Student ID: .....

TE WHARE WĀNANGA O TE ŪPOKO O TE IKA A MĀUI

VUW VICTORIA UNIVERSITY OF WELLINGTON

EXAMINATIONS – 2019

TRIMESTER ONE

COMP 307

********** **WITH**

SOLUTIONS **********

INTRODUCTION TO

ARTIFICIAL INTELLIGENCE

Time Allowed: TWO HOURS

CLOSED BOOK

Permitted materials: Only silent non-programmable calculators or silent programmable calculators with their memories cleared are permitted in this examination.
Non-electronic foreign language translation dictionaries may be used.

Instructions: There are a total of 120 marks on this exam.
Attempt all questions.
The appendix on the last sheet can be torn off for reference for questions 2-6.

Questions

Question 1. Search [15 marks]

(a) [2 marks] Give four real-world applications that require search to be solved.

TSP, VRP, timetabling, job shop scheduling, robotics

(b) [3 marks] In the lectures, we have discussed *uniform cost search*, *greedy best-first search* and *A* search*. Briefly describe how each of them selects the next node to expand.

uniform cost: node with the minimal cost, or g(n); greedy best-first search, h(n); A* search: g(n)+h(n)

(c) [5 marks] *Depth-first search*, *breadth-first search* and *iterative deepening (depth-first) search* are three commonly used ununiformed search methods.

(i) Briefly state the main idea of iterative deepening (depth-first) search.

(ii) For each search, state a common data structure it uses in the general tree-search algorithm.

(iii) John Smith said that the sequence of nodes obtained by iterative deepening (depth-first) search is the same as the sequence obtained by breadth-first search. Do you agree or not? Justify your answer.

(i) it is a state space/graph search strategy in which a depth-limited version of depth-first search is run repeatedly with increasing depth limits until the goal is found. (ii) stack, queue, stack. (iii) same as breadth-first search/can also be different if not skipping the previously visited nodes.

(d) [4 marks] During the lectures, we discussed *hill climbing*, *gradient descent* and *genetic beam search*. For each of the following problems, state the most proper method to use. Briefly justify your answers.

(i) Find a local optimum of $f(x_1, x_2, x_3)$, where x_i $(i = 1, 2, 3)$ is a continuous number between 0 and 1. $f(x_1, x_2, x_3)$ is a continuous and differential function.

(ii) Find the global optimum of $f(x_1, x_2, x_3)$, where x_i $(i = 1, 2, 3)$ is a continuous number between 0 and 1. $f(x_1, x_2, x_3)$ is a continuous and differential function.

(iii) Find a local optimum of $f(x_1, x_2, x_3)$, where x_i $(i = 1, 2, 3)$ is an integer between -100 and 100.

(iv) Find the global optimum of $f(x_1, x_2, x_3)$, where x_i $(i = 1, 2, 3)$ is an integer between -100 and 100.

(i) genetic beam search. (ii) hill climbing. (iii) genetic beam search. (iv) genetic beam search.

(e) [1 mark] Both *hill climbing* and *simulated annealing* keeps one state (node) during the search. State the difference between them in terms of moving to the next state.

simulated annealing has chance to move to worse states.

Student ID: .....

SPARE PAGE FOR EXTRA ANSWERS

Cross out rough working that you do not want marked.
Specify the question number for work that you do want marked.

Question 2. Machine Learning Basics [20 marks]

(a) [3 marks] For each of the following machine learning tasks,

(i) state whether it is *supervised* learning or *unsupervised* learning,

(ii) give a real-world application for it.

(1) Classification: supervised

(2) Clustering: unsupervised

(3) Regression: supervised

(b) [1 mark] Briefly describe the difference between classification and regression in terms of the output of the learned model.

classification: class label, regression: real-number

(c) [4 marks] There are several different paradigms in machine learning. Briefly describe the mechanism of each of the following paradigms:

(i) Induction learning: induce a rule

(ii) Case-based learning: matching the training cases

(iii) Genetic/evolutionary learning: based on genetic selection

(iv) Connectionist learning: based on human brain behaviour

(d) [3 marks] In supervised machine learning systems, a dataset is typically separated into a *training set*, a *test set*, and a *validation* set. Briefly describe the role of each of them.

training set is used to train the model. Test set is used to measure the performance of the learned model. Validation set is to monitor the training process and control overfitting.

