B.Sc. Engg. CSE 5th Semester (35)

22 May 2017

ISLAMIC UNIVERSITY OF TECHNOLOGY (IUT)
ORGANISATION OF ISLAMIC COOPERATION (OIC)
Department of Computer Science and Engineering (CSE)

SEMESTER FINAL EXAMINATION **WINTER SEMESTER 2016-2017**

DURATION: 3 Hours **FULL MARKS: 150**

CSE 4541: Machine Learning

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

There are **8 (eight)** questions. Answer any **6 (six)** of them.

Figures in the right margin indicate marks.

---

1. a) The following diagram shows training data for a binary concept where positive examples are denoted by a heart. Also shown are three decision stumps (A, B and C) each of which consists of a linear decision boundary. Suppose that *AdaBoost* chooses A as the first stump in an ensemble and it has to decide between B and C as the next stump. Which will it choose? Explain why. 5

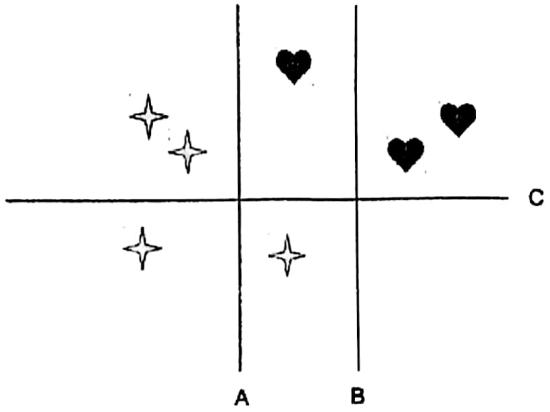


Figure 1: Data distribution for question 1(a)

b) Imagine your hypothesis is not one rectangle but a union of two (or m > 1) rectangles like the following diagram. What is the advantage of such hypothesis? What could be scenario in the worst case? 5

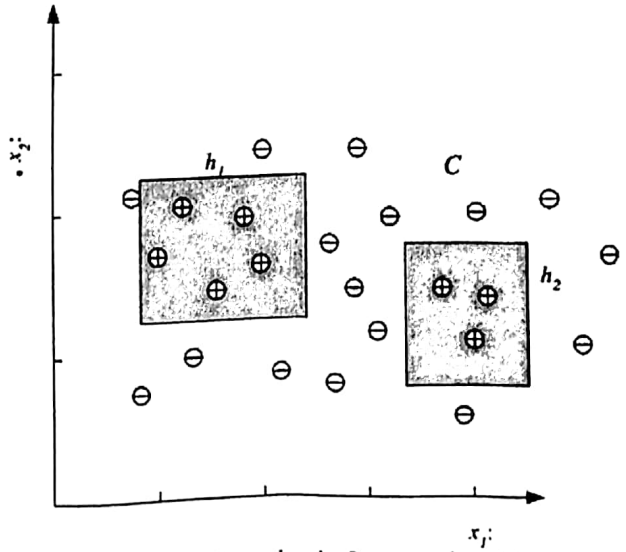


Figure 2: Hypothesis for question 1(b)

c) Suppose *Classifier A* has 90% accuracy on the training set and 75% accuracy on the test set. *Classifier B* has 78% accuracy on both the training and test sets. Which classifier is better? Explain why. 5

d) Consider the three examples given below in the order in which they appear. 10

Table 1: Table for Question 1(d)

Example	Instance	Class
1	(Large, Red, Circle)	negative
2	(Small, Red, Circle)	positive
3	(Small, Green, Hexagon)	negative

i. Run the *Candidate-Elimination* algorithm on the above training examples and generate the sequence of ***S*** and ***G*** boundaries step by step.

ii. If we want to minimize the version space of the *Candidate-Elimination* algorithm, then what is the optimal way? Explain.

