☑ By submitting this assessment, I agree to the following declaration:

As a member of the University's student body, I will complete this assessment in a fair, honest, responsible and trustworthy manner. This means that:

- I will not seek out any unauthorised help in completing this assessment. (NB. Unauthorised help includes a tutorial or answer service whether in person or online, friends or family, etc.)
- I will not discuss or share the content of the assessment with anyone else in any form, including but not limited to, Canvas, Piazza, Chegg, Facebook, Twitter, Discord or any other social media within the assessment period.
- I will not reproduce and/or share the content of this assessment in any domain or in any form where it may be accessed by a third party.
- I am aware the University of Auckland may use Turnitin or any other plagiarism detecting methods to check my content.
- I declare that this assessment is my own work, except where acknowledged appropriately (e.g., use of referencing).
- I declare that this work has not been submitted for academic credit in another University of Auckland course, or elsewhere.
- I declare that I generated the calculations and data in this assessment independently, using only the tools and resources defined for use in this assessment.
- I declare that I composed the writing and/or translations in this assessment independently, using only the tools and resources defined for use in this assessment.

I understand the University expects all students to complete coursework with integrity and honesty. I promise to complete all online assessment with the same academic integrity standards and values.

Any identified form of poor academic practice or academic misconduct will be followed up and may result in disciplinary action.

○ I agree

1 **[4 marks]** You are given a dataset X with class attribute C. Explain the general procedure to train and evaluate a SVM model when the parameters of the SVM need to be optimised. SVM is just an example model here, the procedure would be the same with any model.

Fill in your answer here

Format

Words: 0/200

Maximum marks: 4

2 **[4 marks]** Give a decision tree for the following Boolean function:

- $\neg A \wedge \neg B \vee C$

Upload your file here. Maximum one file.

The following file types are allowed: **.pdf** Maximum file size is **1 GB**

Select file to upload

Maximum marks: 4

3 [10 marks]

If you have a fully connected neural network with 3 variable inputs, 1 hidden layer with 2 nodes and an output layer with 5 nodes, how many weights will you learn? Remember a weight is associated with an edge between two nodes. You must explain your working to get full marks.

Fill in your answer here

How many bias terms will you learn? You must explain your working to get full marks.

Fill in your answer here

What will be the form of the hypothesis returned by this neural network algorithm? You must explain your working to get full marks.

Fill in your answer here

If you add a second layer of hidden units with 2 nodes, how many more numbers will there be in your hypothesis. You must explain your working to get full marks.

Fill in your answer here

What is the size of the set of all possible hypotheses? You must explain your working to get full marks.

Fill in your answer here

Maximum marks: 10

4 **[8 marks]** If you have a population with two individuals 101010 and 010100:

What is the minimum number of applications of single point cross-over to get 101100? You must explain your working to get full marks.

Fill in your answer here

What is the minimum number of applications of single point cross-over to get 101101? You must explain your working to get full marks.

Fill in your answer here

What is the minimum number of applications of two point cross-over to get 000000? You must explain your working to get full marks.

Fill in your answer here

What is the minimum number of applications of mutation to get 111111? You must explain your working to get full marks.

Fill in your answer here

Maximum marks: 8

5 [9 marks] Calculate the feature ranking based on Relief for the following data set:

Instance	Features			
	F1	**F2**	**F3**	**Class**
I1	10	0	0	A
I2	10	10	0	A
I3	0	10	10	A
I4	0	0	5	B
I5	10	0	5	B
I6	0	0	0	B

Use the following distance matrix in your calculations:

	I1	**I2**	**I3**	**I4**	**I5**	**I6**
I1	0	10	30	15	5	10
I2	10	0	20	25	15	20
I3	30	20	0	15	25	20
I4	15	25	15	0	10	5
I5	5	15	25	10	0	15
I6	10	20	20	5	15	0

Assume that the weights will be calculated based only on two iterations of random sampling. Assume that I2 was the randomly selected instance in the first iteration, and I4 was the randomly selected instance in the second iteration. Remember that the weight w_i for an instance x_i is calculated by:

$w_i = w_i - (x_i - h_i)^2 + (x_i - m_i)^2$,

with *nearHit* instance h_i, and *nearMiss* instance m_i.

