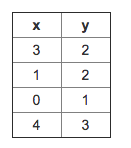
Q1.

Consider the problem of predicting how well a student does in her second year of college/university, given how well she did in her first year.

Specifically, let x be equal to the number of "A" grades (including A-. A and A+ grades) that a student receives in their first year of college (freshmen year). We would like to predict the value of y, which we define as the number of "A" grades they get in their second year (sophomore year).

Here each row is one training example. Recall that in linear regression, our hypothesis is , , and we use to denote the number of training examples.



For the training set given above (note that this training set may also be referenced in other questions in this quiz), what is the value of ? In the box below, please enter your answer (which should be a number between 0 and 10).

A1.

4, because – it’s count of records

Q2.

Consider the following training set of training examples:

|  |  |
| --- | --- |
| **x** | **y** |
| 1 | 0.5 |
| 2 | 1 |
| 4 | 2 |
| 0 | 0 |

Consider the linear regression model . What are the values of and that you would expect to obtain upon running gradient descent on this model? (Linear regression will be able to fit this data perfectly.)

A2.

5

Q3.

Suppose we set in the linear regression hypothesis from Q1. What is ?

A3.

11

Q4.

Let be some function so that outputs a number. For this problem, is some arbitrary/unknown smooth function (not necessarily the cost function of linear regression, so may have local optima).

Suppose we use gradient descent to try to minimize as a function of and . Which of the following statements are true? (Check all that apply.)

1. No matter how and are initialized, so long as is sufficiently small, we can safely expect gradient descent to converge

to the same solution.

1. If the first few iterations of gradient descent cause to increase rather than decrease, then the most likely cause is that we have set the learning rate to too large a value.
2. Setting the learning rate to be very small is not harmful, and can only speed up the convergence of gradient descent.
3. If and are initialized at the global minimum, then one iteration will not change their values.

A4.

2,4

Q5.

Suppose that for some linear regression problem (say, predicting housing prices as in the lecture), we have some training set, and for our training set we managed to find some , such that .

Which of the statements below must then be true? (Check all that apply.)

1. For these values of and that satisfy , we have that for every training example
2. This is not possible: By the definition of , it is not possible for there to exist and so that
3. For this to be true, we must have and so that
4. We can perfectly predict the value of even for new examples that we have not yet seen. (e.g., we can perfectly predict prices of even new houses that we have not yet seen.)

A5.

1