

Name: _____
Section: _____

Reg #: _____

National University of Computer and Emerging Sciences, Lahore Campus



Course: Data Science
Program: BS(Computer Science)
Duration: 90 Minutes
Paper Date: 16-Oct-20
Section: A, B
Exam: Mid-I

Course Code: CS481
Semester: Fall 2020
Total Marks: 41
Weight: 15 %
Page(s): 8

Instruction/Notes: Attempt the examination on the question paper and write concise answers. You can use extra sheet for rough work. Do not attach extra sheets used for rough with the question paper. Don't fill the table titled Questions/Marks.

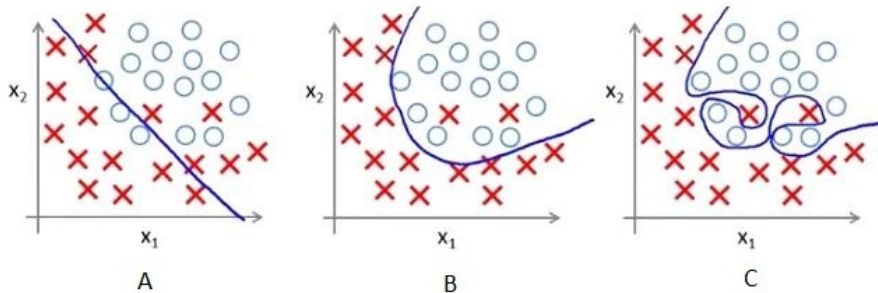
Question	Objective	1	2	3	4	Total
Marks	16 /	/ 6	/ 6	7 /	/6	41 /

Section 1

(Objective part) [points 16]

Clearly circle the correct options and explain your choice with reasoning.

Q1. Below are the three scatter plot (A,B, C left to right) and hand drawn decision boundaries for logistic regression.



Suppose, above decision boundaries were generated for the different value of regularization. Which of the above decision boundary shows the maximum regularization?

- A) A B) B C) C D) All have equal regularization

Reason:

Name: _____

Reg #: _____

Section: _____

Q2. Suppose you have a dataset with $n = 20$ features, and $m = 5000000$ examples. You want to train the parameters of this regression problem using linear regression. Should you prefer gradient descent or normal equations?

- A. Gradient descent, since $(X^T X)^{-1}$ will be very slow to compute in the normal equation.
- B. Gradient descent, since it will always converge to the optimal θ .
- C. The normal equation, since gradient descent might be unable to find the optimal θ .
- D. The normal equation, since it provides an efficient way to directly find the solution.

Reason:

Q3. Let f be some function so that $f(\theta_0, \theta_1)$ outputs a number. For this problem, f is some arbitrary/unknown smooth function (not necessarily the cost function of linear regression, so f may have local optima). Suppose we use gradient descent to try to minimize $f(\theta_0, \theta_1)$ as a function of θ_0 and

θ_1 . Which of the following statements are true? (select all that apply.)

(A) If the first few iterations of gradient descent cause $f(\theta_0, \theta_1)$ to increase rather than decrease, then

the most likely cause is that we have set the learning rate α to too large a value.





(B) If the learning rate α is too small, then gradient descent may take a very long time to converge.

(C) Even if the learning rate α is very large; every iteration of gradient descent will decrease the value of $f(\theta_0, \theta_1)$.

(D) No matter how θ_0 and θ_1 are initialized, so long as α is sufficiently small, we can safely expect gradient descent to converge to the same solution.

