THE UNIVERSITY OF AUCKLAND

SEMESTER ONE 2021
Campus: City

COMPUTER SCIENCE 762

Machine Learning

NOTE:
This quiz is out of **30** marks.
Attempt **ALL** questions.
Upload your answers as a pdf document.
This quiz contributes 10% to your final grade.

We will give you the online quiz in both Word and PDF format. You will need to put your answers into an editor or a document preparation system and save it as a pdf. Upload the pdf to Canvas, like you do for the assignments.

You may wish to print out the quiz write on it and upload your answers. Or you may wish to draw a picture and upload the picture into your document. Please have the means to accomplish this, at your disposal during the quiz. Also arrange in advance a quiet location for completing the quiz undisturbed.

This quiz has been designed as an open book one hour quiz. However you have 1 hour and 15 minutes to upload this quiz. It is an open internet/open book quiz. You will need to confirm that you will not discuss the content of the quiz with anyone else for 24 hours following the quiz. You will not give any assistance to another student taking this quiz. You will not receive any assistance from any person or tutoring service. You will not post any information about this quiz on the Internet.

The work still must be your own. It will be run through Turnitin, but you will not receive the originality report until after the assignment is due. You will be asked to include a sentence which states that the quiz is your own work.

Make sure you look how much the questions are worth so you can decide which questions to spend your time on.

Answer all questions in a document and save it as a pdf document. Name your document with your UPI followed by CS762Quiz.pdf e.g., abcd123CS762Quiz.pdf. Upload the pdf to Canvas.

1. **Integrity Statement [The quiz will be marked as 0 if you don't add your name]**

"*For the duration of this quiz, I, (insert your name here), confirm that I will not discuss the content of the quiz with anyone else for the next 24 hours. I will not give any assistance to another student taking this quiz. I will not receive any assistance from any person or tutoring service. I will not post any information about this quiz on the Internet.*"

2. **Machine Learning Foundations [6 marks]**

You are running a business. You had some data to analyse and have 3 scenarios:

Scenario 1: They took your data, divided it in ½, fit one model, and reported an accuracy of 85% on the second ½ of the data.

Scenario 2: They took your data, divided it in ½, fit 10 models, and one of the models reported an accuracy of 85% on the second ½ of the data.

Scenario 3: They took your data, divided it in ½, fit 10,000 models, and one of the models reported an accuracy of 85% on the second ½ of the data.

Scenario 4: They repeat Scenario 1 ten times and report the average accuracy of the ten models.

What are the pros and cons of each scenario? Which model will you finally use and why? You will get most of the points for explaining the "why".

The answer should be put here.

3. **Machine Learning Foundations [5 marks]**

If your model does not do well on the test set, can you just put the data back into one pot and divide it into training and test sets again? If it is appropriate under some circumstances, what are these circumstances? If it is not appropriate under some circumstances, what are these circumstances? You will get most of the points for explaining the "why".

The answer should be put here.

4. **ROC curves (10 Marks)**

Plot a ROC curve for the thresholds 0, 0.2, 0.4, 0.6, 0.8, and 1. Make sure to annotate the plot (x-axis, y-axis, …) and label the curves. Show the calculations for each threshold. Please explain whether you would choose to use classifier 1 or classifier 2 and why.

Class	**Prediction for Classifier1**	**Prediction for Classifier2**
P	**0.83**	**0.92**
N	**0.78**	**0.62**
P	**0.66**	**0.89**
P	**0.62**	**0.52**
N	**0.48**	**0.49**
N	**0.32**	**0.38**

The answer should be put here.

5. **Artificial Neural Networks (5 Marks)**

Neural Networks use regularization. What problem is regularization meant to fix? How does regularization fix this problem? Describe one type of regularization. You will get most of the points for the explanations.

The answer should be put here.

6. Decision Trees [4 marks]

Assume you have a dataset with two-dimensional points from two different classes C_1 and C_2. The points from class C_1 are given by $A=\{(i,i^2)|i\in\{1,...,100\}\}\subseteq R^2$, while the points from class C_2 are $B=\{(i,125/i)|i\in\{1,...,100\}\}\subseteq R^2$. So the data set will look like this:

class	x_1	x_2
C_1	1	1
C_1	2	4
C_1	3	9
...	...	...
C_1	100	10000
C_2	1	125
C_2	2	62.5
C_2	3	41.666
...	...	...
C_1	100	1.25

Construct a decision tree of minimal depth that assigns as many data points as possible to the correct class. Provide for each split the feature and corresponding thresholds. How many and which data points are miss-classified?

The answer should be put here.