(e) [2 marks] Briefly describe a reason for doing k-fold cross validation.

dataset is small, the performance is sensitive to the data separation.

(f) [2 marks] The performance of the k nearest neighbour classification method is sensitive to the k value. Briefly describe *two* methods to address this issue.

(1) enumerate a range of k, and select the best one. (2) use the weighted voting depending on the distance of all the neighbours

(g) [5 marks] A company is shortlisting a number of candidates for interview based on their *degree*, *work experience* and whether *a visa is needed*. The following dataset contains 8 candidates, of which four are shortlisted, and the other four are not.

Candidate	Degree	Work Experience	Visa Needed	Shortlisted
1	PhD	High	No	Yes
2	Master	High	No	Yes
3	Master	Low	No	Yes
4	Master	High	Yes	Yes
5	Master	Low	Yes	No
6	PhD	Low	Yes	No
7	Bachelor	High	Yes	No
8	Bachelor	Low	Yes	No

The company wants to build a decision tree to help make decisions using the weighted average impurity measure. The impurity function is defined as $P(Yes) * P(No)$ to select attributes.

(i) Calculate the weighted impurity of the three features at the *root* of the decision tree.

(ii) Which attribute should be chosen for the *root*?

Degree: 2/8 * (1/2 * 1/2) + 4/8 * (3/4 * 1/4) + 2/8 * (0/2 * 2/2) = 0.15625
WE: 4/8 * (3/4 * 1/4) + 4/8 * (1/4 * 3/4) = 0.1875
VN: 3/8 * (0/3 * 3/3) + 5/8 * (1/5 * 4/5) = 12/50 = 0.1;
Visa Needed has the lowest score, and it's the root.

Student ID: .....

SPARE PAGE FOR EXTRA ANSWERS

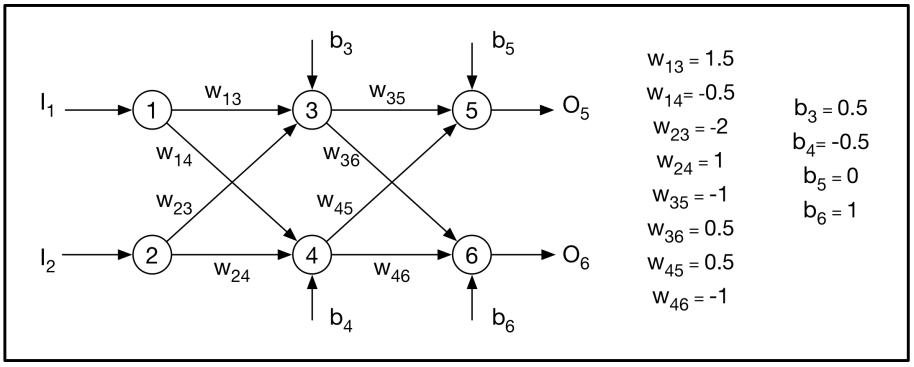
Cross out rough working that you do not want marked.
Specify the question number for work that you do want marked.

SPARE PAGE FOR EXTRA ANSWERS

Cross out rough working that you do not want marked.
Specify the question number for work that you do want marked.

Question 3. Neural Networks [10 marks]

(a) [5 marks] Consider the following feed forward neural network that uses the sigmoid/logistic transfer function (see Appendix B),



(i) What will be the outputs of node 5 (O_5) and node 6 (O_6) for the input vector $(1.0, 1.0)$?

(ii) Assume that the training set consists of only a single instance with an input vector $(1.0, 1.0)$ representing input feature values for I_1 and I_2, and an output vector $(1.0, 0.0)$ representing the target output values for node O_5 and O_6, and that the learning rate η is 0.2. What will be the new value of the weight w_{13} after one epoch of training using the back propagation algorithm?