Reason:

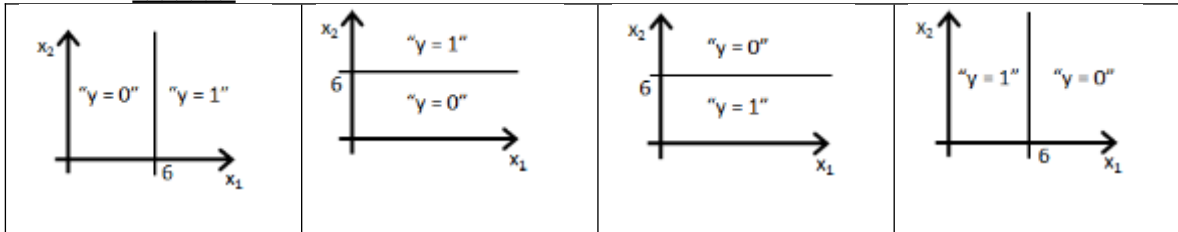
Q4. Suppose you train a logistic classifier $h_{\theta}(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$. Suppose $\theta_0 = 6$, $\theta_1 = -1$, $\theta_2 = 0$. Which of following figures represents the decision boundary found by your classifier?

(A)	(B)	(C)	(D)
			

Name: _____

Reg #: _____

Section: _____

**Reason (Explanation):****Q5.** Which of the following statements are true? Select all that apply.

- (A) The one-vs-all technique allows you to use logistic regression for problems in which each $y^{(i)}$ comes from a fixed, discrete set of values.
- (B) The cost function $J(\theta)$ for logistic regression trained with $m \geq 1$ examples is always greater than or equal to zero.
- (C) Since we train one classifier when there are two classes, we train two classifiers when there are three classes (and we do one-vs-all classification).
- (D) Linear regression always works well for classification if you classify by using a threshold on the prediction made by the linear regression.

Reason (Explanation):

Q6. Imagine, you are solving a classification problems with highly imbalanced class. The majority class is observed 99% of times in the training data. Your model has 99% accuracy after taking the predictions on test data. Which of the following is true in such a case?

- A) Accuracy metric is not a good idea for imbalanced class problems.
- B) Accuracy metric is a good idea for imbalanced class problems.
- C) Precision and recall metrics are good for imbalanced class problems.
- D) Precision and recall metrics aren't good for imbalanced class problems.

Reason:

Q7. Suppose you execute logistic regression twice, once with $\lambda = 1$, and once with $\lambda = 0$.

One of the times, you got parameters $\theta \begin{bmatrix} 71.25 \\ 11.45 \end{bmatrix}$, and the other time you got $\theta \begin{bmatrix} 13.65 \\ 0.85 \end{bmatrix}$.

However, you forgot for which values λ corresponds to which value of θ . Which one do you think to $\lambda = 1$?

a) $\theta \begin{bmatrix} 71.25 \\ 11.45 \end{bmatrix}$

b) $\theta \begin{bmatrix} 13.65 \\ 0.85 \end{bmatrix}$

Name: _____
Section: _____

Reg #: _____

Reason:

Q8. Suppose you have $m=50$ training examples with $n=10$ features (excluding the additional all-ones feature for the intercept term, which you should add). The normal equation is $\theta = (X^T X)^{-1} X^T y$. For the given values of m and n , what are the dimensions of θ , X , and y in this equation?

Answer:

Section 2 (Subjective part) (points 25)

Q1. [4+2 Marks]:

- A) [4 marks] Suppose you train a logistic regression classifier in order to predict if the patient has cancer or not. Given the test data ($m_{\text{test}} = 500$), we already know that 100 patients have actually cancer. On testing, our hypothesis predicted that 40 patients have cancer. Among the predicted ones, only 16 patients are those which actually have cancer. Draw the confusion matrix, and calculate the precision and recall for the case mentioned above.

Name: _____
Section: _____

Reg #: _____

B) [2 marks] A Data Scientist is evaluating different binary classification models. A false positive result is 5 times more expensive (from a business perspective) than a false negative result.

The models should be evaluated based on the following criteria:

- 1) Must have a recall rate of at least 80%
- 2) Must have a false positive rate of 10% or less
- 3) Must minimize business costs

After creating each binary classification model, the Data Scientist generates the corresponding confusion matrix. Which confusion matrix represents the model that satisfies the requirements?

