TE WHARE WĀNANGA O TE ŪPOKO O TE IKA A MĀUI
VICTORIA
UNIVERSITY OF WELLINGTON

COMP307

Surname ..... Andersen

First Name ..... Hayden

Student Number 300375260

Course Title ..... Introduction to Artificial Intelligence

Course Code and No. ..... ~~30M~~ COMP307

Examiners use only	
1	15 ✓
2	18 ✓
3	10 ✓
4	12 ✓
5	10 ✓
6	19 ✓
7	17.5 ✓
8	5.5 ✓
Total	107

This booklet must be handed in at the end of the examination.

PLEASE READ THESE INSTRUCTIONS CAREFULLY

1. Fill in clearly all blanks for insertion of surname, first name, student number, course title and code.
2. Be sure to read the instructions at the head of each section very carefully; also any further instructions for any specific question.
3. Cross out neatly any work not intended for the examiner.
4. Supplementary books are available from the supervisor.
5. Tie this and any supplementary books together and hand the whole script to the supervisor.
6. Any script or portion of a script which is removed from the examination room, whether inadvertently or otherwise, will not be marked.
7. Rough work must be done in the answer book and any work not intended for the examiner should be crossed out.
8. Do not communicate with the examiner, either on the script or in any manner except through the Faculty Administration Manager, Victoria University of Wellington.

Student ID: 300375260

TE WHARE WĀNANGA O TE ŪPOKO O TE IKA A MĀUI
VICTORIA
UNIVERSITY OF WELLINGTON

EXAMINATIONS – 2019

TRIMESTER 1

COMP 307

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 to English dictionaries are permitted.

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.

2

travelling salesman problem
circuit analysis
video game AI
optimization AI (~~regression~~ eg. regression)

(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.

3

uniform cost - the node with the shortest path to the node

greedy best-first - the node with the smallest heuristic function ie $\min(h(n))$

A* - the node with the smallest distance from the initial ~~not~~ node to the node + heuristic function

ie $\min(d(n) + h(n))$

distance to node n

heuristic (estimate) from node n to goal

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

5

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

(ii) For each of the three search methods, state a common data structure (e.g. stack, queue, priority queue) it uses in the general tree-search algorithm.

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

i. Repeatedly perform a depth limited search starting from depth 0 and increasing depth by 1 each time until either a solution is found or maximum depth is reached

ii. depth-first: stack
breadth-first: queue
iterative deepening: stack

iii. I agree that the sequence of new (unchecked) nodes checked is the same as breadth first search. Each new depth checked will check one layer deeper than the previous iteration, which in essence gives the same ordering as breadth-first search.

(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.

4

(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 differentiable 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 differentiable 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. gradient descent, as it is based on continuous data and a local optimum is required

ii. genetic beam search, as a global optimum is required (the other two can only find local)

iii. hill climbing, as it is based on discrete data and local optimum required

iv. genetic beam search, same as ii.

(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.

1

hill climbing will find the ~~be~~ neighbouring state with the best improvement and move in that direction

simulated annealing will choose a random neighbouring state, and if it is better move towards it otherwise move towards it anyway based on some probability effected by the temperature

Student ID: 300175260

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
classifying species of plant
(2) Clustering: un supervised
grouping stars into star clusters
(3) Regression: supervised calculating value that a piece of meat is worth given input factors
~~creating a function to model given data~~

3

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

classification will output the predicted class of the input
regression will output a continuous number based on the input

1

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

(i) Induction learning:
generate a set of rules based on training data

(ii) Case-based learning:
use values of ~~training~~ training data to determine output

(iii) Genetic/evolutionary learning:
darwinian evolution/survival of the fittest in a population

(iv) Connectionist learning:
simulated neurons/simulate brain activity

4

(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 ~~class~~ system

test set is run against a trained system in order to evaluate effectiveness

validation set is checked against the trained system every n epochs in order to ensure training is not overfitting to the data, but is not directly used in ~~testing~~ training.

3

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

Reduces the risk of over-fitting to the data provided.

1

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

2) ~~keep~~ the k value ~~relatively~~ small as small as required

2) use up to k nearest neighbours instead of always k, and only ~~check~~ check a certain distance from the added node.