Show your working.

$O_1 = I_1 = 1; O_2 = I_2 = 1$
$I_3 = 1.5 - 2 + 0.5 = 0; O_3 = f(0) = 0.5;$
$I_4 = -0.5 + 1 - 0.5 = 0; O_4 = 0.5$
$I_5 = 0.5 * (-1) + 0.5 * 0.5 + 0 = -0.25; O_5 = f(-0.25) = 0.44$
$I_6 = 0.5 * 0.5 + 0.5 * (-1) + 1 = 0.75; O_6 = f(0.75) = 0.68$

$\beta_5 = 1 - 0.44 = 0.56$
$\beta_6 = 0 - 0.68 = -0.68$
$\beta_3 = w_{35}O_5(1 - O_5)\beta_5 + w_{36}O_6(1 - O_6)\beta_6 = -1 * 0.44 * 0.56 * 0.56 + 0.5 * 0.68 * 0.32 * (-0.68) = -0.21$
$\Delta w_{13} = \eta O_1 O_3 (1 - O_3)\beta_3 = 0.2 * 1 * 0.5 * 0.5 * (-0.21) = -0.01$
$(w_{13})_{new} = 1.5 - 0.01 = 1.49$

(b) [1 mark] Briefly describe the reason why perceptron cannot solve the XOR problem.

Perceptron can only solve data that is linearly separable, but XOR is not linearly separable.

(c) [4 marks] Jack wants to develop a feedforward neural network to classify the Iris dataset. The dataset has 60 instances, each with 4 features. There are 3 classes. Among the 60 instances, there are 20 instances belonging to each class.

He splits the data into a training set with 15 instances (5 instances per class), and a test set with 45 instances (15 instances per class).

(i) How many nodes should the input layer and output layer have? Justify your choices.

(ii) Jack designed a neural network with a single hidden layer, and 50 hidden nodes. The trained neural network performs very well on the training set, but poorly on the test set. Suggest two ways for improving the test performance of Jack's neural network.

(i) 4 input nodes, 3 output nodes
(ii) re-split data to include more training instances, or reduce the number of hidden nodes

SPARE PAGE FOR EXTRA ANSWERS

Cross out rough working that you do not want marked.
Specify the question number for work that you do want marked.

Question 4. Evolutionary Computation and Learning [15 marks]

(a) [2 marks] Evolutionary computation and learning can be broadly categorised as evolutionary algorithms (EAs), swarm intelligence (SI) and other techniques (Other). Genetic algorithms and genetic programming are in the EAs category. State two techniques in each of the other two categories.

(b) [2 marks] Briefly compare genetic algorithms and genetic programming in terms of the representation of the individuals.

(c) [5 marks] Selection methods, genetic operations, and termination criteria are three important components of the evolutionary process.

(i) State two kinds of commonly used selection methods for selecting parents. Briefly describe their selection process.

(ii) State three criteria that are commonly used for terminating a standard evolutionary process.

(d) [6 marks] The tree based genetic programming approach has been applied to many symbolic regression and classification tasks.

(i) A primitive set in genetic programming consists of a terminal set and a function set. Describe what are terminal set and function set (give an example if necessary), and what are they used for.

(ii) Briefly explain the terms sufficiency and closure in the context of choosing a primitive set in genetic programming.

(iii) State a commonly used fitness function for symbolic regression and classification, respectively.

(iv) Briefly explain what is classification strategy/translation rule in genetic programming for classification. Suggest three possible translation rules for using tree-based GP to perform multi-class classification.

Student ID: .....

Question 5. Reasoning under Uncertainty [10 marks]

Use the probability rules discussed during the lectures to answer questions (a) to (d). Show your working.

Reminder: the $\neg$ symbol stands for "not", the $\perp$ symbol stands for "independence", and the vertical line stands for "given". If no value is given to the variable, e.g. just P(A), you can read it as P(A=true).

