17P Register Number:

Question Paper Code: 18615PD32C

KARPAGAM COLLEGE OF ENGINEERING
(Autonomous)
Coimbatore – 32
END SEMESTER EXAMINATIONS – MARCH 2019
B.E /B.TECH
15PD32/15LD32/15ED32/15ND32/15TD32/15MD32/15AD32/15FD32 – MACHINE LEARNING TECHNIQUES

Time: 3 hrs

Answer ALL questions **Max. Marks: 100**

PART- A **(10x2 = 20 marks)**

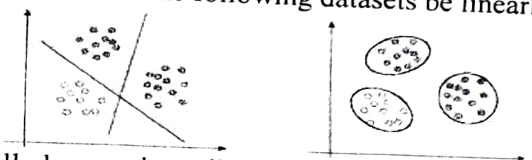
1. Given the following test scores, calculate the Sample Mean and sum of squared differences.
Scores: { 60, 70, 80, 90, 100, 70, 80, 90, 75, 85, 80 }.
2. The following code fragment is used to find the eigen values and eigen vectors of a matrix. Fill in the blanks to complete the code.

```
from numpy import array
from numpy_____ import _____
A = array([[1, 2, 3], [4, 5, 6], [7, 8, 9]])
print(A)
values, vectors = _____
print(values)
print(vectors)
```

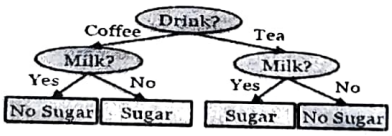
3. Given the following table of observations, calculate the information gain IG(Y|X) that would result from learning the value of X.

X	Y
Red	True
Green	False
Brown	False
Brown	True

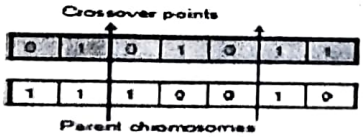
4. What approachs can be used to make the following datasets be linearly separable?



5. Is a K-nearest neighbor will always give a linear decision boundary? Justify your answer.
6. Construct the possible rules from the following decision tree.



7. Draw the offspring chromosomes after performing 2-point cross over on the following parent Chromosomes.



8. Say True or False : Zero correlation between any two random variables implies that the two random variables are independent.
9. Write the properties of Linear Discriminant Analysis.
10. Fill in the blanks:
Following code fragment is the partially filled implementation of Linear Congruential Generator:

```
def lcg(x0,n):
a = 23
m = 162
c = 0
rnd = np.zeros((n))
rnd[0] = np.mod(_____*_____ + _____,_____)
for i in range(_____,_____):
rnd[i] = np.mod(a*_____+c,_____)
```

2 c

PART - B

(5x16 = 80 marks)

11. a) Given that a loaded coin has the following probability for coming up heads: P(Loaded coin: heads) 0.25

i) What is the probability that the loaded coin will come up tails

ii) Train the perceptron model by using the following data points and explain

sample_id	x(1)	x(2)	y
1	0	0	-1
2	1	0	1
3	0	1	1
4	1	1	1

(OR)

b) i) Perform Linear Regression on the following dataset and find the coefficients.

x	y
1	0.5
2	1
4	2
0	0

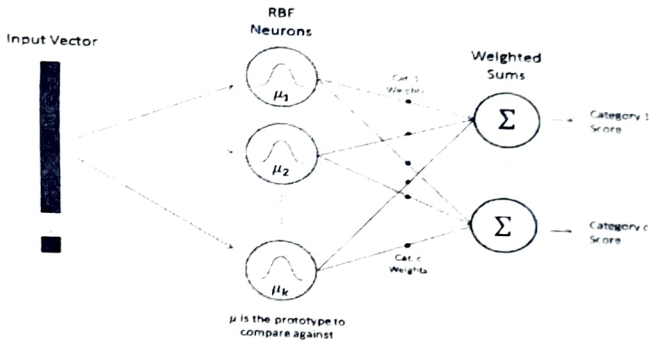
ii) Write the need for bias and learning rate in perceptron model. Explain with suitable examp

12. a) i) Write Back-propagation procedure for Multilayer Perceptron (MLP) and explain in detail

ii) The Back-Propagation learning algorithm for training feed-forward neural networks requires activation and error functions to be differentiable. What is the meaning of the above sentence why it is true?