1

(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*?

i. Degree $= \frac{2}{8} \times \frac{1}{2} \times \frac{1}{2} + \frac{4}{8} \times \frac{3}{4} \times \frac{1}{4} + \frac{2}{8} \times \frac{0}{2} \times \frac{2}{2} = 0.15625$

work experience $= \frac{4}{8} \times \frac{3}{4} \times \frac{1}{4} + \frac{4}{8} \times \frac{1}{4} \times \frac{3}{4} = 0.1875$

visa needed $= \frac{3}{8} \times \frac{3}{3} \times \frac{0}{3} + \frac{5}{8} \times \frac{1}{5} \times \frac{4}{5} = 0.1$

ii. visa needed

Student ID: 300375260

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

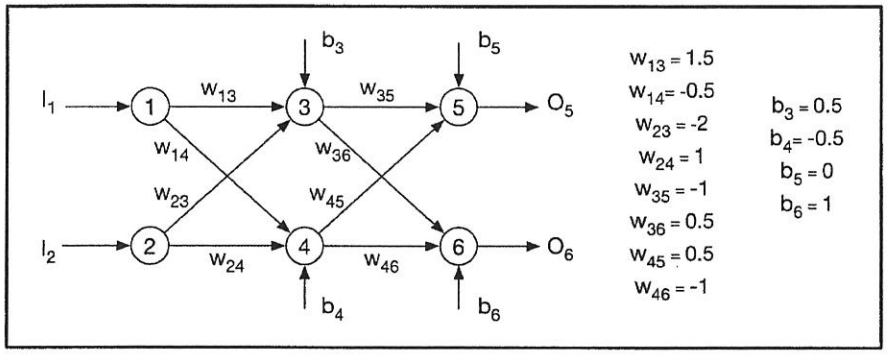
Cross out rough working that you do not want marked.
Specify the question number for work that you do want marked.

Student ID: 30037526 0

Question 3. Neural Networks [10 marks]

5

(a) [5 marks] Consider the following feedforward 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.

i. $I_1 = O_1 = 1$ $I_2 = O_2 = 1$ $I_3 = 1.5 \times O_1 + 0.5 = 2$

~~$O_3 = sig(2) = 0.881$~~ I

$1.5 \times O_1 + -2 \times O_2 + 0.5 = 0$

$O_3 = sig(0) = 0.5$ $I_4 = -0.5 \times O_2 + 1 \times O_2 + -0.5 = 0$

$O_4 = sig(0) = 0.5$ $I_5 = -1 \times O_3 + 0.5 \times O_4 + 0 = \sout{0.75}\ -0.25$

$O_5 = sig(-0.25) = 0.4378$ $I_6 = 0.5 \times O_3 + -1 \times O_4 + 1 = 0.75$

$O_6 = sig(0.75) = 0.6792$

ii. $\beta_5 = 1 - sig(-0.25) = 0.5622$

$\beta_6 = 1 - sig(0.75) = 0.3208$

$\beta_3 = -1 \times O_5 \times (1 - O_5) \times \beta_5 + 0.5 \times O_6 \times (1 - O_6) \times \beta_6$

$= -0.1034$

$\Delta w_{1 \to 3} = 0.2 \times 1 \times 0.5 \times (1 - 0.5) \times \beta_3$

$= -0.00517$

$w_{13}' = 1.5 + \Delta w_{1 \to 3} = 1.4948$

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

The xor problem is not linearly ~~be~~ 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 the 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. input layer: 4, one for each feature
output layer: 3, one for each class

ii. ~~Reduce the~~
- Reduce the number of hidden nodes
- Use a larger training set.

Student ID: 300375260

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.

0

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

2

genetic algorithms - ~~bits~~ strings
genetic programming - trees of nodes

(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.

5

i. # tournament selection - select k random individuals the select the fittest of those
roulette wheel - choose a random individual, fitter ~~fittest~~ individuals have higher probability of being chosen

ii. # of epochs
~~small~~ good enough fitness
overfitting detected.

Student ID: 300375260

(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 the terminal set and function set are (give an example if necessary), and what they are 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 the role of the classification strategy/translation rule in genetic programming for classification. Suggest three possible translation rules for using tree-based GP to perform multi-class classification.