[Note: Regarding the calculation, you should show your *working* of the calculation in the form like $P(X=0|Y=1) = \frac{P(X=0,Y=1)}{P(Y=1)}$, to demonstrate that you *know how to calculate* them. It does not matter whether you compute the final numbers here – you can write your final answers in expression form like $\frac{0.55+0.08}{0.89}$.]

(a) [2 marks] Consider two *Boolean* variables A and B, and assume that we know $P(A) = 0.5$ and $P(A,B) = 0.2$. For each of the following probabilities, state whether we can calculate it or not. If yes, calculate its value.

(i) $P(A|B)$

(ii) $P(B|A)$

(iii) $P(\neg B|A)$

(iv) $P(B)$

(i) no; (ii) $P(B|A) = P(A,B)/P(A) = 0.2/0.5 = 0.4$; (iii) 0.6; (iv) no

(b) [3 marks] Consider three *Boolean* variables A, B and C, we have $P(A,B|C) = 0.32$, $P(A,C) = 0.4$, and $P(B,C) = 0.2$, $P(C) = 0.5$. State whether A and B are conditionally independent given C. Justify your answer.

Yes. $P(A,B|C) = P(A|C)P(B|C) = P(A,C)P(B,C)/P^2(C)$

(c) [5 marks] A school surveyed 100 high-school students on which major they would like to choose in university. The following table shows the survey results.

Major	**Male**	**Female**	**Total**
Science	12	9	21
Engineering	30	3	33
Art	8	20	28
Others	10	8	18
Total	60	40	100

(i) Calculate the probability that the student chose engineering as their major.

(ii) Calculate the probability that the student was male and the student chose art as their major.

(iii) Calculate the probability that the student chose art as their major, given that the student was male.

(iv) Calculate the probability that the student is female, given that the student chose science as their major.

(v) Are the events "The student selects others as their major" and "the student is male" independent? Justify your answer.

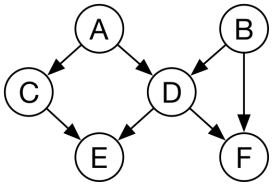
(i) $33/100$
(ii) $8/100$
(iii) $8/60$
(iv) $9/21$
(v) $P(O) = 18/100$, $P(O|M) = 10/60 = 1/6$, not independent

SPARE PAGE FOR EXTRA ANSWERS

Cross out rough working that you do not want marked.
Specify the question number for work that you do want marked.

Question 6. Bayesian Networks [20 marks]

(a) [2 marks] In the following Bayesian network, answer the following true/false questions.



(iii) A and F are conditionally independent given D.

(ii) D and F are conditionally independent given B.

(iii) A and E are conditionally independent given C and D.

(iv) C and D are conditionally independent given A and E.

(i) true
(ii) false
(iii) true
(iv) false

(b) A patient goes to a doctor for a medical condition. The doctor suspects three diseases D1, D2, D3 as the cause of the condition. The three diseases are independent from each other. The doctor wants to check for four symptoms S1, S2, S3, S4 to find the most probable disease. The symptoms are conditionally dependent to the three diseases as follows: S1 depends only on D1, S2 depends on D1 and D2. S3 is depends on D1 and D3, whereas S4 depends only on D3.

(i) [2 marks] Draw the Bayesian network for this problem.

draw

(ii) [2 marks] Write the expression for the joint probability distribution as a product of conditional probabilities.

$$P(D1)P(D2)P(D3)P(S1|D1)P(S2|D1,D2)P(S3|D1,D3)P(S4|D3)$$

(c) [6 marks] The following dataset contains examples of diagnosis for flu.

Chill (C)	**Running nose (R)**	**Headache (H)**	**Fever (F)**	**Flu**
Y	N	Mild	Y	N
Y	Y	Mild	N	Y
Y	N	Strong	Y	Y
N	Y	Mild	Y	Y
N	N	No	N	N
N	Y	Strong	Y	Y
N	Y	Strong	N	N
Y	Y	Mild	Y	Y

Use the Naive Bayes method to classify the following new examples. Show your working.

