Final Assessment Test – April 2018

VIT
Vellore Institute of Technology

Course: CSE4020 - Machine Learning

Class NBR(s): 3796 / 3799 / 3802 / 3805 / 4279 Slot: E1

Time: Three Hours Max. Marks: 100

Students are permitted to use Scientific Calculator

Answer ALL Questions
(10 X 10 = 100 Marks)

1. a) Write three computer applications for which machine learning approaches seem appropriate and two applications for which they seem inappropriate. Include a one-sentence justification for each. [5]

   b) Prove the polynomial complexity of PAC-learnable problems. [5]

2. Compute the version space containing all hypotheses that are consistent with an observed sequence of training examples given below, using Candidate Elimination algorithm. [Note: Step by step solution is expected]

Table 1: Likes Database

hair	*body*	*likesSimon*	*pose*	*smile*	*smart*
blond	thin	yes	arrogant	toothy	no
brown	thin	no	natural	pleasant	yes
blond	plump	yes	goofy	pleasant	no
black	thin	no	arrogant	none	no
blond	plump	no	natural	toothy	yes

3. The values of x and the corresponding values of y are shown in the table below.

x	0	1	2	3	4
y	2	3	5	4	6

Find the regression line $y = ax + b$ using gradient descent algorithm with four iterations and 0.1 as the learning rate.

4. a) Assume our hypothesis class as the set of intervals on the real number line. H is the set of hypotheses of the form $a < x < b$, where a and b may be any real constants. Compute its VC dimension. [3]

   b) Suppose that we take a data set, divide it into equally-sized training and test sets, and then try out two different classification procedures. First we use logistic regression and get an error rate of 20% on the training data and 30% on the test data. Next we use 1-nearest neighbors (i.e. K=1) and get an average error rate (averaged over both test and training data sets) of 18%. Based on these results, which method should you prefer to use for classification of new observations? Justify your answer. [3]

   c) Assume that the data is collected from a group of students in a statistics class with variables X_1= hours studied, X_2= undergrad GPA, and Y= receive an A grade. A logistic regression is fit which produced estimated coefficients, β_0=−6, β_1=0.05, and β_2=1 [4]

      i. Estimate the probability that a student who studies for 40 hours and has an undergrad GPA of 3.5 gets an A grade in the class.

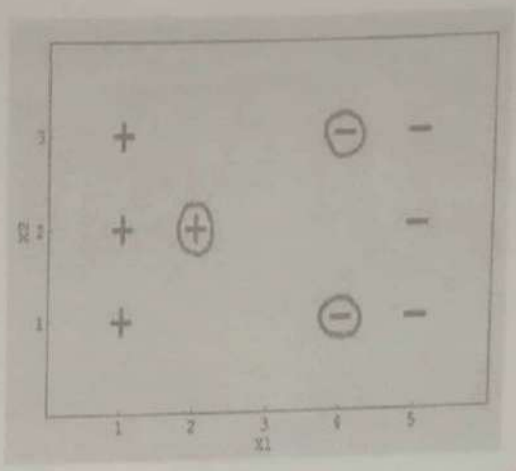
      ii. How many hours would the student need to study to have a 50% chance ofice in the class?

5. a) Suppose you are using a Linear SVM classifier with 2 class classification problem. Now you have been given the following data in which some points are circled. Will the decision boundary change? [3]

   i. If the circled points are removed from the data?

   ii. If the non - circled points are removed from the data?

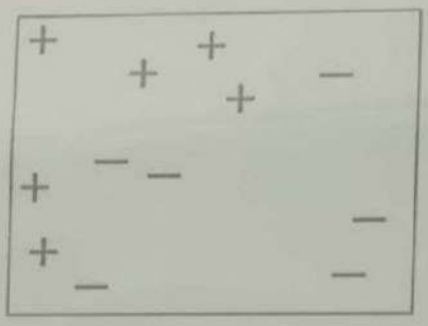
   Justify your answer.



   b) Give a note on general purpose kernel functions and application specific kernels with examples. [3]

   c) Bagging is the first effective method of ensemble learning; explain how bagging uses averaging and voting to solve classification and regression problems. [4]

6. a) Use AdaBoost algorithm with decision stumps to classify the samples given below. Present how the week classifier in each step with misclassification errors results in a strong classifier. [6]



   b) Suppose we have height, weight and T-shirt size of some customers and we need to predict the T-shirt size of a new customer given only height and weight information we have. Data including height, weight and T-shirt size (class attribute) information is shown below. [4]

	Height (in cms)	Weight (in kgs)	T Shirt Size
1	158	59	M
2	158	63	M
3	160	59	M
4	160	60	M
5	163	60	M
6	163	61	M
7	160	64	L
8	163	64	L
9	165	61	L
10	165	62	L
11	165	65	L
12	168	62	L

What T shirt size would 5-NN model using Euclidean distance return for a new customer of height 161 cm and weight 61 kg?

7 a) How do you user self-organizing map to perform clustering?

b) Partition the given data into two clusters using k-means algorithm. Choose the centres as the first two points.

Data: (2, 2), (4, 4), (5, 5), (6, 6), (9, 9), (0, 4), (4, 0)

8 a) Given below is an example of the output of a factor analysis looking at indicators of wealth, with six variables and two resulting factors. Comment on the association of these variables with the factors. [3]

Variables	Factor 1	Factor 2
Income	0.65	0.11
Education	0.59	0.25
Occupation	0.48	0.19
House value	0.38	0.60
Number of public parks in neighborhood	0.13	0.57
Number of violent crimes per year in neighborhood	0.23	0.55

b) Explain how Principal Components Analysis is used for dimensionality reduction. [7]

9 a) Propose a suitable test to compare the errors of two regression algorithms? [2]

b) Given two learning algorithms, describe two methods to compare and test whether they construct classifiers that have the same expected error rate. [8]

10 a) How does Deep learning techniques influence machine learning algorithms? [8]

b) Describe four applications of Deep learning.