School of Computer Science and Applied Mathematics
University of the Witwatersrand, Johannesburg

ADAPTIVE COMPUTATION AND MACHINE LEARNING (COMS4030A/COMS7047A)

Class Test I

Date: 23rd April, 2021

1 Instructions

Please read the instructions carefully, before attempting the ACML(COMS4030A) class test.

1. This is an open book, open notes assessment.
2. Please make sure you attempt all the questions.
3. This is an online assessment created using Gradescope, all answers to be typed in the space provided.
4. If you need clarification for any of the questions, then post your queries on the "Test queries" discussion forum on Ulwazi. I will monitor this forum frequently during the scheduled ACML slot.
5. If you have made any assumptions in answering a question, please state it explicitly.
6. Even though the test is open for 24hrs, you are expected to use your weekly ACML slot (10:00-12:00) for attempting the test.
7. Do not leave the assessment until the last minute. Treat it as an on-campus scheduled test and complete the test as soon as possible.
8. No excuses will be entertained unless supported with proper documentation.
9. This assessment is out of 50 marks.
10. There are 6 questions in this paper, and you should attempt all 6 questions to collect 50 marks.
11. You should ideally finish this assessment in 2 hrs (this includes 30 mins extra time).
12. This assessment allows 2 attempts, only to cover for situations around network/power failures. Make sure you are fully aware of any issues before you attempt your test.
13. By attempting this assessment, you agree to the Honor code stated below. Any violations will results in a Failed Component (FCom) code for this assessment.
    Honor code:
    I, a student of COMS4030A, attest that I have not given or received aid for any questions from my fellow students/colleagues/other parties during this assessment. I have treated this assessment as a purely open book, open notes test. I understand that any violations will result in my assessment being coded as FCom based on Wits policies around plagiarism.

.

2 Questions

1. As the number of training examples goes to infinity, you model trained on the data will have {higher/lower/same} variance, {higher/lower/same} bias.{5 marks}

   Select the appropriate option and provide a justification for your selection.

2. Given a problem space where the output variable/label y_i consists of a range of real values (eg., y_1=0.5, y_2=1.0, y_3=2.5 and y_4=3.0), which machine learning paradigms listed below can be used and why? {5 marks}
   (a) supervised
   (b) unsupervised
   (c) both

3. Assume we implement an AND function using a single neuron whose activation function is denoted as

$$f(x) = \begin{cases} 0 \text{ for x} < 0 \\ 1, \text{ for x} \geq 0 \end{cases} \tag{1}$$

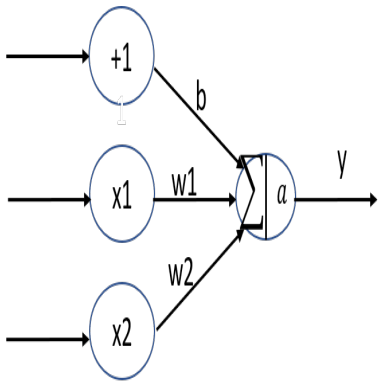

Figure 1: Neuron

For which vales of W1, W2 and b given below does the given neuron implement an AND function? Provide an intuition behind your selection. {5 marks}

(a) b = -1.5, w1 = 1, w2 = 1
(b) b = 1.5, w1 = 2, w2 = 2
(c) b = 1, w1 = 1.5, w2 = 1.5
(d) None of these

4. Looking at the below graphs, what conclusions can you make about how the regularization parameter λ affects a model? {5 marks}

