1. 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. 1 point

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).

Refer to the following training set of a small sample of different students' performances (note that this training set may also be referenced in other questions in this quiz). Here each row is one training example. Recall that in linear regression, our hypothesis is $h_\theta(x) = \theta_0 + \theta_1 x$, and we use m to denote the number of training examples.

x	y
3	4
2	1
4	3
0	1

For the training set given above, what is the value of m? In the box below, please enter your answer (which should be a number between 0 and 10).

4

2. For this question, assume that we are 1 point

using the training set from Q1. 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 $J(0, 1)$? In the box below,

please enter your answer (Simplify fractions to decimals when entering answer, and '.' as the decimal delimiter e.g., 1.5).

0.5

3. Suppose we set $\theta_0 = -2, \theta_1 = 0.5$ in the linear regression hypothesis from Q1. What is $h_\theta(6)$? 1 point

1

4. Let f be some function so that 1 point

$f(\theta_0, \theta_1)$ 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? (Check all that apply.)

- [x] 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.
- [ ] 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.
- [x] If θ_0 and θ_1 are initialized at

  the global minimum, then one iteration will not change their values.
- [ ] Setting the learning rate α to be very small is not harmful, and can

  only speed up the convergence of gradient descent.

5. 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 θ_0, θ_1 such that $J(\theta_0, \theta_1) = 0$. 1 point

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

- [x] For these values of θ_0 and θ_1 that satisfy $J(\theta_0, \theta_1) = 0$,

  we have that $h_\theta(x^{(i)}) = y^{(i)}$ for every training example $(x^{(i)}, y^{(i)})$
- [ ] We can perfectly predict the value of y 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.)
- [ ] For this to be true, we must have $\theta_0 = 0$ and $\theta_1 = 0$

  so that $h_\theta(x) = 0$
- [ ] This is not possible: By the definition of $J(\theta_0, \theta_1)$, it is not possible for there to exist

  θ_0 and θ_1 so that $J(\theta_0, \theta_1) = 0$