Quiz 1: Introduction to Data science: Tools and Techniques **Total Marks: 10** 2019-02-18

Name: ---------- **Registration #: ----------** **Section: ------**

Consider the problem of predicting how well a student does in her last year of university, given how well they did in their third year. Specifically, let x be equal to the number of "A" grades that a student receives in their third year of university.

We would like to predict the value of y, which we define as the number of "A" grades in last year.

Questions 1 through 2 will use the following training set of a small sample of different students' performances.

x	y
4	3
3	4
2	3.5
0	2

Hypothesis: $h_\theta(x) = \theta_0 + \theta_1 x$

Here each row is one training example and total number of training examples are denoted by m.

Question 1: (3 marks)

For this question, continue to assume that we are using the training set given above. Recall our definition of the cost function was $J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2$. What is $J(0,1)$?

Question 2: (2 marks)

Suppose we set $\theta_0 = -1, \theta_1 = 0.5$ What is $h_\theta(4)$?

Part 1: Draw points (crosses) for all training examples given above (on y-axis we have y and on x-axis we have independent variable x).

Part 2: draw graph (linear line) for h(x). Now on y-axis we have h(x) and on x-axis we have independent variable x). Note: for both parts use same graph.

Quiz 1: Introduction to Data science: Tools and Techniques **Total Marks: 10** 2019-02-18

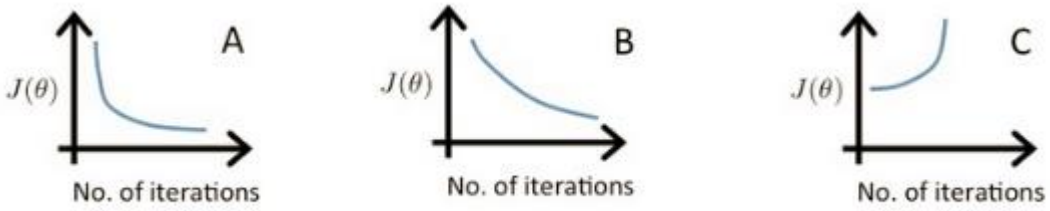
Name: ---------- **Registration #: ----------** **Section: ------**

Question 3: (5 marks) Select the correct options and explain your choice with reasoning.

i) The _____ is often the preferred measure of central tendency if the data are severely skewed.

a. Mean b. Median c. Mode d. Range

ii) Which of the following is true about below graphs (A,B, C left to right) between the cost function and Number of iterations?

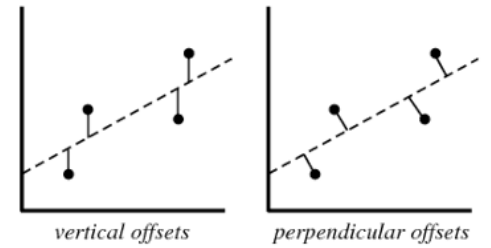


Suppose l1, l2 and l3 are the three learning rates for A,B,C respectively. Which of the following is true about l1, l2 and l3?

A) l2 < l1 < l3 B) l1 > l2 > l3 C) l1 = l2 = l3 D) It depends on derivative

Reason:

iii) Which of the following offsets, do we use in linear regression's least square line fit? Suppose horizontal axis is independent variable and vertical axis is dependent variable.



A) Vertical offset

B) Perpendicular offset

C) Both, depending on the situation

D) None of above

Reason:

iv) A multiple regression model has the form: $y = 2 + 3x_1 + 4x_2$. As x1 increases by 1 unit (holding x_2 constant), y will

(A) decrease by 4 units (B) increase by 4 units (C) decrease by 3 units (D) increase by 3 units

v) Focusing on describing or explaining data versus going beyond immediate data and making inferences is the difference between _____.

a. Central tendency and common tendency b. Mutually exclusive and mutually exhaustive properties

c. Descriptive and inferential d. Positive skew and negative skew

Reason: