$$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2 \times 0.75 \times 0.6}{0.75 + 0.6} = 0.625$$