Quiz 1: Data science
Total Marks: 10 2018-02-23

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

Question 1: (4 marks)

Let's say we try to predict blood fat based on age and weight of a child. The training dataset is as follows:

Weight (x_1)	Age (x_2)	blood fat
3	2	6
5	4	10
6	3	8
4	3	7

You would like to use linear regression to predict the blood fat from weight and age of the child. Concretely, suppose you want to fit a model of the form $\mathbf{h_\theta(x)} = \theta_0 + \theta_1 x_1 + \theta_2 x_2$, assuming $\alpha = 0.2$, and initial $\theta_0 = 1, \theta_1 = 2$, and $\theta_2 = 3$.

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

Your task is to execute gradient descent algorithm and compute updated values of thetas (θ_0 and θ_1, θ_2) and associated cost (J) for first the iteration. Moreover write the values of thetas, which will be used in the 2nd iteration.

Quiz 1: Data science
Total Marks: 10 2018-02-23

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

Question 2: (4 marks) Select the correct options and explain your choice with reasoning.

i) We can also compute the coefficient of linear regression with the help of an analytical method called “Normal Equation”. Which of the following is/are true about Normal Equation?

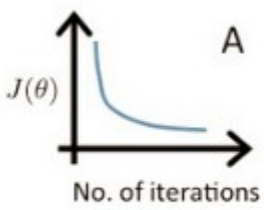
1) We don’t have to choose the learning rate
2) It becomes slow when number of features is very large
3) There is no need to iterate

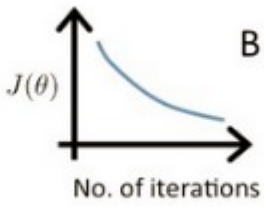
A) 1 and 2 B) 1 and 3 C) 2 and 3 D) 1, 2 and 3

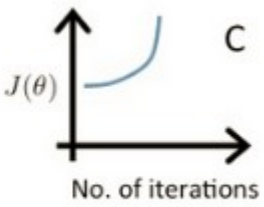
Reason:

ii) Which of the following is one of the key data science skill?
a) Statistics b) Machine Learning c) Data Visualization d) All of the Mentioned

iii) 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?

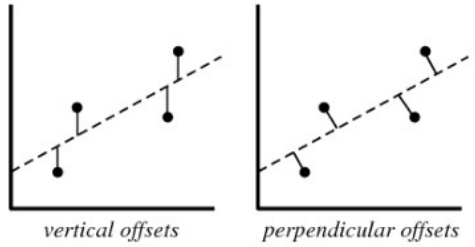
A) l2 < l1 < l3 B) l1 > l2 > l3 C) l1 = l2 = l3 D) It depends on derivative

Reason:

iv) 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.

Quiz 1: Data science
Total Marks: 10 2018-02-23

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

	A) Vertical offset B) Perpendicular offset C) Both, depending on the situation D) None of above

Reason: