

CHUKA UNIVERSITY

EXAMINATION FOR MASTER OF SCIENCE IN COMPUTER SCIENCE Y1S2

COSC 843 : MACHINE LEARNING

DATE:

TIME: 2 HRS

INSTRUCTIONS:

Answer Question ONE and ANY other TWO Questions

Question one (30 Marks)

- a) With aid of a diagram discuss the differences between machine learning and deep learning. (4 marks)
- b) Explain any four applications of Machine Learning in the modern healthcare. (4 marks)
- c) Discuss the difference between a Perceptron and Logistic Regression as applied in machine learning. (4 Marks)
- d) During Machine learning class you were tasked to build a random forest model with 8000 trees. Through its training, your model achieved a training error of 0.00. However, on testing the validation error was 34.23. Explain the phenomena and the solution to this phenomenon. (6 Marks)
- e) Discuss with an aid an of an illustration how you will design an e-mail spam filter? (6 marks)
- f) In class we discussed three different kinds of Unsupervised Learning problems. List the three types of problems and for each name a method which addresses that problem. (6 Marks)

Question 2 (15 Marks)

- a) With reference to the ID3 algorithm, explain clearly how the decision is made as to which attribute is selected to partition a set of examples. (5 marks)
- b) Use the ID3 algorithm to devise a set of rules for identifying bird's eggs (see table below). (10 marks)

Bird	Colour	Texture	Size
Herring Gull	Brown	Speckled	Medium
Starling	Blue	Smooth	Small
Mallard	Grey	Smooth	Medium
Swan	Grey	Smooth	Large
Guillemot	Brown	Speckled	Medium
Sparrow	Grey	Speckled	Small

Question 3 (15 Marks)

a) With aid of examples in each case differentiate between parametric and non-parametric models.

(3 Marks)

b) You are working on a time series data set. Your manager has asked you to build a high accuracy model. You start with the decision tree algorithm since you know it works fairly well on all kinds of data. Later, you tried a time series regression model and got higher accuracy than the decision tree model. Explain the occurrence.

(6 Marks)

c) The following is a table that shows the items bought in different transactions. Calculate (as a percentage) the support and confidence of:

(6 Marks)

- i. Steak \rightarrow Beer
- ii. Beer \rightarrow Steak

Transaction-id	Items bought
10	Steak, Sugar, Beer
20	Steak, Milk, Beer
30	Steak, Beer, Coffee
40	Sugar, Coffee, Bread

Question 4 (15 Marks)

- (a) With the aid of example discuss how a K-Nearest Neighbour is used as an algorithm in Machine Learning. **(5 marks)**
- (b) Considering a long list of Machine Learning Algorithms, given a Data Set, how do you decide which machine learning algorithm to apply for a given task. **(5 Marks)**
- (c) Discuss five applications of clustering in modern business. **(5 Marks)**

Question 5 (15 Marks)

- (a) With the aid of diagram illustrate the different layers on Convolution Neural Network(CNN). **(6 Marks)**
- (b) Explain what is Pooling on CNN, and How discuss how it works in deep learning. **(4 Marks)**
- (c) You built a machine learning model, and while training it, you noticed that after a certain number of epochs, the accuracy is decreasing. What's the problem and how to fix it? **(5 Marks)**