THE UNIVERSITY OF AUCKLAND

SEMESTER ONE 2020
Campus: City

COMPUTER SCIENCE 361

Machine Learning

NOTE:
This test is out of **45** marks.
Attempt **ALL** questions.
Upload your answers as a pdf document.
This test contributes 22% to your final grade.

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.

This test has been designed as an open book one hour test. However you have a full 24 hours to work on this test. You are allowed to use the full 24 hours and you are allowed to use books, notes and the internet. It is possible that you will have to look outside the notes provided to answer the questions. It is an open internet/open book test. You will need to confirm that you will not discuss the content of the test with anyone else. You will not give any assistance to another student taking this test. You will not receive any assistance from any person or tutoring service. You will not post any information about this test on the Internet.

The work still must be your own. We will run the test through Turnitin. 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 test is your own work.

Answer all questions in a document and save as a pdf document. Name your document with your upi followed by CS361Test.pdf e.g., abcd123CS361Test.pdf. Upload the pdf to Canvas.

1. Integrity Statement (1 Marks)

Please include as the answer to question 1.

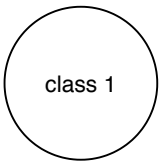
"*For the 24 hour duration of this test, I, (insert your name here), confirm that I will not discuss the content of the test with anyone else. I will not give any assistance to another student taking this test. I will not receive any assistance from any person or tutoring service. I will not post any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

2. Overfitting (10 Marks)

Given the 4 decision trees shown below. Explain for each tree why you think it might be (a) overfit (b) underfit or (c) a good model. Each tree does not necessarily have a single answer.

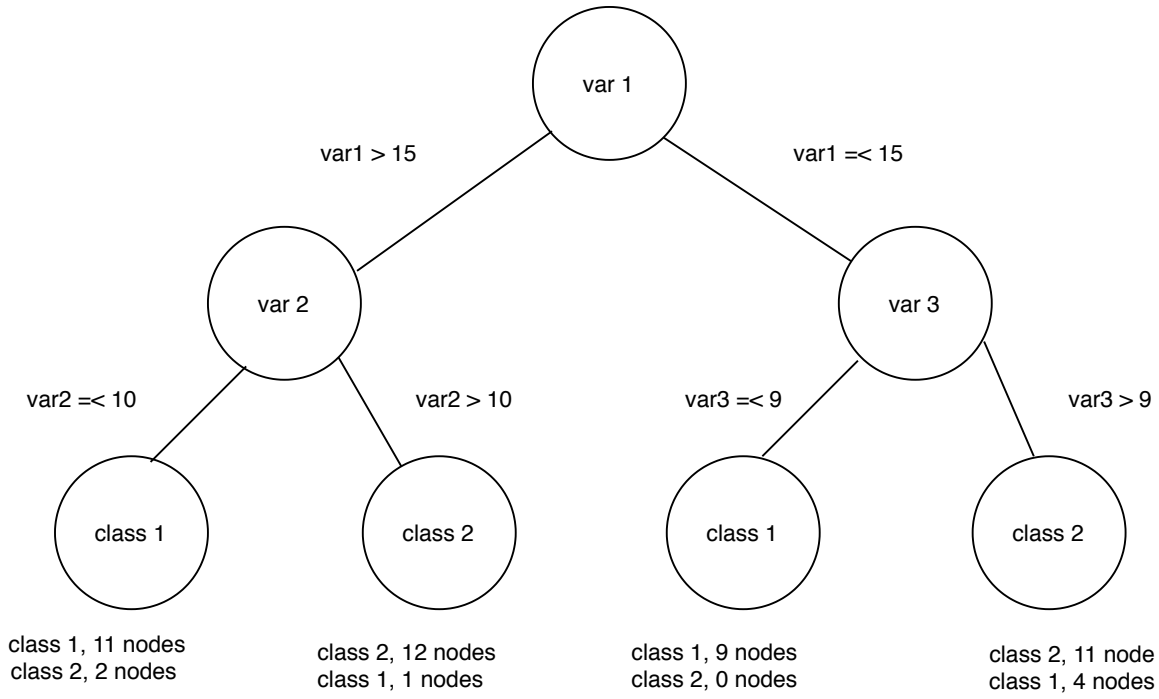
[10 marks]

