Feedback — II. Linear regression with one variable

You submitted this quiz on **Fri 3 May 2013 11:46 AM PDT -0700**. You got a score of **5.00** out of **5.00**.

Question 1

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

Questions 1 through 4 will use the following training set of a small sample of different students' performances. 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	2
1	2
0	1
4	3

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).

You entered:

4

Your Answer		Score	Explanation
4	✔	1.00	
Total		1.00 / 1.00	

Question Explanation

m is the number of training examples. In this example, we have m=4 examples.

Question 2

For this question, continue to assume that we are using the training set given above.

Recall our definition of the cost function was $J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$. What is $J(0, 1)$? In the box below, please enter your answer (use decimals instead of fractions if necessary, e.g., 1.5).

You entered:

0.5

Your Answer		Score	Explanation
0.5	✔	1.00	
Total		1.00 / 1.00	

Question Explanation

When $\theta_0 = 0$ and $\theta_1 = 1$, we have $h_\theta(x) = \theta_0 + \theta_1 x = x$. So,

$$
\begin{aligned}
J(\theta_0, \theta_1) &= \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2 \\
&= \frac{1}{2*4} \left((1)^2 + (1)^2 + (1)^2 + (1)^2\right) \\
&= \frac{4}{8} \\
&= 0.5
\end{aligned}
$$

Question 3

Suppose we set $\theta_0 = -2, \theta_1 = 0.5$. What is $h_\theta(6)$?

You entered:

1

Your Answer		Score	Explanation
1	✔	1.00	
Total		1.00 / 1.00	

Question Explanation

Setting $x = 6$, we have $h_\theta(x) = \theta_0 + \theta_1 x = -2 + 0.5 * 6 = 1$

Question 4

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? (Check all that apply.)

Your Answer	Score	Explanation

☐ If θ_0 and θ_1 are initialized so that $\theta_0 = \theta_1$, then by symmetry (because we do simultaneous updates to the two parameters), after one iteration of gradient descent, we will still have $\theta_0 = \theta_1$.	✔	0.25	The updates to θ_0 and θ_1 are different (even though we're doing simultaneous updates), so there's no particular reason to expect them to be the same after one iteration of gradient descent.
☐ 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.	✔	0.25	This is not true, because depending on the initial condition, gradient descent may end up at different local optima.
☑ If θ_0 and θ_1 are initialized at the global minimum, the one iteration will not change their values.	✔	0.25	At the global minimum, the derivative (gradient) is zero, so gradient descent will not change the parameters.
☑ If the learning rate is too small, then gradient descent may take a very long time to converge.	✔	0.25	If the learning rate is small, gradient descent ends up taking an extremely small step on each iteration, and therefore can take a long time to converge.
Total		1.00 / 1.00	

Question 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 $J(\theta_0, \theta_1) = 0$. Which of the statements below must then be true? (Check all that apply.)

Your Answer	Score	Explanation

☐ Gradient descent is likely to get stuck at a local minimum and fail to find the global minimum.	✔	0.25	The cost function $J(\theta_0, \theta_1)$ for linear regression has no local optima (other than the global minimum), so gradient descent will not get stuck at a bad local minimum.
☐ For this to be true, we must have $\theta_0 = 0$ and $\theta_1 = 0$ so that $h_\theta(x) = 0$	✔	0.25	If $J(\theta_0, \theta_1) = 0$, that means the line defined by the equation "$y = \theta_0 + \theta_1 x$" perfectly fits all of our data. There's no particular reason to expect that the values of θ_0 and θ_1 that achieve this are both 0 (unless $y^{(i)} = 0$ for all of our training examples).
☑ 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 $(x^{(i)}, y^{(i)})$	✔	0.25	$J(\theta_0, \theta_1) = 0$, that means the line defined by the equation "$y = \theta_0 + \theta_1 x$" perfectly fits all of our data.
☐ 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 $J(\theta_0, \theta_1) = 0$	✔	0.25	If all of our training examples lie perfectly on a line, then $J(\theta_0, \theta_1) = 0$ is possible.
Total		1.00 / 1.00	