STAMFORD UNIVERSITY BANGLADESH

Department of Computer Science and Engineering

Final Exam, Summer 2022 Semester

MCE 642 : Data Mining

CT: Dr. Md. Tauhid Bin Iqbal

Date and Time: 11/12/2022 and 08:00 PM ~ 10.30 PM

Batch: MCE-S-MIXED **Campus:** Siddheswari

Duration: 2 hours and 30 mins **Full Marks:** 40

---

*(There are **FOUR** questions. Answer **ALL**. Numbers in the right indicate corresponding marks.)*

1 (a) Perform min-max normalization within a new range [10, 20] for the samples: [34, 21, 56, 76, 98]. [5]

(b) Use K-Means Algorithm to cluster the following four points (with (x, y) representing locations) into two-clusters: [5]
A(1,1), B(1,2), C(3,4), D(4,5).

Note: Consider the initial cluster-centroids as A and B. Also, use the Eucledian distance as the distance measure.

2 (a) For the following given transaction dataset, generate rules using Apriori-algorithm. Consider the values as Support= 30%, confidence=65%. [5]

Transaction ID	Item Purchased
1	A, B, E
2	B, D
3	A, B, D
4	B, C
5	A, B, C, E

(b) Following table represents the sample-values of a Feature X (in the left column) for a two-class problem (yes/no). [5]

Feature-X	Class
High	No
High	No
High	Yes
Medium	Yes
Low	Yes
Low	No
Low	Yes
Medium	No
Low	Yes
Medium	Yes
Medium	Yes
Medium	Yes
High	Yes
Medium	No

Calculate Information Gain and Gain-ratio for the given Feature-X.

3. (a) What are the issues of K-fold cross validation? Which method solves the issues, and how? Explain graphically, if needed. [5]

(b) Suppose, you got the following class-predictions for 10 test data (A~J). The outputs are generated for a 2-class problem (1: positive class, 0: negative class). [5]

Data	A	B	C	D	E	F	G	H	I	J
Actual Level	1	0	1	1	1	0	0	1	1	0
Predicted level	1	0	0	1	0	1	0	1	1	0

i) From the given outputs, construct a confusion matrix comprising True-positives, True-negatives, False-positives ad False-negatives.

ii) Calculate *Accuracy* and *F1-score*.

4.

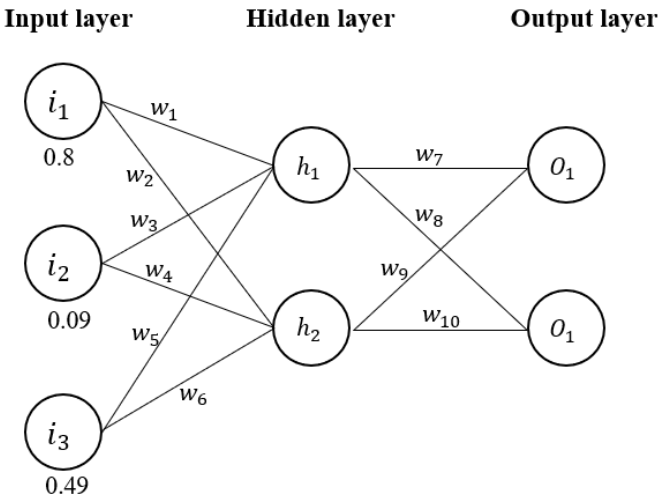

Figure 1: A Neural Network representation

Fig. 1 represents a simple Neural Network with one hidden layer (with two nodes h_1 and h_2) and an output layer having two output nodes O_1 and O_2. The weights for this network are given as:

$w_1 = 0.04$ $w_2 = 0.5$ $w_3 = 0.07$ $w_4 = 0.63$ $w_5 = 0.89$

$w_6 = 0.12$ $w_7 = 0.34$ $w_8 = 0.03$ $w_9 = 0.44$ $w_{10} = 0.59$

(a) Conduct a forward-pass to calculate the values of h_1, h_2, O_1 and O_2. Use the sigmoid activation function for the calculation when needed. [5]
The formula for sigmoid is given as: $\frac{1}{1+e^{-x}}$, where x is the input.

(b) Using the backpropagation technique, update the weight of $\boldsymbol{w_7}$. Use the information you got after the forward-pass you have conducted in Q.4(a). [5]

Note: Use the following given values for calculation:
true level for $O_1 = 0.8$ and $O_2 = 0.2$; learning rate=0.5