aimlcZG565
@ wilp.bits-pilani.ac.in

Format of Question paper for Midsemester Examination

Birla Institute of Technology & Science, Pilani
Work Integrated Learning Programmes Division
Second Semester 2022-2023

Comprehensive Examination
(EC-2 Regular)

Course No.	: AIMLCL ZG565
Course Title	: Machine Learning
Nature of Exam	: Open Book
Weightage	: 30%
Duration	: 2 Hours
Date of Exam	: 22nd July 2023

No. of Pages = 4
No. of Questions = 6

Note to Students:
1. Please follow all the *Instructions to Candidates* given on the cover page of the answer book.
2. All parts of a question should be answered consecutively. Each answer should start from a fresh page.
3. Assumptions made if any, should be stated clearly at the beginning of your answer.

Q.1. Study the following data and answer the following questions. **[4+1.5+1.5 = 7 Marks]**

Q1

Note: Wherever applicable use only Min-Max scaling to [0-5] range. Round all the calculation to 4 decimal places. Use default threshold >=0.5 for positive class (i.e., Capacity = High)

Product	Power Density (in W/L)	Average Temperature (in Celcius)	Energy Density (in Wh/L)	Capacity	
LI105	600	50	100	Low	Training Data
LI102	100	20	100	Low	
LI103	9	-10	5	High	
LI101	1050	45	50	High	
LI115	3	10	2	Low	
LI112	10000	70	500	Low	
LI108	5	-5	5	High	
LI109	5000	15	180	High	
LI201	200	-2	250	High	Test Data
LI210	50	60	200	Low	
LI205	25	45	10	High	
LI207	150	15	40	High	

A. Rechargeable lithium-ion batteries manufactured with variety of specification, constraints the use of it in all electronic commodities. Above are few such observations captured. Apply logistic regression with batch gradient descent for only one iteration to learn the pattern to determine the capacity of the battery based on operating setup and manufacturing properties. Use learning rate = 0.35, W_0 = 1.5, $W_{energyDensity}$ =5, $W_{powerDensity}$ = 2.5, $W_{avgTemp}$ =-2, $W^TX = W_0 + W_1 X_{powerDensity} + W_2 X^2_{avgTemp} + W_3 X_{energyDensity}$. Assuming no

2.5 -2 5

1.5

regularization is used, write the gradient update equation with new weights at the end of 1st iteration.

B. Predict the capacity level of a new battery model with 10 W/L power density, 15 Wh/L energy density and average temperature is -2 Celsius. Interpret the effect of power density on the capacity of the battery. Assume the results obtained in previous part a) is the final weights. Using this construct the confusion matrix for given test data and find the recall and precision of the positive class.

Q.2. Study the following data and answer the following questions. **[3.5+2.5 = 6 Marks]**

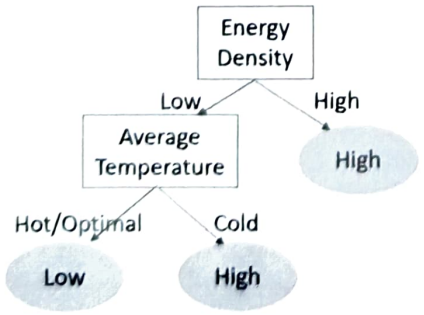
Q2

Product	Power Density (in W/L)	Average Temperature (in Celcius)	Energy Density (in Wh/L)	Capacity (Target)	
LI105	Low	Hot	Low	Low	Training Data
LI102	Low	Optimal	Low	Low	
LI103	Low	Cold	Low	High	
LI101	Low	Hot	Low	High	
LI115	Low	Optimal	Low	Low	
LI112	High	Hot	High	Low	
LI108	Low	Cold	Low	High	
LI109	High	Optimal	High	High	
LI201	High	Cold	High	High	Test Data
LI210	Low	Hot	High	Low	
LI205	Low	Hot	Low	High	
LI207	High	Optimal	High	High	

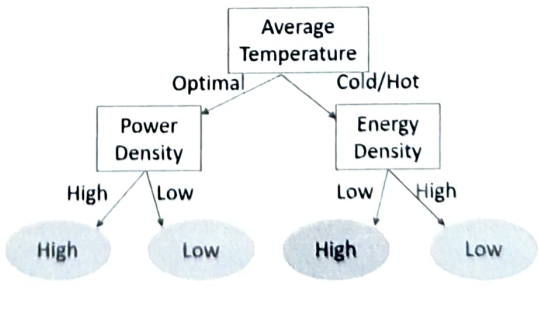
A. If the decision tree is constructed using ID3 classifier with entropy based information gain as a criterion for attribute selection, Identify the attribute used at the root. Show all the calculations involved. Draw the resultant tree after the first iteration.

B. Consider the decision trees given below in the table. Which of the trees is more suitable for the given data? Use F-Score as the decision criteria and explain with necessary computations.

Decision Tree 1



Decision Tree 2



Q.3. Suppose a logistic regression classifier σ(x,y)=1/(1+exp(-w0-5*x-5*y)) is used to classify the following dataset. What is the range of values w0 can take for 100% classification accuracy? Show the steps clearly. **[2+2=4 Marks]**

Q6

x	y	class
0	0	0
0	1	1
1	0	1
1	1	1

Q5

Q.4. Linear regression was performed on a dataset and it was found that the best least square fit was obtained by the line $y = 3x + 1$. The dataset on which regression was performed was corrupted in storage and it is known that the points are (x, y): (-2,α), (0,1), (2, β). Can we recover unique values of α, β so that the line y=3x+1 continues to be the best least square fit? If yes, find those unique values of α, β. Otherwise, give a mathematical justification for your answer.
[2 + 2 = 4 Marks]

Q3

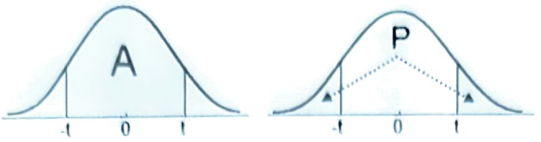
Q.5. It is known that a natural law obeys the quadratic relationship, $y = A * x^2$, where predictor x is a continuous random variable, uniformly distributed between -1 and +1 and y is a real valued response variable. What is the best linear curve of the form y=a*x+b that can be used to model this data? That is, find the optimal values of a and b in terms of A that minimize the loss function! **[2+2=4 Marks]**

Q4

Q.6. Following table shows the effect of advertising spend on revenue of a product. Assuming a linear model is used to measure the impact, determine the optimal parameters. Calculate the 90% confidence interval. With 90% confidence can we say that increase in spend will increase the revenue? Show all calculation steps, otherwise marks will not be awarded. **[2+2+1=5 Marks]**

Advertising Spend	Revenue
0.8	2.4
1.2	3.0
1.6	4.5
2.0	6.0
2.4	6.5
3.0	8.0

You can refer to the following t distribution (two sided) table.



DF	A P	0.80 0.20	0.90 0.10	0.95 0.05	0.98 0.02	0.99 0.01	0.995 0.005	0.998 0.002	0.999 0.001
1		3.078	6.314	12.706	31.820	63.657	127.321	318.309	636.619
2		1.886	2.920	4.303	6.965	9.925	14.089	22.327	31.599
3		1.638	2.353	3.182	4.541	5.841	7.453	10.215	12.924
4		1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5		1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6		1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7		1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8		1.397	1.860	2.306	2.897	3.355	3.833	4.501	5.041
9		1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10		1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587