1. What will be the weight for each feature (attribute) after two iterations?
2. Provide the feature ranking from most important to least important feature.
3. What are the two most important features?

Show your working step-by-step (you must show how you calculated the weights and the ranks) and explain your answers.

Fill in your answer here

Format

Words: 0/200

Maximum marks: 9

6 **[13 marks]** You are given a cancer data set that describes patients with four Boolean attributes: tumor stage above 2 (*TS*), chemotherapy (*CT*), drug therapy (*DT*), and the outcome of the treatment (*O*), which is the target/class variable.

TS	CT	DT	O
F	F	T	pos
F	T	F	pos
T	T	F	pos
T	F	F	neg
T	F	T	neg

1. Use Naive Bayes to predict the target/class for the following new case - $case_1$:
   $case_1(TS = T, CT = T, DT = T)$
   What are the model predictions for this case?
2. What is the probability that the model predicts a positive prognosis for $case_1$?
3. What is the likelihood of $case_1$, given that the positive class hypothesis is true?
4. The trained Naive Bayes model mostly predicts one of the classes. Explain this effect. Explain which examples would be predicted as the other class?

Show your working step-by-step (you must show how you calculated all probabilities) and explain your answers.

Fill in your answer here

Format

Words: 0/400

Maximum marks: 13

7 [12 marks] You are given a small training data set with two input features (F_1, F_2):

ID	F_1	F_2	Target
A	-1	2	positive
B	-1	1	positive
C	-1	-1	positive
D	-1	-2	positive
E	1	0	negative
F	2	-2	negative
G	5	2	negative

1. If you train a linear hard margin support vector machine on this data set, what will the decision boundary look like? Draw and properly annotate both the decision boundary and the margins in the input space. What are the support vectors for this data set? *Hint: The fastest way is to draw it by hand.*
2. What is the minimum number of examples that you can remove to change the decision boundary? Give the ID of these examples. Which example can you safely remove without changing the decision boundary?
3. How many examples will be misclassified on the training set with this hard margin classifier? What is the leave-one-out error rate of this classifier (average number of misclassified examples)?
4. Given a 1-Nearest Neighbour (1-NN) classifier using this distance matrix:

	A	**B**	**C**	**D**	**E**	**F**	**G**
A	0	1	9	16	8	25	36
B	1	0	4	9	5	18	45
C	9	4	0	1	5	10	45
D	16	9	1	0	8	9	52
E	8	5	5	8	0	5	20
F	25	18	10	9	5	0	25
G	36	45	45	52	20	25	0

   Will the 1-NN be better or worse than the SVM in terms of the leave-one-out error rate?

Show your workings step-by-step and explain your answers.

Upload your PDF file here. Please ensure your PDF file is in portrait mode. Maximum one file.

The following file types are allowed: **.pdf** Maximum file size is **1 GB**

Select file to upload

Maximum marks: 12

8 [10 marks] Consider a vacuum cleaner robot that moves and cleans your floor. The robot's goal is to clean the floor as best as possible while moving fast from an initial point to a destination point. The possible state transitions, actions, and rewards are shown in the following table:

state s_t	action a_t	reward	next state s_{t+1}	$P(s_{t+1}\|a_t, s_t)$
on.cool	normal	4	on.cool	1
on.cool	turbo	10	on.cool	1/4
on.cool	turbo	10	on.hot	3/4
on.hot	normal	4	on.cool	1/2
on.hot	normal	4	on.hot	1/2
on.hot	turbo	10	on.hot	7/8
on.hot	turbo	10	off	1/8

Note the rewards are deterministic, only the state transitions are probabilistic. Once the robot reaches the state off, it will stay in that state with zero reward.

1. Assuming that the discount factor, $\gamma = 0.8$ and the robot follows a conservative control strategy π^c, i.e. always selects normal mode of operation. What is the value of $V^{\pi^c}(on.\,cool)$ under the conservative policy? Recall that $V(s)$ is the expected discounted sum of rewards when starting at state s.
2. Specify the optimal policy π^* for each state, for $\gamma = 0.8$.
3. Now assume the robot does not know the probabilistic transition function, and uses the Q-learning method to achieve its goal. Assuming the initial values of the Q-table are 0, and the agent performs the following steps $(on.\,cool, turbo) \rightarrow (on.\,hot, turbo)$ ending in state $on.\,hot$. Calculate the values for the Q-table after these steps. What is the value of $V(on.\,hot)$ after these steps?