5.

i. terminal set is a set of possible inputs to functions (i.e. the leaves of the tree). Usually features of data and a random number.
function set is a set of the possible functions that can be used (all the non-leaf nodes). The function set can take any terminal and the result of any other function.

ii. sufficiency: the primative set can produce a correct output
closure: every element in the function set can take all of the terminal set and all returned values in the function set as arguments

iii. Symbolic regression: mean-squared error
symbolic classification: ~~# of~~ % of correct classifications

iv. the classification strategy takes a single (usually) floating point output and translates it into a classification
- choosing multiple thresholds say, t_1, t_2, t_3 and if output $< t_1$ then C_1, else if output $< t_2$ then C_2 and so on
- return a vector from the GP and use the vector to classify (increase dimensionality)
- run output through multiple binary GP programs to 'narrow down' the classification

Question 5. Reasoning under Uncertainty [10 marks]

Use the probability rules discussed during the lectures to answer questions (a) to (c). 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. $P(B|A) = \frac{P(A,B)}{P(A)} = \frac{0.2}{0.5} = 0.4$

iii. $P(\neg B|A) = 1 - P(B|A) = 1 - 0.4 = 0.6$

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.

$P(A|C) = \frac{P(A,C)}{P(C)} = \frac{0.4}{0.5} = 0.8$

$P(B|C) = \frac{P(B,C)}{P(C)} = \frac{0.2}{0.5} = 0.4$

if $A \perp B|C$ then $P(A,B|C) = P(A|C)P(B|C)$

$0.32 = 0.8 \times 0.4$, so A and B are independent given C.

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

5

Degree	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 degree.

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

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

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

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

i. $\frac{33}{100} = 0.33$

ii. $\frac{8}{100} = 0.08$

iii. $\frac{8}{60} = 0.133$

iv. $\frac{9}{21} = 0.429$

v. If $A \perp B$ then $P(A,B) = P(A)P(B)$

$P(A,B) = \frac{10}{100} = 0.1$

$P(A) = \frac{10}{60} = 0.167$

$P(B) = \frac{10}{18} = 0.556$

$0.1 \neq \frac{10}{60} \times \frac{10}{18}$ so others and male are not independent.

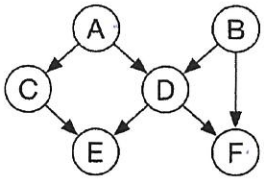
SPARE PAGE FOR EXTRA ANSWERS

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

Student ID: 300375260

Question 6. Bayesian Networks [20 marks]

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



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

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

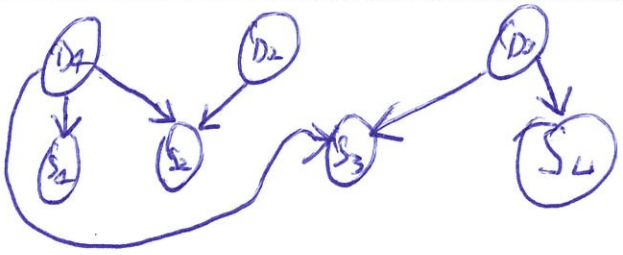
(iii) A and E are conditionally independent given C and D.

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

i. true ii. true
iii. true iv. ~~false~~ true

4 **(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.



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

$$P(D_1, D_2, D_3, S_1, S_2, S_3, S_4) = P(D_1)P(D_2)P(D_3)P(S_1|D_1)$$
$$P(S_2|D_1,D_2)P(S_3|D_1,D_3)P(S_4|D_3)$$

$P(C|D) = \frac{P(D|C)\,P(C)}{P(D)}$

6

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

initializing all at 1 for both problems

i. $Y - \frac{4}{7} \times \frac{3}{7} \times \frac{4}{8} \times \frac{5}{7} \times \frac{6}{10} = 0.035$

$N - \frac{2}{5} \times \frac{3}{5} \times \frac{2}{5} \times \frac{2}{5} \times \frac{4}{10} = 0.0128$

$Y > N$ so classify as Y

ii. $Y - \frac{4}{7} \times \frac{3}{7} \times \frac{1}{8} \times \frac{5}{7} \times \frac{6}{10} = 0.02187$

$N - \frac{2}{5} \times \frac{2}{5} \times \frac{2}{8} \times \frac{2}{5} \times \frac{4}{10} = 0.00853$

