Birla Institute of Technology & Science, Pilani
Work Integrated Learning Programmes Division
2020-2021
M.Tech. (Data Science and Engineering)
Midsem Examination (Regular Solution)

Course No.	: DSECLZG565
Course Title	: MACHINE LEARNING
Nature of Exam	: Open Book
Weightage	: 30%

No. of Pages = 4
No. of Questions = 3

Note: Assumptions made if any, should be stated clearly at the beginning of your answer.

Question 1. **[3+2+3+2=10 marks]**

A. Define task, performance measure and the training experience for a Traffic prediction problem Which type of machine learning algorithms with reasoning [**3 Marks**]

B. In MTech DSE, 55% of the students opted for Natural Language Processing (NLP) and Information Retrieval (IR). 65% of the students opted for IR. What is the probability of a student opted for NLP given she/he is already opted for IR. [**2 Marks**]
Solution
P(NLP and IR) = 0.55
P(IR) = 0.65
P(NLP|IR) = P(NLP and IR)/P(IR) = 0.55/0.65 = 2/3 = 0.846

C. Spam mail detection software claims that it can detect 96% of spam emails. 40% of Emails in the inbox are spam emails. The probability of a non-spam email detected as spam is 4%. Now if an email is detected as spam, then what is the probability that it is in fact a non-spam email?
[3 Marks]

Solution:
Define events A = event that an email is detected as spam,
S = event that an email is spam, S_c = event that an email is not spam.
We know P (S) =0.4, P (S_c) = .0.6, P (A | S) = 0.96, P (A | S_c) = 0.04
Hence by the Bayes's formula we have
P(S_c | A) = P(A | S_c) P(S_c) / (P(A | S)P(S) + P(A| S_c)P(S_c))
= (0.04 × 0.6) / (0.96 × 0.4 + 0.04 × 0.6)
= 0.024 / (0.384 + 0.024) = 0.024 / 0.408 = 0.059

D. In a small covid dataset, 39 out of 40 patients recovered successfully from covid. What is the probability (MLE) of a new person detected with covid recovering successfully? Will this estimate change if we have prior probability of successful recovery as 90%? ? **[2 Marks]**

Solution
MLE estimate = 39/40 =0.975
MAP estimate=39/40 * 0.8 = 0.78

Question 2.

A. Given a data set comprising customer feedbacks on various products and their corresponding categories in terms of the feedback:

Feedback Text	rating
durable product	Good
High cost product	Poor
Nice product features	Good
Value for money	Good
Not upto expectation	Poor

Consider a feedback text : Text_new=" Product features are amazing though cost is high" and apply Naïve Bayes classification technique to compute the probabilities, p('Good'), P('Poor'), Prob(Good | Test_new), Prob(Poor | Test_New). Which class does Text_new belongs to?

[5 Marks]

Solution:

Total number of words in Vocabulary: |V| =12; Total words in Good class: 8;
Total words in Poor class:6; P(Good) =3/5; P(Poor)=2/5

Word	P(Word|Good)	P(word|Poor)
Product	(2+1)/(8+12)=3/20	(1+1)/(6+12)=2/18
Features	(1+1)/20=2/20	(1+1)/18=2/18
are	(1/20)	(1/18)
amazing	(1/20)	(1/18)
though	(1/20)	(1/18)
cost	(1/20)	(1+1)/18=2/18
is	(1/20)	(1/18)
high	(1/20)	(1+1)/18=2/18

P(Good|New_Text) =p(New_Text|Good) * p(Good) =3/20* 2/20 * $(1/20)^6$ =$(6/20^8)$ *3/5

P(Poor|New_Text) =p(New_Text|Poor) * p(Poor) =$(8/18^8)$ * 2/5

Class =argmax {18/(20^8 *5) , 16/(18^8 *5)} = argmax { 3.6/(20^8), 3.2/(18^8)}

Clearly the text is classified as poor rating -considering denominator.

B. Consider conditional distribution of generating data that maximizes $h(w,x)$; Apply gradient ascent technique to estimate the weight update for the model parameter $w = w + \alpha \frac{\partial}{\partial w} L(w)$ where
$L(w,x) = y \log h(w,x) + (1-y) \log(1 - h(w,x)$. Here $h(w,x)$ is the logistic regression function given by $h(w,x) = (1 + e^{-wx})^{-1}$

We have $w = w + \alpha \frac{\partial}{\partial w} L(w,x) = w + \alpha \{ y \frac{\partial}{\partial w} h + (1-y) \frac{\partial}{\partial w} \log(1 - h(w,x)\}$

This yields:

$$w + \alpha \{ y\, h(1-h)[-w] + \frac{(1-y)}{[1-h(w,x)]} \{\frac{\partial}{\partial w} - h(w,x)\}$$

After simplifying we get: $w = w + \alpha\, \{y - h(w,x)\}x$

Question 3.

A. Explain if the gradient descent algorithm can be applied to non-convex problems? What would be the outcome? How does learning rate impact the algorithm? **[3 Marks]**

B. Which classifier is better and why? **[2 Marks]**

Classifier 1 training accuracy 100% and test accuracy is 75%

Classifier 2 training accuracy is 80% and test accuracy is 80%

Solution:

Classifier 2 is better since Classifier 1 tends to overfit and hence the lower test accuracy.

C. Imagine we are training a decision tree and we are at the node. Each point is (x1, x2, y) where x1, x2 are attributes and y is the label. The data at this node is

x1	**x2**	**y**
0	1	+1
1	0	-1
0	1	+1
1	1	+1

Which feature results in the best split? And Why? [1+2+2]

Solution

Entropy of the collection E(S) = -¾ log ¾ - ¼ log ¼ = 0.311+0.5= 0.811
E(x1 = 0) = - 2/2 log 2/2 - 0/2 log 0/2 = 0
E(x1 = 1) = -½ log ½ -½ log ½ = 1
Info Gain = E(S) - 2/4*E(x1=0) - 2/4*E(x1=1) = .811-0-.5 = 0.311
E(x2 = 0) = - 1/1 log 1/1 - 0 = 0
E(x2 = 1) = -3/3 log 3/3 - 0 = 0
Info Gain = E(S) - 1/4*E(x2=0) - 3/4*E(x2=1) = .811-0-0 = 0.811

We maximize for the information gain which is best in using x2.