CHUKA  **UNIVERSITY**

UNIVERSITY EXAMINATIONS

EXAMINATION FOR THE AWARD OF DEGREE OF MASTER OF SCIENCE IN COMPUTER SCIENCE

COSC 843: MACHINE LEARNING

STREAMS: Y1S2 **TIME:2 HOURS**

DAY/DATE: FRIDAY 6/12/2019 **8.30 A.M – 10.30 A.M**

INSTRUCTIONS:

Answer all questions in section A and any two in section B

SECTION A (30 MARKS)

QUESTION 1(30 MARKS)

a. Define the following terms as used in machine learning: (4 Marks)
   a. Convergence
   b. Generalization
   c. Underfitting
   d. ID3

b. Explain *Linear Regression* and *Logistic regression* stating the application area for each? (2 Marks)

c. In class we discussed two different kinds of Unsupervised Learning problems. Discuss the two types of problems and for each name a method which addresses that problem.

   (2 marks)

d. What is overfitting? What strategies can help to reduce overfitting? (2 marks)

e. What problem will result from using a learning rate that is too high and too low? How would you detect this problem? (2 Marks)

f. Ensemble methods combine together many simple, poorly performing classifiers in order to produce a single, high quality classifier. Name two ensemble methods. (2 Marks)

g. Differentiate between CNN and ANN? (2 Marks)

h. Explain the advantages of linear rectified activation compared to logistic sigmoid activation. (2 Marks)

i. What is the principal assumption in the Naive Bayes' model, and when is this assumption useful? (2 marks)

j. Explain the training, validation and test sets as used in supervised learning? (3 marks)

k. Describe the K-fold cross-validation algorithm for model selection. (2 Marks)

l. Differentiate between likelihood and maximum likelihood. (2 marks)

m. How does one use boosting and decision stumps to feature selection? (3 marks)

SECTION B

QUESTION 2 (15 MARKS)

a. Consider the following set of training examples:

Instance	Classification	A1	A2
1	+	T	T
2	+	T	T
3	-	T	F
4	+	F	F
5	-	F	T
6	-	F	T

What is the information gain of a2 relative to these training examples? Provide the equation for calculating the information gain as well as the intermediate results.(3 Marks)

b. Explain four Applications of deep learning. (4 marks)

c. What are biological neurons? How do they help in creating artificial neuron model? (2 Marks)

d. Explain the principle of the gradient descent algorithm. Accompany your explanation with a diagram. Explain the use of all the terms and constants that you introduce and comment on the range of values that they can take. (6 Marks)

QUESTION 3 (15 MARKS)

a. Distinguish between non-parametric and parametric learning algorithms. (2 marks)

b. Derive the gradient descent training rule assuming that the target function representation is:

$$od = w0 + \ + \ldots + .$$

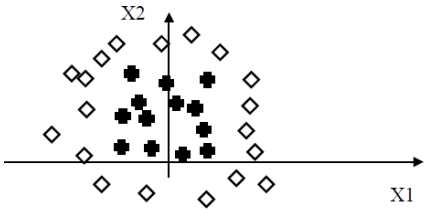
Define explicitly the cost/error function E, assuming that a set of training examples D is provided, where each training example d $\in$ D is associated with

the target output td. (5 marks)

c. Prove that the LMS training rule performs a gradient descent to minimize the cost/error function E defined in (4). (8 Marks)

QUESTION 4 (15 MARKS)

a. For the K-Means algorithm: What are the inputs? Which parameters are usually specified by the user, and what does the algorithm estimate? (3 marks)
b. What objective function does the K-Means algorithm minimize? (2 marks)
c. Specify the steps of the K-Means algorithm. (5 marks)
d. Suppose that we want to build a neural network that classifies two dimensional data (i.e., X = [x1, x2]) into two classes: diamonds and crosses. We have a set of training data that is plotted as follows:



Draw a network that can solve this classification problem. Justify your choice of the number of nodes and the architecture. Draw the decision boundary that your network can find on the diagram. (5 marks)

QUESTION 5 (15 MARKS)

a) What is a cost function? How is it important for learning algorithms? (3 marks)

b) With the aid of an example explain what is meant by non-parametric models. (5 marks)

c) What is the "Curse of Dimensionality?" as applied in dimensionality reduction. (2 marks)

d) Explain the Bias-Variance Tradeoff. (2 marks)

e) What are 3 ways of reducing dimensionality? (3 marks)
