



## Final Assessment Test – April 2023

Course: CSE4020 - Machine Learning

Class NBR(s): 3108/3113

Time: Three Hours

Slot: B1+TB1

Max. Marks: 100

**KEEPING MOBILE PHONE/SMART WATCH, EVEN IN 'OFF' POSITION IS TREATED AS EXAM MALPRACTICE**

**Answer ALL Questions**

**(10 X 10 = 100 Marks)**

1. Consider the hypothesis space defined over these instances; Provide a hand trace of the candidate elimination algorithm learning from the Loan approval prediction dataset shown in the following table.

In particular, Show the specific and general hypothesis for class attribute of the version space training example.

Our Data: Loan Approval Prediction

Target Attribute

ID	AGE	JOB_STATUS	OWNS_HOUSE	CREDIT_RATING	CLASS (Yes or No)
1	Young	False	False	Fair	No
2	Young	False	False	Good	No
3	Young	True	False	Good	Yes
4	Young	True	True	Fair	Yes
5	Young	False	False	Fair	No
6	Middle	False	False	Fair	No
7	Middle	False	False	Good	No
8	Middle	True	True	Good	Yes
9	Middle	False	True	Excellent	Yes
10	Middle	False	True	Excellent	Yes
11	Old	False	True	Excellent	Yes
12	Old	False	True	Good	Yes
13	Old	True	False	Good	Yes
14	Old	True	False	Excellent	Yes
15	Old	False	False	Fair	No

2. Consider the Sunburn training dataset given in the table with Result Sunburned (positive) and None (Negative) be the class attribute. Construct a decision tree from the given data based on ID3 algorithm with entropy and Information Gain measure.

Name	Hair	Height	Weight	Lotion	Result
Sarah	Blonde	Average	Light	No	Sunburned
Dana	Blonde	Tall	Average	Yes	None
Alex	Brown	Short	Average	Yes	None
Annie	Blonde	Short	Average	No	Sunburned
Emily	Red	Average	Heavy	No	Sunburned
Pete	Brown	Tall	Heavy	No	None
John	Brown	Average	Heavy	No	None
Kate	Blonde	Short	Light	Yes	None

Further classify the following tuple to specify outcome as Sunburned/None for Result attribute.

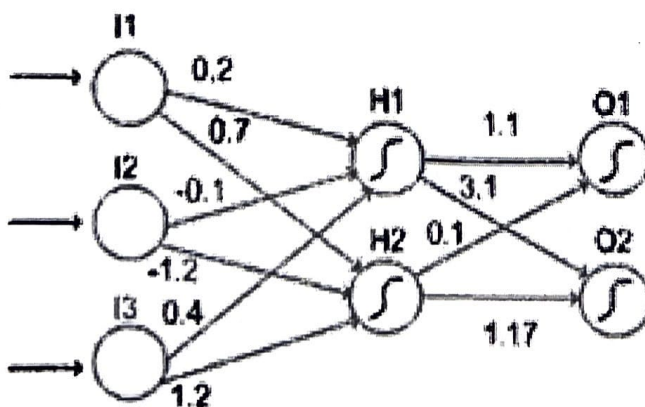
Name	Hair	Height	Lotion	Result
Mary	Red	Tall	Yes	???

3. Many substances that can burn (such as gasoline and alcohol) have a chemical structure based on carbon atoms; for this reason they are called hydrocarbons. A chemist wants to understand how the number of carbon atoms in a molecule affects how much energy is released when that molecule combusts (meaning that it is burned). The chemist obtains the dataset below. In the column on the right, "kJ/mol" is the unit measuring the amount of energy released

Name of molecule	Number of hydrocarbons in molecule (x)	Heat release when burned (kJ/mol) (y)
methane	1	-890
ethene	2	-1411
ethane	2	-1560
propane	3	-2220
cyclopropane	3	-2091
butane	4	-2878
pentane	5	-3537
benzene	6	-3268
cyclohexane	6	-3920
hexane	6	-4163
octane	8	-5471
naphthalene	10	-5157

Use the linear regression to estimate the amount of energy released (y) as a function of the number of carbon atoms (x).

4. a) Classify the instance (15, 20, 35) using Multi-Layer Perceptron model by considering the weights as given below in the neural network model. [8]



- b) Why XOR problem could not be solved by a simple perceptron?

[2]

5. Consider the following training data.

Class	X1	X2
+	1	1
+	2	2
+	2	1
-	0	0
-	1	0
-	0	1

- a) Plot these six training points. Are the classes {+, -} linearly separable or not? [2]
- b) Identify the proper support vector and find the optimal hyper plane using the appropriate SVM technique. [8]

6. Perform the Adaboost for the following dataset and classifier.

X	Y	TARGET
-0.55	0.85	+1
-0.60	0.70	-1
-0.45	0.75	-1
-0.35	-0.2	-1
0.3	0.2	+1
0.2	-0.3	+1

CLASSIFIER/STUMP
X > -0.84
X > -0.19
X < -0.62
X < 0.35
X > 0.35
X > -0.61

Execute the algorithm for 3 epochs.



7. Consider the following Coin toss problem in which identities of the coins are unknown in each toss.

H	T	T	H	H	T	H	H	T	T
T	T	H	H	T	H	H	H	H	H
H	H	H	T	T	T	T	T	T	H
H	H	H	T	H	T	H	H	T	T
T	T	T	T	H	H	H	T	T	H

Calculate the probability of getting head for coin A and coin B  $\phi_A$  and  $\phi_B$  using EM algorithm. Assume the initial probability of  $\phi_A = 0.6$  and  $\phi_B = 0.4$ .

8. Cluster the following eight points:  $x_1=(2,3)$ ,  $x_2=(5,6)$ ,  $x_3=(8,7)$ ,  $x_4=(1,4)$ ,  $x_5=(2,2)$ ,  $x_6=(6,7)$ ,  $x_7=(3,4)$  and  $x_8=(8,6)$ . Suppose, you assigned  $x_1$  and  $x_2$  as initial cluster centres for K-means clustering ( $k=2$ ). Using K-means with the Euclidean distance, compute the two clusters for each round of the algorithm until convergence.
9. Consider the two dimensional patterns  $(2.5, 2.4)$ ,  $(0.5, 0.7)$ ,  $(2.2, 2.9)$ ,  $(1.9, 2.2)$ ,  $(3.1, 3.0)$ ,  $(2.3, 2.7)$ ,  $(2, 1.6)$ ,  $(1, 1.1)$ ,  $(1.5, 1.6)$ ,  $(1.1, 0.9)$ . Use the PCA Algorithm to reduce the dimension from 2 to 1.
10. a) Explain in detail about various types of Feature selection approaches. [5]  
 b) Enumerate the various Metrics used for evaluating the Machine Learning models with an example. [5]