Tree 1

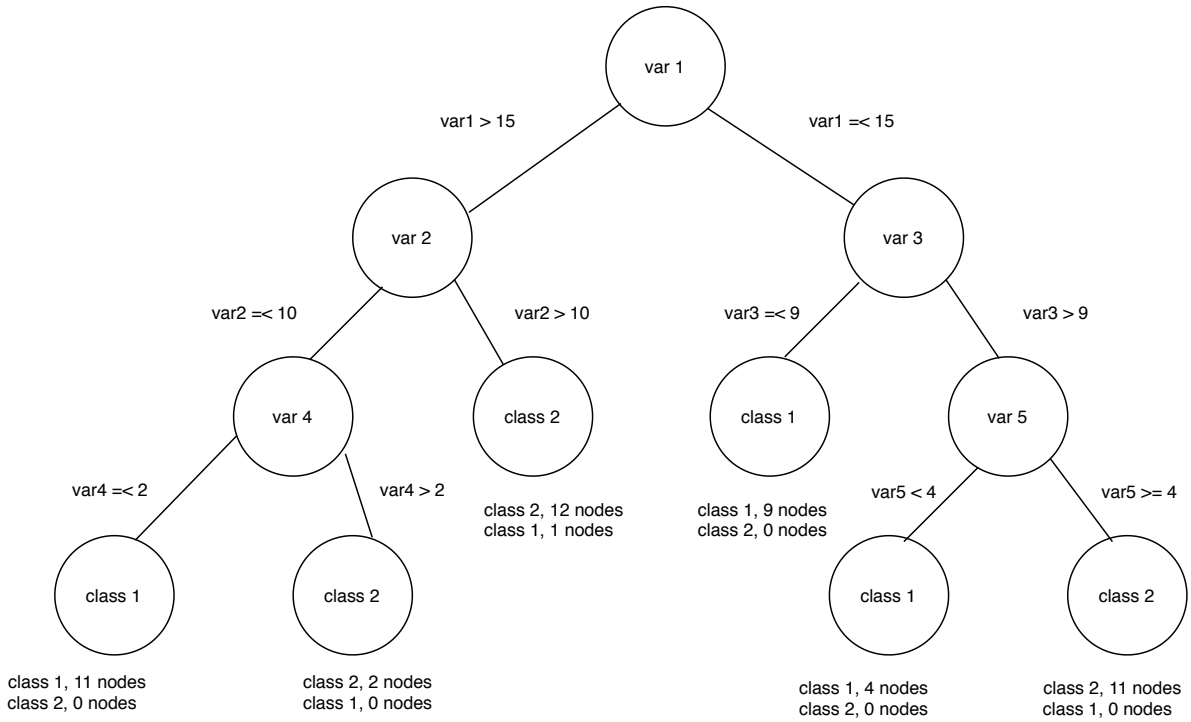


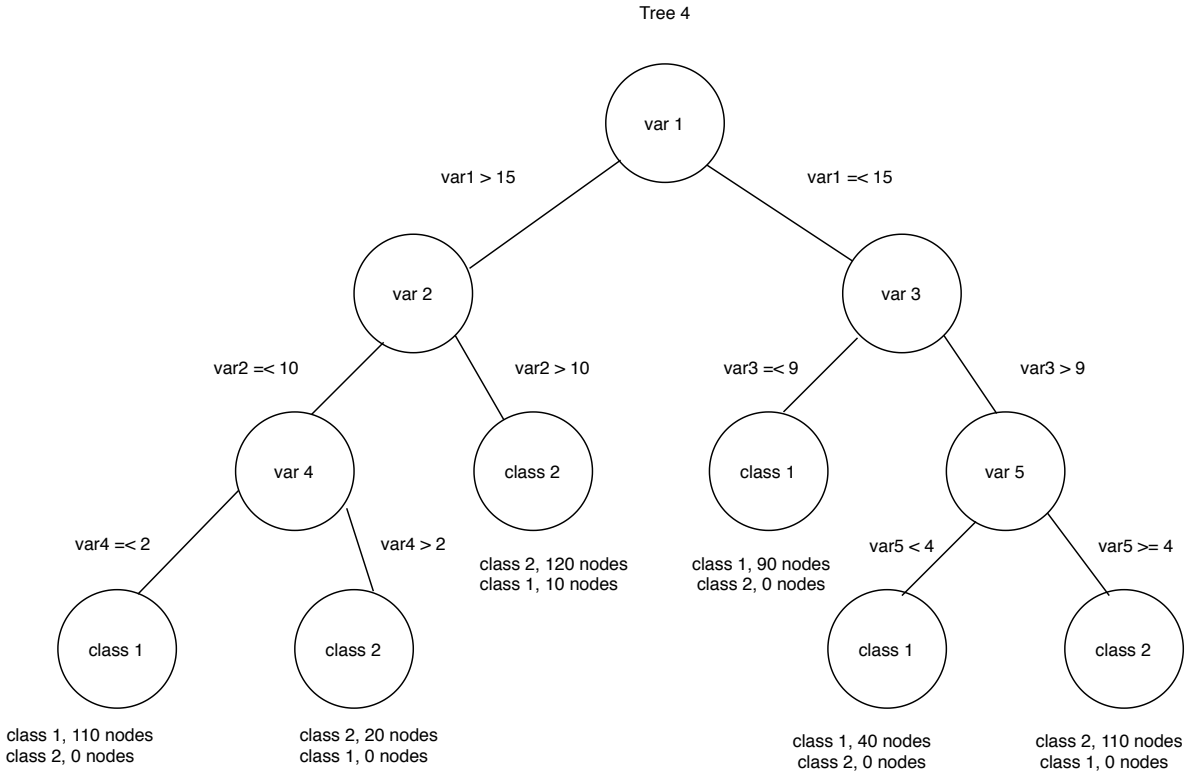
class 1, 27 nodes
class 2, 23 nodes

Tree 2


var 1
var1 > 15
var1 =< 15
var 2
var 3
var2 =< 10
var2 > 10
var3 =< 9
var3 > 9
class 1
class 2
class 1
class 2
class 1, 11 nodes
class 2, 2 nodes
class 2, 12 nodes
class 1, 1 nodes
class 1, 9 nodes
class 2, 0 nodes
class 2, 11 node
class 1, 4 nodes

Tree 3


var 1
var1 > 15
var1 =< 15
var 2
var 3
var2 =< 10
var2 > 10
var3 =< 9
var3 > 9
var 4
class 2
class 1
var 5
var4 =< 2
var4 > 2
var5 < 4
var5 >= 4
class 1
class 2
class 1
class 2
class 2, 12 nodes
class 1, 1 nodes
class 1, 9 nodes
class 2, 0 nodes
class 1, 11 nodes
class 2, 0 nodes
class 2, 2 nodes
class 1, 0 nodes
class 1, 4 nodes
class 2, 0 nodes
class 2, 11 nodes
class 1, 0 nodes



Each answer should be 100 words or less.

3. Bias (4 Marks)

Each Machine Learning Algorithm has a bias it uses to learn. We have discussed how a model that is underfit will have a large bias on the training set. Please discuss how these two uses of the word bias are connected to each other.

[4 marks]

The answer should be 100 words or less.

4. ANN – Activation Functions (8 Marks)

There are four main activation functions we have discussed in the class: step function, linear function, sigmoid function, rectified linear function. For each of them explain what their pros and cons are and what kind of models they can learn.

[8 marks]

Each answer should be 100 words or less.

5. ANN and local minima (4 Marks)

In neural networks we talked about getting stuck in a local minima. If you got stuck in a local minima would that be overfitting or underfitting or could it be either?

[4 marks]

The answer should be 100 words or less.

6. GAs (4 Marks)

In genetic algorithms we talked about overcrowding causing early convergence. Explain what this is and why it happens in genetic algorithms. Also explain whether this is more like to cause overfitting or underfitting.

[4 marks]

The answer should be 100 words or less.

7. Particle Swarms (4 Marks)

Particle swarms and genetic algorithms both have a population of solutions. Discuss how the information is shared between individuals in both algorithms and how they are similar and different.

[4 marks]

The answer should be 100 words or less

8. Practical Machine Learning (10 Marks)

To do a practical task for machine learning there are many things you must decide upon. Think of the checkers task in the early lectures. Let us say you were going to learn to play Connect Four (https://en.wikipedia.org/wiki/Connect_Four). What are the decisions you would have to make? Please specify the choices you would make to learn this game.

[10 marks]

The answer should be 300 words or less.