2. a) NASA wants to be able to discriminate between Martians (M) and Humans (H) based on the following characteristics: Green $\in \{N, Y\}$, Legs $\in \{2, 3\}$, Height $\in \{S, T\}$, Smelly $\in \{N, Y\}$. Our available training data is as follows. 16

Table 2: Table for Question 2(a)

	Species	Green	Legs	Height	Smelly
1	M	N	3	S	Y
2	M	Y	2	T	N
3	M	Y	3	T	N
4	M	N	2	S	Y
5	M	Y	3	T	N
6	H	N	2	T	Y
7	H	N	2	S	N
8	H	N	2	T	N
9	H	Y	2	S	N
10	H	N	2	T	Y

Learn a decision tree using the *ID3* algorithm and draw the tree.

b) Imagine that you are given the following set of training examples. Each feature can take on one of three nominal values: a, b, or c. 9

Table 3: Table for Question 2(b)

F1	**F2**	**F3**	**Category**
a	c	a	+
c	a	c	+
a	a	c	-
b	c	a	-
c	c	b	-

How would a *Naive Bayes* system classify the test example *[F1 = a, F2 = c, F3 = b]*? Show your classification step by step.

3. a) A Travel Agent has to suggest spot to the customers that will best suit their interests. These interests are the attributes on which a spot will be evaluated. They may be different for every customer. In your group most of you are youngsters. You would like a place full of adventures. But your parents are also coming along. It will be great if there is something less adventures as well. Most of you are non-vegetarians but there are a few veggies. A place with more adventures and more non vegetarian food. Major concerns are very well addressed but minor ones are left out. Most of the people from group will be Happy but some will be extremely unhappy. But the Travel Agent wants everyone to be Happy to some degree. 12

So design a legitimate *Fuzzy logic system* describing the fuzzifier, rules, inference engine, and defuzzifier for the travel agent.

b) What are the difference between the *Traditional* learning method and *Ensemble* method for classification? Briefly describe *Bagging*, *Boosting* and *Radom forests*. 8

c) What is the problem of hard margin in the design of *SVM*? How can we solve it? Explain mathematically. 5

4. a) Let us analyze the following 3-variate dataset with 10 observations in the following table. Each observation consists of 3 measurements on a wafer: thickness, horizontal displacement, and vertical displacement. 21

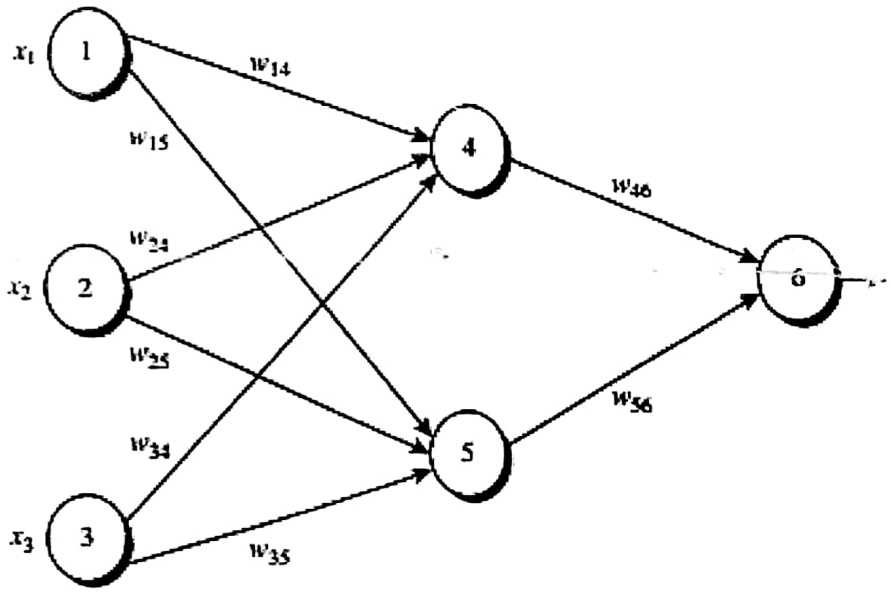
Table 4: Table for question 4(a)

Thickness	7	4	6	8	8	7	5	9	7	8
Horizontal displacement	4	1	3	6	5	2	3	5	4	2
Vertical displacement	3	8	5	1	7	9	3	8	5	2

Perform the *PCA* step by step on the above dataset and generate the final version of the dataset.

b) Write down the differences between *PCA* and *SVD*. 4

5. a) Following figure shows a multilayer feed-forward neural network. Let the learning rate be 0.9. The initial weight and bias values of the network are given, along with the first training tuple, $X = (1, 0, 1)$, with a class label of 1. 14



Initial Input, Weight, and Bias Values

x_1	x_2	x_3	w_{14}	w_{15}	w_{24}	w_{25}	w_{34}	w_{35}	w_{46}	w_{56}	θ_4	θ_5	θ_6
1	0	1	0.2	−0.3	0.4	0.1	−0.5	0.2	−0.3	−0.2	−0.4	0.2	0.1

Figure 3: Neural Network for question 5(a)

Perform the *multilayer feed-forward neural network (Back-propagation)* algorithm step by step and update all the corresponding values for at least two iterations. 8

b) Define the following terms:
i. Perceptron, ii. Sigmoid function, iii. Feed-forward network,
iv. Feed-back network 3

c) Weights are modified at each step according to the Perceptron training rule, which revises the weight w_i associated with input x_i according to the rule

$$w_i \leftarrow w_i + \Delta w_i$$
$$\Delta w_i = \eta\,(t - o)\,x_i$$

Why should this update rule converge toward successful weight values?

6. a) You are going to spend a month in the wilderness. You're taking a backpack with you, however, the maximum weight it can carry is 20 kilograms. You have a number of survival items available, each with its own number of "survival points". Your objective is to maximize the number of survival points. Your items are as follows. 12

Table 5: Table for question 6(a)

Item	Survival Point	Weight
pocketknife	10.00	1.00
beans	20.00	5.00
potatoes	15.00	10.00
unions	2.00	1.00
sleeping bag	30.00	7.00
rope	10.00	5.00
compass	30.00	1.00

Perform the *Genetic algorithm* using two point crossover and one point mutation for each pair set and select the best fitted sample using at least two iterations.

b) A budget airline company operates 3 planes and employs 5 cabin crews. Only one crew can operate on any planes on a single day, and each crew cannot work for more than two days in a row. The company uses all planes every day. A Genetic Algorithm is used to work out the best combination of crews on any particular day. Suggest what chromosome, alphabet and fitness function can be used represent this algorithm? Is it necessary to use Genetic Algorithms for solving it? Explain. 8

c) What is *Genetic Programming*? Define the *Terminal* and *Primitive* functions in case of GP with suitable example. 5

7. a) Discuss about the *Partitioning, Hierarchical, Density-based* and *Grid-based* clustering methods in a comparative manner. 8

b) Discuss about the following terms in case of clustering:
i. PAM, ii. BIRCH, iii. AGNES, iv) DIANA 12

c) How does *Chameleon* dynamic clustering model works? Briefly discuss with figures. 5

8. a) To compare the capabilities of some popular clustering and classification algorithms, provide sample datasets that cannot be dealt with accurately by one algorithm but by the other. In every case, explain why one of the algorithms fails to discover the correct clusters or classes. 20

i. Draw a 2-dimensional dataset with two clusters that can be discovered with 100% accuracy by *DBSCAN*, but not by *k-means*.
ii. Draw a 2-dimensional dataset with two clusters that can be discovered with 100% accuracy by *k-means*, but not by *DBSCAN*.
iii. Draw a 2-dimensional dataset with two classes that can be classified with 100% accuracy (on the training dataset) by a *decision tree*, but not by a *linear SVM*.
iv. Draw a 2-dimensional dataset with two classes that can be classified with 100% accuracy (on the training dataset) by a *linear SVM*, but not by a *3-Nearest Neighbor* classifier.

b) "Which method is more robust for clustering—*k-means* or *k-medoids*?" Discuss in terms of noise and complexity. 5