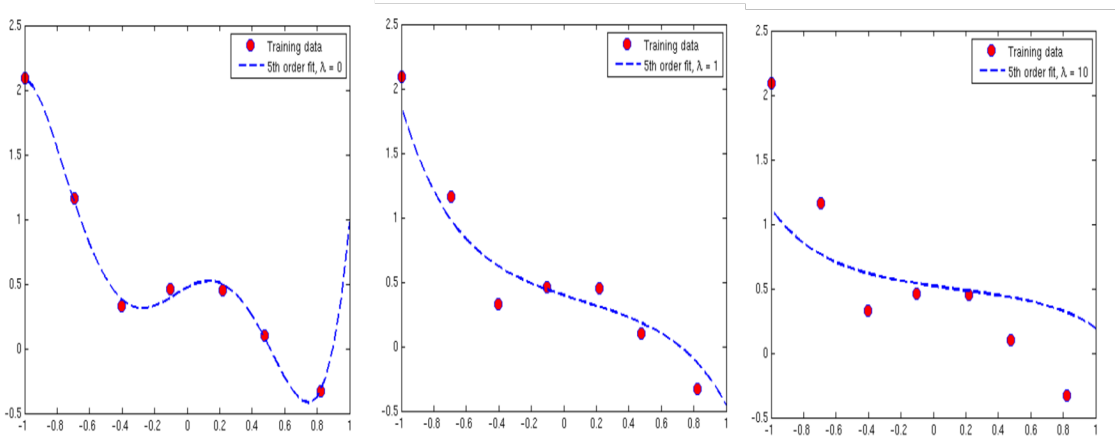

Figure 2: Reguralization

5. Given the below neural network architecture

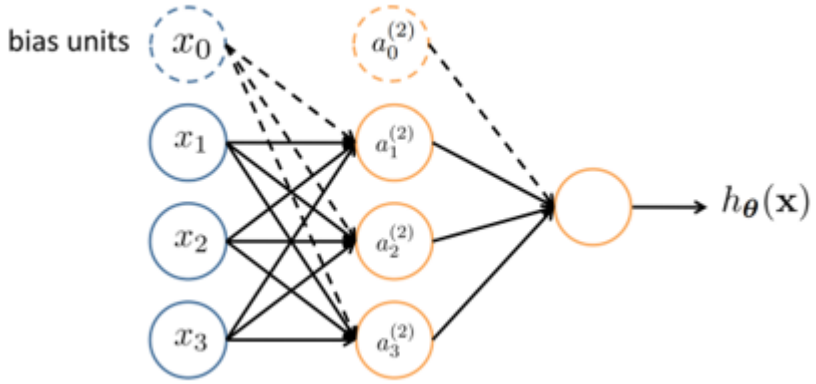

Figure 3: A simple neural network

(a) How many parameters does this network have? Explain your answer. {5 marks}

(b) If you were asked to now extend the network given in Fig 3, to recognize the presence/absence of cat in a given image, how would you modify the above network architecture (in terms of size of the input and output layers) and how should your data(input/output) be represented? {5 marks}

Assume that the images in the data set are of size 10*10 pixels RGB.

(c) A new data set containing images of only cats, lions or leopards is provided and you are tasked to perform a multi-class classification problem. What modifications would you introduce to the network in question (b), and why? {5 marks}

Assume that there is a single instance of either a cat or a lion or a leopard in a given image.

(d) Given the new architecture in question (c), do you think it will be robust enough to classify images containing instances of cat and lion, lion and a leopard, or cat and a leopard? Why or why not? Justify your answer. {5 marks}

6. The back-propagation algorithm requires us to calculate the error terms $\delta()$ in order to calculate the gradients (see below fig for reference), which will then be used to update the parameters using gradient descent.

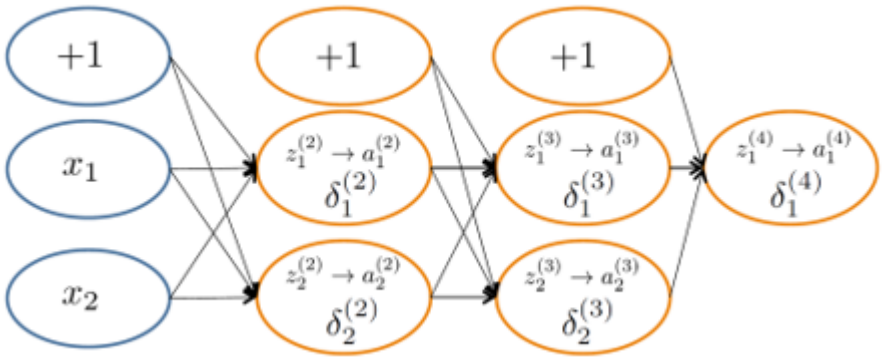

Figure 4: NN learning using backprop

(a) Complete the following chain rules used for calculating the errors. {5 marks}

$$\frac{\partial J(\theta)}{\partial z^3} = \frac{\partial J(\theta)}{\partial \_} \cdot \frac{\partial \_}{\partial \_} \cdot \frac{\partial \_}{\partial \_} \cdot \frac{\partial \_}{\partial \_}$$

$$\frac{\partial J(\theta)}{\partial z^2} = \frac{\partial J(\theta)}{\partial \_} \cdot \frac{\partial \_}{\partial \_} \cdot \frac{\partial \_}{\partial \_} \cdot \frac{\partial \_}{\partial \_} \cdot \frac{\partial \_}{\partial \_} \cdot \frac{\partial \_}{\partial \_}$$

Note: Do not bother about typing answers in mathematical formats. Use simple notation as a2, z3, J(theta), etc.

(b) In class, we used sigmoid as our activation function for all the neurons in the above network. In other words, a^2, a^3, a^4 were sigmoid/logistic functions. If we instead use linear functions for all the above activations (eg. a^2, a^3, a^4), how would the network behave while performing gradient descent? {5 marks}

Hint: Explain your answer w.r.t cost function.