$Y > N$ so classify as Y

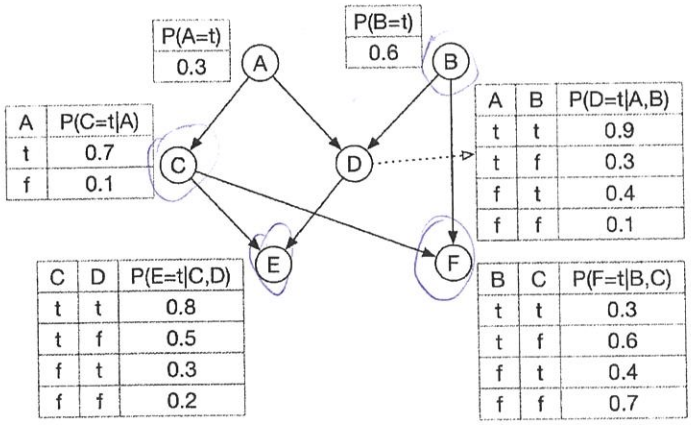
Student ID: 30037526O

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 $\frac{0.41+0.17}{0.32}$.]



P(A=t)
0.3

P(B=t)
0.6

A	P(C=t|A)
t	0.7
f	0.1

A	B	P(D=t|A,B)
t	t	0.9
t	f	0.3
f	t	0.4
f	f	0.1

C	D	P(E=t|C,D)
t	t	0.8
t	f	0.5
f	t	0.3
f	f	0.2

B	C	P(F=t|B,C)
t	t	0.3
t	f	0.6
f	t	0.4
f	f	0.7

2

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

16

Student ID: 300375260

Shorthand

$P(B) = P(B=t)$
$P(\neg B) = P(B=f)$

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

$$P(B, C, E, \neg F) = P(A, B, C, D, E, \neg F) + P(\neg A, B, C, D, E, \neg F) + P(A, B, C, \neg D, E, \neg F) + P(A, B, C, \neg D, E, \neg F)$$

$$= 0.3 \times 0.6 \times 0.7 \times 0.9 \times 0.8 \times 0.7 + 0.7 \times 0.6 \times 0.1 \times 0.4 \times 0.8 \times 0.7 + 0.3 \times 0.6 \times 0.7 \times 0.1 \times 0.5 \times 0.7 + 0.7 \times 0.6 \times 0.1 \times 0.6 \times 0.5 \times 0.7$$

$$= 0.086142$$

Question 7. Planning and Scheduling [20 marks]

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

choosing the best sequence of actions to reach a given 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) does not remove a part of the ~~goal~~ initial state from the current state

2) adds a portion of the initial state to the current state

(c) [5 marks] Consider a robot to change a flat tyre. The initial state has a flat tyre on the axle and a good spare tyre in the boot. The goal is to have a good spare tyre properly mounted onto the car's axle. There are four actions: removing the spare from the boot, removing the flat tyre 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 spare and flat tyres both 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*.

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

Goal($At(Spare, Axle)$)

$Action(Remove(Spare, Boot),$

PRECOND : $At(Spare, Boot)$

EFFECT : $\neg At(Spare, Boot) \wedge At(Spare, Ground))$

$Action(Remove(Flat, Axle),$

PRECOND : $At(Flat, Axle)$

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

$Action(PutOn(Spare, Axle),$

PRECOND: $At(Spare, Ground) \wedge \neg At(x, Axle)$

EFFECT: $\neg At(Spare, Ground) \wedge At(Spare, Axle)$

$Action(LeaveOvernight,$

PRECOND : ~~$At(Spare, x) \wedge At(Spare, Ground)$~~ $\wedge At(Spare, x) \wedge At(Flat, y)$