(OR)

b) Apply the NOR-Logic function for the classification of the target value in the following network and explain briefly.



13. a) i) Given the matrix whose rows represent different data points, you are asked to perform a K-Means clustering on this dataset using the Euclidean distance as the distance function. Here is chosen as 2.

	Driver_ID	Distance_Feature	Speeding_Feature
0	3423311935	71.24	28
1	3423313212	52.53	25
2	3423313724	64.54	27
3	3423311373	55.69	22
4	3423310999	54.58	25

ii) List the various feature selection methods used in Machine Learning.

(OR)

b) The following dataset will be used to learn a decision tree for predicting if people have risk buying car (High) or not (Low), based on their machine learning Age (Young, Middle_A Old) and Car Type (Family, Sports, Truck).

Tid	Age	Car Type	Class
0	Young	Family	High
1	Young	Sports	High
2	Middle_Aged	Sports	High
3	Old	Family	Low
4	Middle_Aged	Truck	Low
5	Young	Family	High

i. What is the entropy H(Class)? Briefly justify

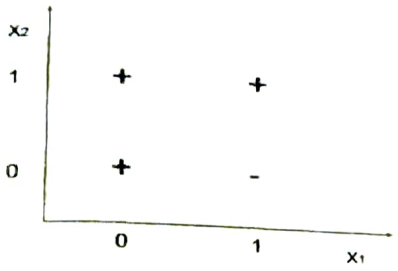
ii. What is the entropy H(Class | Age = Young)? Briefly justify
iii. Draw the full decision tree that would be learned for this data

(ii) Consider, we have data from the questionnaires survey (to ask people opinion) and objective testing with two attributes (acid durability and strength) to classify whether a special paper tissue is good or not. Here is four training samples. (10)

Acid Durability(sec)	Strength (kg/square meter)	Classification
7	7	Bad
7	4	Bad
3	4	Good
1	4	Good

Now the factory produces a news paper tissue that pass laboratory test with Acid Durability=3 and Strength=7. Without another expensive survey, can you guess what the classification of this new tissue is by using K-NN approach? Consider K=3.

(ii) A set of reasonably clean sample records was extracted by Barry Becker from the 1994 Census database. We are interested in predicting whether a person makes over 50K a year. For simplicity suppose we model the two features with two Boolean variables $X_1, X_2 \in \{0,1\}$ and label $Y \in \{0,1\}$ where Y = 1 indicates a person makes over 50K. In the following figure we show three positive samples ("+" for Y = 1) and one negative samples ("-" for Y = 0). Please complete the following questions. (6)



I) If we train a KNN classifier (K=1) based on data in Figure 1, and then try to classify the same data. Which sample(s) must be misclassified by this classifier?

II) Suppose we have trained a linear regression model y = ax+b where a = 0.5 and b = 1.0. on a set of training data points D = {(1.0,1.6), (1.5,1.5), (3.0;2.4)}. Calculate the mean squared errors of this model on D.

(OR)

b) Write a python code to Train a following Machine Learning Model using genetic algorithm with its suitable operators. (16)

Data:

x_1	x_2	x_3	x_4	x_5	x_6	y
6	-2	7	3	11	2	45

Machine Learning Model:

$$\sum_i w_i x_i$$

a) Given the following 2D input data, identify the principal component. (16)

x	y
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

(OR)

(4)

i) Write the implementation of Linear Congruential Generator.

ii) If P(A) = 0.3, P(B) = 0.5, and P(A|B) = 0.6 (4)

I) Compute $P(A \cap B)$ (8)

II) Compute P(B|A).

i) Explain in detail about Reinforcement Learning.
