B.Sc. Engg. CSE 7th Semester

04 March 2019 (Afternoon)

ISLAMIC UNIVERSITY OF TECHNOLOGY (IUT)

ORGANISATION OF ISLAMIC COOPERATION (OIC)

Department of Computer Science and Engineering (CSE)

MID SEMESTER EXAMINATION **WINTER SEMESTER, 2018-2019**

DURATION: 1 Hour 30 Minutes **FULL MARKS: 75**

CSE 4709: Machine Learning

Programmable calculators are not allowed. Do not write anything on the question paper.

There are **4 (four)** questions. Answer any **3 (three)** of them.

Figures in the right margin indicate marks.

---

1. a) According to Tom Mitchel (1998), a computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E. Identify T, P, and E for the following problems:
   i. A program that will tell which tweets will get retweets. 3
   ii. A program that will predict the traffic pattern at a busy intersection. 3

   b) Briefly explain the steps of developing a machine learning application with a real life example. 14

   c) What do you mean by feature vector? Suppose in a two class problem the feature vectors are normally distributed with a covariance matrix, $\Sigma = \begin{bmatrix} 1.2 & 0.4 \\ 0.4 & 1.8 \end{bmatrix}$. The mean vectors of the classes are, $\mu_1 = [0.5, 0.5]^T$ and $\mu_2 = [1.1, 1.5]^T$. The Mahalanobis distance from a vector x to the class mean is given by the equation, $d_i^2 = (x - \mu_i)^T \Sigma^{-1} (x - \mu_i)$. Classify the test point, $x = [1, 1]^T$ using Mahalanobis distance. [$\Sigma^{-1} = \begin{bmatrix} 0.9 & -0.2 \\ -0.2 & 0.6 \end{bmatrix}$ is given for your convenience.] 1+4

2. a) Consider a linear regression problem $y = w_1 x + w_0$, with a training set having m examples $(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)$. Suppose that we wish to minimize the mean *fifth* degree error (loss function) given by:

$$Loss = \frac{1}{m} \sum_{i=1}^{m} (y_i - w_1 x_i - w_0)^5$$

   i. Derive the equation to calculate the gradient with respect to the parameters w_1 and w_0. 6
   ii. Write the pseudo-code of the gradient descent algorithm for this problem. 6

   b) Consider the following set of points: $\{(-2, -1), (1, 1), (3, 2)\}$
   i. Find the least square regression line for the given data points. 4
   ii. Plot the given points and the regression line in the same rectangular system of axes. 4

   c) What is the use of basis function in linear regression? 5

3. a) What is overfitting problem? How does regularization solve the overfitting problem? Explain with example. [Hint: Ridge regression] 3+5

   b) What is logistic function? Why do you need to use logistic function in linear regression? Explain with example. 2+5

   c) Explain the concept of bias-variance trade-off. What will be the effect on bias and variance if we regularize the weights in linear/logistic regression model? Explain in brief. 5+5

4. a) Consider the dataset in Table 1. *Grade*, *Bumpiness* and *Speed-limit* are the features and *Speed* is label.

Table 1: Dataset for decision tree

SN	Grade	Bumpiness	Speed-limit	Speed
1	steep	bumpiness	yes	slow
2	steep	smooth	yes	slow
3	flat	bumpiness	no	fast
4	steep	smooth	no	fast

Answer the followings:

i. Determine the entropy of **Speed.** 4

ii. Which attribute should be selected as a root of the decision tree? 3

iii. Construct the decision tree for this dataset based on information gain. 6

b) What do you mean by clustering? Consider the following sample points, $A\ (1,1), B\ (2,-2), C\ (2,3), D\ (3,3)$. Perform k-means clustering, show the calculation of distance matrix and group assignment matrix for two epochs only. [Assume k=2.] 2+10