Unique Paper Code   :   32347607

Name of the Course   :   B. Sc. (H) Computer Science DSE-3

Name of the Paper   :   Machine Learning

Semester   :   VI

Year of Admission   :   2015, 2016, 2017, 2018

Duration   :   3 Hours   Maximum Marks: 75

Instructions for candidates:

*Attempt any **FOUR** questions.*

*All questions carry **equal** marks.*

*The complete answer to a question **MUST** be uploaded as a single **PDF** file.*

1.
- A classification model performs with a high accuracy on training data but generalizes poorly to new instances. Identify the problem and illustrate it with the help of a suitable figure. Enumerate three possible solutions to this problem.

- Indicate whether each scenario given below is a classification or regression problem. Justify your answer in each case.
    - A Company wants to launch a new product and wishes to know whether the product will be successful in the market. For this purpose, data of 20 similar products is collected that were previously launched. For each product marketing budget, price of other products in competition, whether that product was successful, its price, and five other variables were collected.

    - Government is interested in predicting the % change in the USD/INR. Weekly data for the whole year of 2017 was collected. For each week, the % change happened in the USD/INR, the % change in the US market, the % change in the Indian market, and two other variables were studied.

    - A data set relates to top 500 companies in the world. The dataset consists of profit, number of employees, industry and salary of the CEO. The problem is to identify and understand factors which affect the salary of the CEO of the company.

    - A music company wants to predict the number of copies of a music album to be sold depending on the popularity rating of the singer, success index of his previous albums and two other variables.

    - A model to identify animal type (mammal vs non-mammal) in images from the animal encyclopaedia.

2.

- Design a neural network of one variable for the boolean function NOT. Show that the function is responding correctly with appropriate weight and bias.

- Give an example of a function, say f, which cannot be implemented using a single perceptron. Suggest a suitable neural network model that can implement the function f.

- Calculate the output y of a three inputs neuron with bias. The input feature vector is $(x_1, x_2, x_3) = (0.8, 0.6, 0.4)$ and weight values are $[w_1, w_2, w_3, b] = [0.2, 0.1, -0.3, 0.35]$. Use binary Sigmoid function as activation function.

3.

Given below is the data of five students who took a proficiency test as well as language course.

S. No.	Marks in proficiency test	Marks in language course
1	95	85
2	85	95
3	80	70
4	70	65
5	60	70

- Use the least square approximation to estimate the linear equation that best predicts language course performance, based on proficiency test scores?

- Compute the sum of squared error (SSE) using the estimated model.

- If a student scored 80 on the proficiency test, what marks would we expect her to obtain in the language course?

4.

Consider the dataset given below, which categorizes an article either as *Technical (Class 1)* or *Non-Technical (Class 0)* based on the time spent in *reading (in Hours)* and the *number of sentences (in multiples of 1000)* in that article.

Time (Hours)	Sentences (In multiples of 1000)	Article Type
2.7	2.5	0
3	3	0
5.9	2.2	1
7.7	3.5	1

- Using the above data, build a logistic regression model to predict the class of an article using gradient descent method. Assume *learning rate* = 0.3. Further, in the first iteration the value of the coefficients is 0, and the *bias* is 1. Use two iterations of the gradient descent process to learn the model parameters.

- Compute the error in prediction.

- Use the above model to predict the article type of an article which requires **6.2 hours** of reading time and contains **3100 sentences**.

5.

What is the assumption of the Naïve Bayes classifier? What are the advantages of the assumption? Consider the following table of ten conditions, with the attribute purchase home as the class label. Assume that the loan amount to be paid follows a normal distribution.

S. No.	PhD Student	Marital status	Loan amount to be paid	Output class (purchase home)
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90k	yes

- Obtain a Naïve Bayes classifier using the above table. Show all the prior and the conditional probabilities required to compute the posterior probabilities.

- Use the Naïve Bayes classifier obtained above to predict the class label purchase home for a given new instance: (**PhD student = No, marital status = Married, Loan to be paid = 120K**).

6.
- Suppose 10000 patients get tested for flu; out of them, 9000 are actually healthy and 1000 are actually sick. For the sick people, a test was positive for 620 and negative for 380. For healthy people, the same test was positive for 180 and negative for 8820. Construct a confusion matrix for the data and compute the True Positivity Rate (TPR), False Positivity Rate (FPR), Specificity, Sensitivity and Accuracy of the test.

- In each case, indicate whether the root mean squared error is a good performance measure and justify your answer:
    - binary classification problems
    - multiclass classification problems

- List out and briefly explain the representation power of feedforward networks.