Roll No											

PRESIDENCY UNIVERSITY
BENGALURU

SET-A

SCHOOL OF ENGINEERING

END TERM EXAMINATION –MAY/JUNE 2024

Semester : Semester VI - B.Tech CSE - 2021
Course Code : CSE3009
Course Name : Optimization Technique for Machine
Program : B.Tech. Computer Science and Engineering

Date : JUNE 06-2024
Time : 01:00PM-04:00PM
Max Marks :100
Weightage : 50%

Note:
1. **Answer ALL 5 FULL Questions.**
2. **Each Full Question carries 20 Marks**
3. **Scientific and non-programmable calculator are permitted.**
4. **Do not write any information on the question paper other than Roll Number.**

1.a. List Data Preprocessing Techniques. **[Knowledge]** **(CO1)** **(04 Marks)**

1.b. Explain the properties of VC – Dimension. **[Comprehension]** **(CO1)** **(06 Marks)**

1.c. Calculate the entropy and Weighted Average Gini Index for the attribute "**Outlook**" in a decision tree for the simple dataset shown below: **[Application]** **(CO1)** **(10 Marks)**

Outlook	Temperature	Humidity	Windy	Play Tennis
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

The target variable is Play Tennis, which can be either Yes or No.

or

2.a. Define VC-Dimension. **[Knowledge]** **(CO1)** **(04 Marks)**

2.b. Explain Machine Learning Techniques **[Comprehension]** **(CO1) (06 Marks)**

2.c. Calculate the entropy and weighted average gini index for the attribute "**Temperature**" in a decision tree for the simple dataset shown below: **[Application]** **(CO1) (10 Marks)**

Outlook	Temperature	Humidity	Windy	Play Tennis
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

The target variable is Humidity, which can be either High or Normal.

3.a. import numpy as np **[Knowledge]** **(CO2) (04 Marks)**

```
# Actual house prices (in $1000s)
y_true = np.array([150, 200, 250, 300, 350])

# Predicted house prices (in $1000s)
y_pred = np.array([160, 210, 240, 310, 340])

x = np.mean((y_true - y_pred) ** 2)
print (f"loss function: {x}")
```

Identify the suitable loss function the above piece of code is used for.

3.b. Explain the framework of multiple kernel learning. **[Comprehension]** **(CO2) (06 Marks)**

3.c. Calculate the Empirical Risk for **[Application]** **(CO2) (10 Marks)**

1) Hypothesis: h(x)=wx+b
2) Loss Function: Mean Squared Error(MSE)
3)Dataset: a single feature House Size (sq. ft) and a target variable Price ($):

House Size(sq ft)	Price($)
800	150,000
1000	200,000
1200	250,000
1500	300,000
1800	350,000

or

4.a.	List the specific requirements of the task, which will help to choose the type of loss function. **[Knowledge]**	**(CO2)**	**(04 Marks)**
4.b.	Explain Singular Value Decomposition. **[Comprehension]**	**(CO2)**	**(06 Marks)**
4.c.	Cluster a simple 2D dataset into k=3 clusters with 3 initial centroids as (2,3), (6,6), and (8,8) until 2iterations for the following points: **[Application]** {(2,3),(3,3),(6,6),(8,8),(7,7),(5,5),(6,4),(3,8),(8,3),(2,2)}	**(CO2)**	**(10 Marks)**
5.a.	List the applications of Low Rank Matrix Factorization. **[Knowledge]**	**(CO3)**	**(04 Marks)**
5.b.	Explain the properties of Convex Quadratic Optimization. **[Comprehension]**	**(CO3)**	**(06 Marks)**
5.c.	Maximize $Z=5x+4y$ Subject to constraints: **[Application]** i) $x+2y\leq 10$ ii) $3x+2y\leq 12$ iii) $x,y\geq 0$	**(CO3)**	**(10 Marks)**
	or		
6.a.	List Sparse Regression Techniques. **[Knowledge]**	**(CO3)**	**(04 Marks)**
6.b.	Explain the challenges with semi-definite optimization. **[Comprehension]**	**(CO3)**	**(06 Marks)**
6.c.	Minimize $Z=6x+5y$ Subject to constraints: **[Application]** i) $x+2y\leq 10$ ii) $2x+2y\leq 12$ iii) $x,y\geq 0$.	**(CO3)**	**(10 Marks)**
7.a	List the advantages of Interior Point Methods. **[Knowledge]**	**(CO4)**	**(04 Marks)**
7.b.	Explain the advantages of using Stochastic Coordinate Gradient Method. **[Comprehension]**	**(CO4)**	**(06 Marks)**
7.c	Find the root of the function $f(x)=-3x+1$ using Newton Method (4 iterations) with initial value as 0.5. **[Application]**	**(CO4)**	**(10 Marks)**
	or		
8.a	List the challenges of Stochastic Gradient Descent. **[Knowledge]**	**(CO4)**	**(04 Marks)**
8.b.	Explain the kind of problems for which Active Set Optimization is suitable. **[Comprehension]**	**(CO4)**	**(06 Marks)**
8.c	Identify a root of the function $f(x)=-2$ using Newton Method (5 iterations) with the initial value as 1. **[Application]**	**(CO4)**	**(10 Marks)**
9.a	List the types of Learning Guarantees in Machine Learning. **[Knowledge]**	**(CO1)**	**(04 Marks)**
9.b	Explain the Structural Risk Minimization models. **[Comprehension]**	**(CO1)**	**(06 Marks)**

9.c Calculate entropy and weighted average gini index for the attribute "**Outlook**" in a decision tree for the simple dataset shown below: **[Application]** **(CO1)** **(10 Marks)**

Outlook	Temperature	Humidity	Windy	PlayTennis
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

The target variable is Windy, which can be either True or False.

or

10.a List the disadvantages of SVM. **[Knowledge]** **(CO2)** **(04 Marks)**

10.b Differentiate between Simple Linear Regression and Multiple Linear Regression with examples. **[Comprehension]** **(CO2)** **(06 Marks)**

10.c Predict the missing ratings in Movie Recommendations systems dataset below using Low Rank Matrix Factorization. Predict the missing ratings in Movie Recommendations systems dataset below using Low Rank Matrix Factorization. **[Application]** **(CO2)** **(10 Marks)**

a) Dataset:

Movie 1	Movie 2		Movie 3	Movie4
User1	5	3	-	1
User2	4	-	-	1
User3	1	1	-	5
User4	1	-	-	4
User5	-	1	5	4

b) k=2c)

c) U as

$$\begin{bmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \\ 0.3 & 0.5 \\ 0.9 & 0.7 \\ 0.5 & 0.4 \end{bmatrix}$$

D.V as

$$\begin{bmatrix} 0.6 & 0.1 \\ 0.7 & 0.2 \\ 0.3 & 0.9 \\ 0.5 & 0.8 \end{bmatrix}$$