- 1) TN = 91, FP = 9 2) TN = 99, FP = 1, FN = 21, TP = 79 3) TN = 96, FP = 4, FN = 10, TP = 90 4) TN = 98, FP = 2, FN = 18, TP = 82
FN = 22, TP = 78

Q2. [6 marks] Suppose $m = 4$ students have taken some machine learning course, and the course had a midterm exam and a final exam. You have collected a dataset of their scores on the two exams, which is as follows:

midterm exam	final exam
10	12
8	8
12	11
7	9

You

would like to use linear regression to predict a student's final exam score from their midterm exam score. Concretely, suppose you want to fit a model of the form $\mathbf{h}_{\theta}(\mathbf{x}) = \theta_0 + \theta_1 X_1$, where

X_1 is midterm score. Assume $\alpha = 0.1$, and initial $\theta_0 = 2$ and initial $\theta_1 = 3$.

The definition of the cost $J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$ function is

Your task is to execute gradient descent algorithm and compute updated values of thetas (θ_0 and θ_1) and associated cost (J) for first iteration.

Name: _____
Section: _____

Reg #: _____

Q3. Multiclass Classification: (1+2+3+1 Marks)

In the table given below, we have labeled data for patients. The output is classified into classes as given below:

$y = 1$ if patient is “**not- ill**”, $y = 2$ if patient has “**cold**”, and $y = 3$ if patient has “**Flu**”

$\mathbf{x_1}$	$\mathbf{x_2}$	\mathbf{y}
2	2	1
4	2	1
2	8	2
3	7	2
7	6	3
8	8	3

- (A) Draw the plot for the training data, where each class should be represented by a different symbol. (keep $\mathbf{x_1}$ on x-axis and $\mathbf{x_2}$ on y-axis)
- (B) How we will train logistic regression classifiers for this data?
- (C) Suppose trained parameters have same values for all models ($\Theta_0 = 0$, $\Theta_1 = 1$, $\Theta_2 = 3$), draw all the decision boundaries. (Assume simple linear Hypothesis $h(x) = \Theta_0 + \Theta_1 X_1 + \Theta_2 X_2$. Predict $y=1$ if $h(x) \geq 0.5$)
- (D) On new input x (new patient), how we will predict if the patient has “flu”, “cold” or is “not-ill”.

Name: _____
 Section: _____

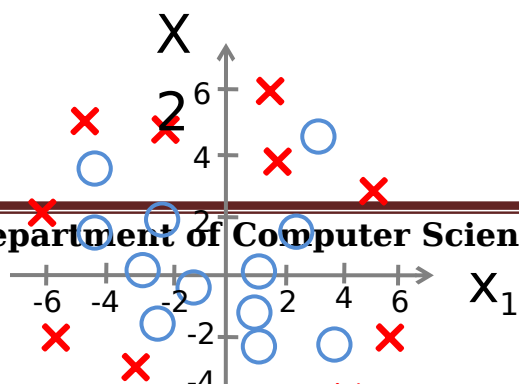
Reg #: _____

Q4. [6 marks] We consider the following model of logistic regression for binary classification

$$g(z) = \frac{1}{1+e^{-z}}$$

with a sigmoid function

Model: $h_{\theta}(x) = g(\theta_0 + \theta_1 x_1^2 + \theta_2 x_2^2 + \theta_3 x_1^2 x_2 + \theta_4 x_1^3 x_2)$



Name: _____
Section: _____

Reg #: _____

Suppose the trained parameter values are $\theta_0 = -50$, $\theta_1 = 2$, $\theta_2 = 2$, $\theta_3 = 0$, and $\theta_4 = 0$.

Draw the decision boundary.

Predict "y = 1" if $h(x) \geq 0.75$

Calculate and Draw the decision boundary according to the threshold given above. Show your working here. If you just draw the boundary without working, you will not get any point.