(Note: need to deal with zero occurrence when necessary).

(i) $C = Y, R = N, H = Mild, F = Y$.

(ii) $C = Y, R = Y, H = No, F = Y$.

(i) flu = Y

$$P(Y \mid C = Y, R = N, H = Mild, F = Y) =$$
$$= \frac{3/5 * 1/5 * 3/5 * 4/5 * 5/8}{P(C = Y, R = N, H = Mild, F = Y)} = 0.036/$$
$$P(N \mid C = Y, R = N, H = Mild, F = Y) =$$
$$= \frac{1/3 * 2/3 * 1/3 * 1/3 * 3/8}{P(C = Y, R = N, H = Mild, F = Y)} = 0.009/$$

(ii) flu = yes.

$$P(Y \mid C = Y, R = Y, H = No, F = Y) =$$
$$= \frac{4/7 * 5/7 * 1/8 * 5/7 * 6/10}{P(C = Y, R = Y, H = No, F = Y)} = 0.022/$$
$$P(N \mid C = Y, R = Y, H = No, F = Y) =$$
$$= \frac{2/5 * 2/5 * 2/6 * 2/5 * 4/10}{P(C = Y, R = Y, H = No, F = Y)} = 0.009/$$

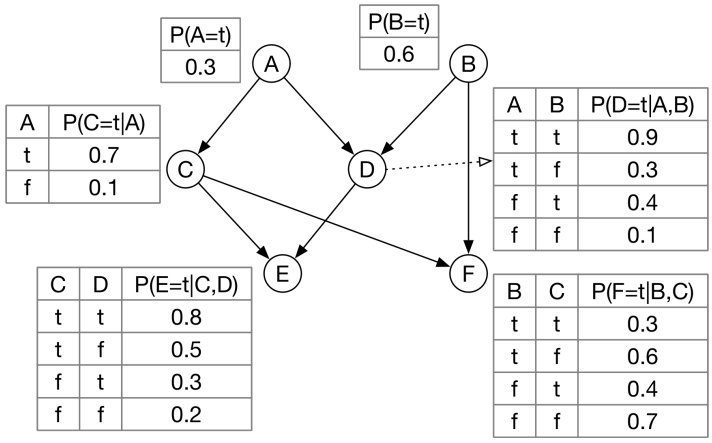
Student ID: .....

SPARE PAGE FOR EXTRA ANSWERS

Cross out rough working that you do not want marked.
Specify the question number for work that you do want marked.

Consider the following Bayesian network along with the conditional probability tables. Each variable takes the value *true (t)* or *false (f)*. Answer questions (d) and (e).

[Note: regarding the calculation, you should show your *working process* of the calculation to demonstrate that you *know how to calculate* them. It does not matter whether you compute the final numbers here – you can write your final answers in expression form like $\frac{0.41+0.17}{0.32}$.]



(d) [2 marks] What is the total number of free parameters in this Bayesian network?

$1+1+2+4+4+4=16$

(e) [6 marks] What is the probability $P(B = t, C = t, E = t, F = f)$?

$$
\begin{aligned}
P(B=t,C=t,E=t,F=f) &= P(t,t,t,t,t,f) + P(f,t,t,t,t,f) + \\
&\quad P(t,t,t,f,t,f) + P(f,t,t,f,t,f) \\
&= P(A)P(B)P(C|A)P(D|A,B)P(E|C,D)P(\neg F|B,C) + \\
&\quad P(\neg A)P(B)P(C|\neg A)P(D|\neg A,B)P(E|C,D)P(\neg F|B,C) + \\
&\quad P(A)P(B)P(C|A)P(\neg D|A,B)P(E|C,\neg D)P(\neg F|B,C) + \\
&\quad P(\neg A)P(B)P(C|\neg A)P(\neg D|\neg A,B)P(E|C,\neg D)P(\neg F|B,C) \\
&= 0.3*0.6*0.7*0.9*0.8*0.7 + 0.7*0.6*0.1*0.4*0.8*0.7 + \\
&\quad 0.3*0.6*0.7*0.1*0.5*0.7 + 0.7*0.6*0.1*0.6*0.5*0.7
\end{aligned}
$$

Question 7. Planning and Scheduling [20 marks]

(a) [1 mark] Briefly describe the definition of *AI planning*.

