Name: ____________________   Reg #: ____________________

National University of Computer and Emerging Sciences, Lahore Campus

Course:	Data Science: Tools and Techniques	Course Code:	DS500
Program:	MS(Computer Science)	Semester:	Spring 2019
Duration:	75 Minutes	Total Marks:	23
Paper Date:	11-March-19	Weight	22-23 %
Section:	ALL	Page(s):	6
Exam:	Mid-I		

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	Total
Marks	/ 9	/ 6	/ 4	/ 4	/ 23

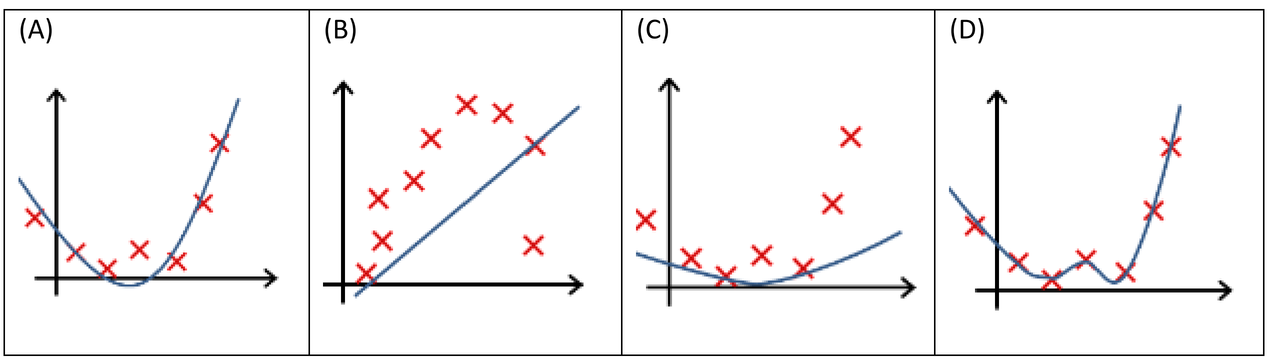
<u>Section 1 (Objective part)</u> [points 9]

Clearly circle the correct options.

Q1. Which of the following statement is true about outliers in linear regression?

A) Linear regression is sensitive to outliers   B) Linear regression is not sensitive to outliers
C) Can't say   D) None of these

Q2. In which of the following figure do you think the hypothesis is over-fitting the training set?



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.)

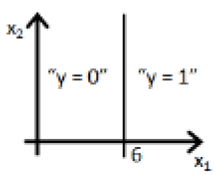
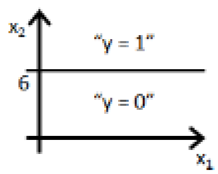
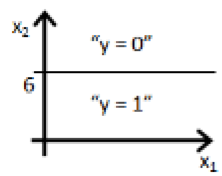
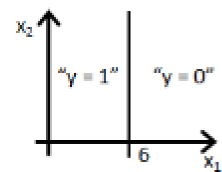
(A) If the first few iterations of gradient descent cause $f(\theta_0, \theta_1)$ to increase rather than decrease, then the most likely cause is that we have set the learning rate α to too large a value.

(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.

Q4. Suppose you train a logistic classifier $\mathbf{h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)}$. 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)
			

Q5. The Supervised learning problem is called Regression problem if

(A) the input is continuous (B) the input is discrete (c) the output is discrete (D) the output is continuous

Q6. Suppose that you have trained a logistic regression classifier, and it outputs a new example **x** a prediction $\mathbf{h_\theta(x) = 0.6}$. This means (select all that apply):

(A) our estimate for $P(y = 0 | x; \theta)$ is 0.4 (B) our estimate for $P(y = 0 | x; \theta)$ is 0.6

(C) our estimate for $P(y = 1 | x; \theta)$ is 0.6 (D) our estimate for $P(y = 1 | x; \theta)$ is 0.4

Q7. You are training a classification model with logistic regression, which of the following statement are true. Select all that apply.

(A) Introducing regularization to the model always results in equal or better performance on examples not in the training set.

(B) Adding many new features to the model helps prevent overfitting on the training set.

(C) Adding many new features to the model makes it more likely to overfit the training set.

(D) Adding a new feature to the model always results in equal or better performance on examples not in the training set.

Q8. Which of the following statement about regularization are true. Select all that apply.

(A) Using too large value of **λ** can cause your hypothesis to overfit the data; this can be avoided by reducing **λ.**

(B) Using too large value of λ can cause your hypothesis to underfit the data.