EFFECT: $\neg At(Spare, x) \wedge \neg At(Flat, y)$

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

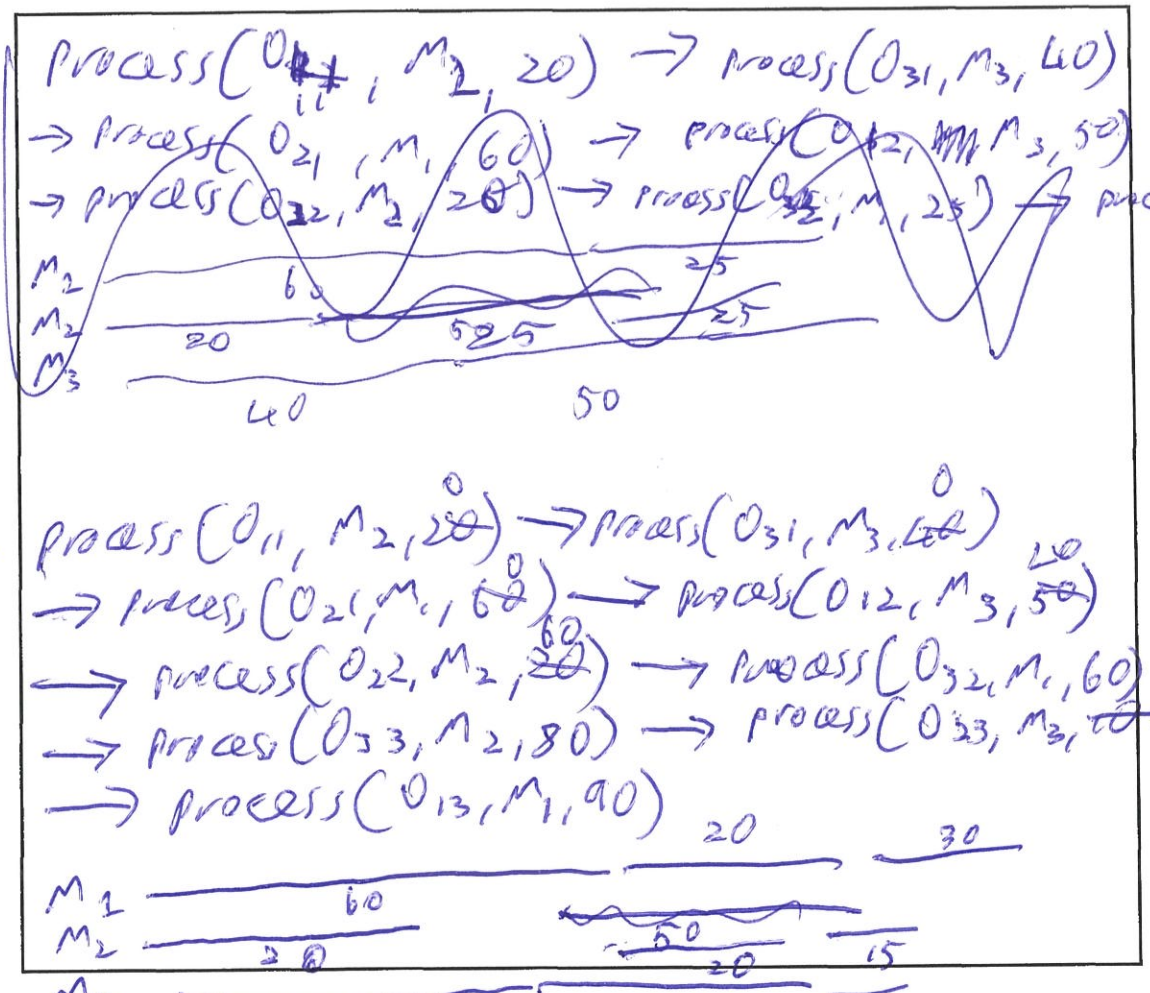
$At(Flat, Axle) \wedge At(Spare, Ground)$

(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? Show your working.



i. Process(O_{11}, M_2, 0) → Process(O_{31}, M_3, 0) → Process(O_{21}, M_1, 0) → Process(O_{12}, M_3, 40) → Process(O_{22}, M_2, 20) → Process(O_{32}, M_1, 60) → Process(O_{33}, M_3, 90) → Process(O_{13}, M_1, 90)

ii. longest part of gantt chart is for M_2
60 + 20 + 30 = 110

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

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

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 $R_1 = (1, \ldots, 1)$, $R_2 = (1, \ldots, 1), \ldots$.

6

$R_1 = (1, 2, 3, 5, 1)$

$R_2 = (1, 6, 7, 9, 1)$

$R_3 = (1, 4, 8, 10, 1)$

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.

AI assistants
translation

1

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

Volume
veracity
~~variety~~ value
velocity
variability

3

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

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

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

~~SVM~~
MP

i). maps it to another dimension where reactions may be linearly separable

1.5

Student ID: 7207726o

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

0

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

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.

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.

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