Show your working step-by-step (you must show how you calculated all values) and explain all your answers.

Fill in your answer here

Format

Words: 0/300

Maximum marks: 10

9 **[2 marks]** Compare FP-Growth with the Apriori algorithm in terms of the following aspects.

- Computational time performance
- Memory performance
- Candidate generation efficiency
- Database scanning overhead

Fill in your answer here

Format

Words: 0/60

Maximum marks: 2

10

1. Consider a transaction database with five transactions below. Use FP-Growth algorithm to find all frequent itemsets (in descending order of frequency) with min_sup = 60%, and min_conf = 80%. **[7 marks]**

TID	Items Bought
T1	A, K, N, T
T2	A, G, N, P, R, S
T3	G, K, N, S, T
T4	G, K, S, T, W
T5	A, G, K, S, W

2. Construct an FP-Tree with a sorting order of your choice. Based on your informed judgement, please discuss the compactness of FP-Tree representation under these two settings. You may choose to use diagram(s) to aid in your discussion. **[5 marks]**

Upload your file here. Maximum one file.

The following file types are allowed: **.pdf** Maximum file size is **1 GB**

Select file to upload

Maximum marks: 12

11

1. Consider 8 examples: A1=(2,10), A2=(2,5), A3=(8,4), A4=(5,8), A5=(7,5), A6=(6,4), A7=(1,2), A8=(4,9) and the corresponding distance matrix based on the Euclidean distance. Use each of the follow algorithms to group the data separately. **[9 marks]**
   - Single link agglomerative clustering (show the dendrograms)
   - DBSCAN with Epsilon=10 and Minpoint=2 (show clusters and outliers)

	A1	A2	A3	A4	A5	A6	A7	A8
A1	0	45	63	57	41	28	95	6
A2		0	55	49	35	11	5	25
A3			0	11	23	54	47	65
A4				0	2	7	26	5
A5					0	5	21	35
A6						0	13	27
A7							0	53
A8								0

2. Draw a 10 by 10 space with all the 8 points to show the clusters for each algorithm above separately. **[4 marks]**

Upload your file here. Maximum one file.

The following file types are allowed: **.pdf** Maximum file size is **1 GB**

Select file to upload

Maximum marks: 13

12 **[3 marks]** Evaluate and compare the discovered clusters in the Question 11 using an internal measurement of your choice. You may need to determine an appropriate cutting threshold for agglomerative clustering task.

Fill in your answer here

Format

Words: 0/75

Maximum marks: 3

13

1. Supervised outlier detection approaches build a classification model to classify unseen events. Discuss a major challenge of classification-based outlier detection. Explain how to resolve this specific challenge. **[4 marks]**
2. Proximity-based techniques are examples of unsupervised learning approaches for point-based anomaly detection. Discuss a scenario where density-based outlier detection may fail and explain how to resolve this. **[4 marks]**
3. Clustering-based anomaly detection assumes normal data belong to large and dense clusters, while anomalies do not belong to any of the clusters or form very small clusters. Discuss a scenario where clustering-based anomaly detection may fail. Explain how to resolve the specific case. **[4 marks]**

Fill in your answer here

Format

Words: 0/200

Maximum marks: 12

14

1. Consider an infinite bit stream with a bit value of either 0 or 1. Discuss one counting approach of your choice to tackle questions like "*how many 1s are there in the last k bits (k ≤ N, and N=1 billion)?*" Explain the advantages of your choice. **[4 marks]**
2. Consider two drift detectors: DDM and ADWIN algorithms. Discuss how they work with respect to the following two aspects: **[4 marks]**
    - How do ADWIN and DDM detect concept drifts?
    - How do ADWIN and DDM remember discarded data?

Fill in your answer here

Format | B I U x_2 x^2 I_x | | | Ω | | | Σ |

Words: 0/150

Maximum marks: 8