(C) Using too small value of λ can cause your hypothesis to overfit the data.

(D) Using very large value of **λ** cannot hurt the performance of your hypothesis; the only reason we do not set **λ** to be too large is to avoid numerical problems.

Q 9: Gradient Descent always finds the global optimum irrespective of the nature or shape of the cost function.

a) True b) False

Section 2 (Subjective part) (marks 14)

Q1. [2+2+2 Marks] **Short Questions**:

A) [2 marks] You are a reviewer for the International conference on Learning Algorithms, and you read papers with the following experimental setups. Would you accept or reject each paper? Provide a one sentence justification. (This conference has short reviews.)
i). **accept/reject** "My algorithm is better than yours. Look at the training error rates!"

ii). **accept/reject** "My algorithm is better than yours. Look at the test error rates. Suppose we have Choosing λ based on the test data."

B) **[2 point]** In classification problems, why do we use equation (ii) for cost instead of equation (i) given below. Here in both equations h(x) is a sigmoid logistic function.

$\mathrm{Cost}(h_\theta(x^{(i)}), y^{(i)}) = \frac{1}{2}\left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$	Equation (i)
$\mathrm{Cost}(h_\theta(x), y) = \begin{cases} -\log(h_\theta(x)) & \text{if } y = 1 \\ -\log(1 - h_\theta(x)) & \text{if } y = 0 \end{cases}$	Equation (ii)

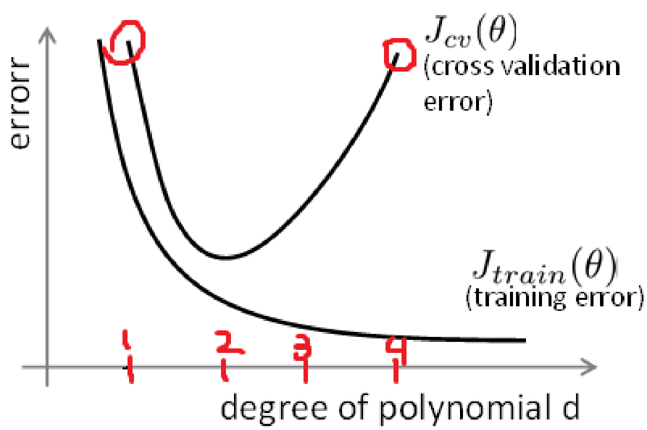
C) [2 marks] **Diagnosing bias vs. variance: Answer the following questions:**

(1). If $J_{cv}(\theta)$ and $J_{train}(\theta)$ are high such that ($J_{cv}(\theta) \approx J_{train}(\theta)$). Is it a bias problem or variance problem?

(2). If $J_{train}(\theta)$ is low and $J_{cv}(\theta) >> J_{train}(\theta)$. Is it a bias problem or variance problem?

(3). For what value of d (degree of polynomial), the problem is underfit?

(4). For what value of d (degree of polynomial), the problem is overfit?



Name: ____________________ Reg #: ____________________

Q2. Multiclass Classification: (1+2+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**"

$\mathbf{x_1}$	$\mathbf{x_2}$	**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)

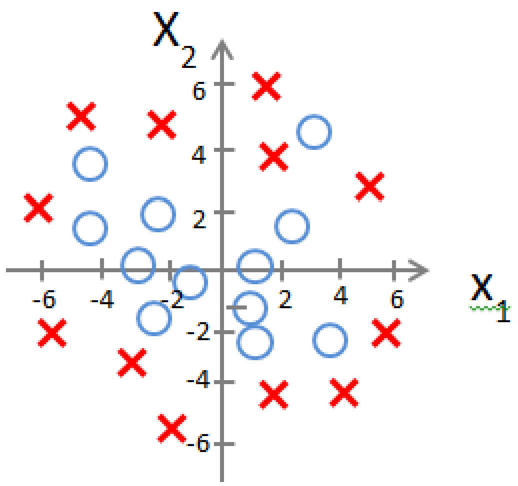
(B) How we will train logistic regression classifiers for this data?

(C) On new input x (new patient), how we will predict if the patient has "flu", "cold" or is "not-ill".

[4 points] Logistic Regression – Decision Boundary:

We consider the following model of logistic regression for binary classification with a sigmoid function as discussed in the course.

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



Suppose the trained parameter values are **ϴ$_0$ = -48, ϴ$_1$ = 3, ϴ$_2$ = 3, ϴ$_3$ = 0,** and **ϴ$_4$ = 0**.

Predict "y = 1" if h(x) >= 0.5

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.