Find a sequence of actions from the initial state to the goal state.

(b) [2 marks] In backward (regression) relevant state-space search, it is important to identify the *relevant* actions, which could be the last step in the plan leading up to the current goal state. State **two criteria** that an action must satisfy to be relevant.

(1) at least one effect must unify with the elments of the goal.
(2) must not have any effect that negates an element of the goal.

(c) [5 marks] Consider a robot to change a flat tire. The initial state has a flat tire on the axle and a good spare tire in the trunk. The goal is to have a good spare tire properly mounted onto the car's axle. There are four actions: removing the spare from the trunk, removing the flat tire from the axle, putting the spare on the axle, and leaving the car unattended overnight. We assume that the car is in a particularly bad neighborhood, so that the effect of leaving it overnight is that the tires disappear.

Using Planning Domain Definition Language (PDDL), the initial and goal states, and the actions are described as follows. Fill in the blanks in each of the following places.

(i) The goal state ($Goal()$).

(ii) The precondition and effect of $PutOn(Spare, Axle)$.

(iii) The effect of $LeaveOvernight$.

$Init(At(Flat, Axle) \wedge At(Spare, Trunk))$

Goal(At (Spare , Axle))

$Action(Remove(Spare, Trunk),$
PRECOND : $At(Spare, Trunk)$
EFFECT : $\neg At(Spare, Trunk) \wedge At(Spare, Ground))$

$Action(Remove(Flat, Axle),$
PRECOND : $At(Flat, Axle)$
EFFECT : $\neg At(Flat, Axle) \wedge At(Flat, Ground))$

$Action(PutOn(Spare, Axle),$
PRECOND: At(Spare, Ground) $\wedge \neg At(Flat, Axle)$
EFFECT: $\neg At(Spare, Ground) \wedge At(Spare, Axle))$

$Action(LeaveOvernight,$
PRECOND :
EFFECT: $\neg At(Spare, Ground) \wedge \neg At(Spare, Axle) \wedge \neg At(Flat, Ground) \wedge$

(iv) Write the resultant state after applying $Remove(Spare, Trunk)$ to the initial state.

$At(Flat, Axle) \wedge At(Spare, Ground)$

(d) [6 marks] We have discussed the job shop scheduling problem during the lectures. Consider the following *static* job shop scheduling problem, where the operations of each job have to be processed in order. For example, O12 cannot start until O11 has completed.

Job	Operation	Machine	ProcTime
1	O11	M2	20
	O12	M3	50
	O13	M1	30
2	O21	M1	60
	O22	M2	20
	O23	M3	10
3	O31	M3	40
	O32	M1	25
	O33	M2	15

(i) Calculate the schedule generated by applying the Shortest Processing Time dispatching rule. The schedule is given by a sequence of processes. $Process(o_i, m_i, t_i)$ is the i^{th} decision made by the dispatching rule, where o_i is the operation, m_i is the machine, and t is the start time of the processing. You can draw a Gantt Chart to help you.

(ii) What is the makespan of the generated schedule?

(i) $Process(O11, M2, 0)$,
$Process(O31, M3, 0)$,
$Process(O21, M1, 0)$,
$Process(O12, M3, 40)$,
$Process(O32, M1, 60)$,
$Process(O33, M2, 85)$,
$Process(O13, M1, 85)$,
$Process(O22, M2, 115)$,
$Process(O23, M3, 135)$.
(ii) 145.

(e) [6 marks] The matrix below is the distance matrix between the nodes in a graph for a vehicle routing problem.

$$D = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \end{array} \begin{array}{c} \begin{array}{cccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \end{array} \\ \left(\begin{array}{cccccccccc} 0 & 1 & 1.5 & 3 & 2 & 1.5 & 2 & 4 & 3 & 5.5 \\ 1 & 0 & 1 & 3 & 1.5 & 2 & 3 & 2 & 6 & 6.5 \\ 1.5 & 1 & 0 & 2 & 1 & 2 & 4 & 1.5 & 6 & 6 \\ 3 & 3 & 2 & 0 & 2 & 3 & 5 & 1.5 & 7 & 6 \\ 2 & 1.5 & 1 & 2 & 0 & 3 & 4.5 & 2 & 7 & 7 \\ 1.5 & 2 & 2 & 3 & 3 & 0 & 2 & 1.5 & 4 & 4 \\ 2 & 3 & 4 & 5 & 4.5 & 2 & 0 & 4 & 3 & 4 \\ 4 & 2 & 1.5 & 1.5 & 2 & 1.5 & 4 & 0 & 5.5 & 5 \\ 3 & 6 & 6 & 7 & 7 & 4 & 3 & 5.5 & 0 & 2 \\ 5.5 & 6.5 & 6 & 6 & 7 & 4 & 4 & 5 & 2 & 0 \end{array} \right) \end{array}$$

The node 1 is the depot. Each node except the depot has a demand of 1. The capacity is 3. Write the solution generated by the decision-making heuristic. The heuristic generates the routes one by one, and appends one node to the current route at a time. The priority function used by the heuristic is "(distance from here + distance to depot)", which means each time it selects the candidate node with the smallest sum of the distance from the current location plus the distance to the depot.

Write the solution as a set of node sequences starting and ending at the depot node (node 1). It should look like $R_1 = (1, \ldots, 1)$, $R_2 = (1, \ldots, 1), \ldots$.

$R_1 = (1, 2, 3, 5, 1)$,
$R_2 = (1, 6, 7, 9, 1)$,
$R_3 = (1, 4, 8, 10, 1)$

Question 8. Other Topics [10 marks]

At the final stage of our lectures, we provided a very brief discussion of *knowledge based systems, natural language processing, support vector machines, data and web mining, big data*, and *deep learning*.

(a) [1 mark] List two real-world applications of natural language processing.

Machine translation, Speech-to-speech translation, Database querying, Instruction following

(b) [3 marks] List **5Vs** of *big data*.

Volume, variety, veracity, velocity, value

(c) [3 marks] Support Vector Machine (SVM) is an effective classification method.

(i) Briefly describe its difference from *perceptron* in terms of objective function.

(ii) Briefly describe how it deal with non-linearly separable data.

(i) maximise margin. (ii) use kernel function to increase dimensionality.

(d) [3 marks] List three commonly used deep learning methods, and state whether each of them is typically supervised or unsupervised.

CNN, autoencoder, deep belief networks, ...

* * * * * * * * * * * * * * *

Student ID: .....

SPARE PAGE FOR EXTRA ANSWERS

Cross out rough working that you do not want marked.
Specify the question number for work that you do want marked.

SPARE PAGE FOR EXTRA ANSWERS

Cross out rough working that you do not want marked.
Specify the question number for work that you do want marked.

Appendix for COMP307 exam

(You may tear off this page if you wish.)

A Some Formulae You Might Find Useful

$$p(C|D) = \frac{p(D|C)p(C)}{p(D)} \tag{1}$$

$$f(x_i) = \frac{1}{1+e^{-x_i}} \tag{2}$$

$$O_i = f(I_i) = f(\sum_k w_{k\to i} \cdot o_k + b_i) \tag{3}$$

$$\Delta w_{i\to j} = \eta o_i o_j (1-o_j)\beta_j \tag{4}$$

$$\beta_j = \sum_k w_{j\to k} o_k (1-o_k)\beta_k \tag{5}$$

$$\beta_j = d_j - o_j \tag{6}$$

B Sigmoid/Logistic Function