Feedback — II. Linear regression with one variable

Help Center

You submitted this quiz on **Sat 24 Jan 2015 8:46 AM CET**. 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

Consider the following training set of $m = 4$ training examples:

x	y
1	0.5
2	1
4	2
0	0

Consider the linear regression model $h_\theta(x) = \theta_0 + \theta_1 x$. What are the values of θ_0 and θ_1 that you would expect to obtain upon running gradient descent on this model? (Linear regression will be able to fit this data perfectly.)

Your Answer		Score	Explanation
$\theta_0 = 0.5, \theta_1 = 0$			
$\theta_0 = 0.5, \theta_1 = 0.5$			
$\theta_0 = 0, \theta_1 = 0.5$	✔	1.00	
$\theta_0 = 1, \theta_1 = 1$			
Total		1.00 / 1.00	

Question Explanation

As $J(\theta_0, \theta_1) = 0$, $y = h_\theta(x) = \theta_0 + \theta_1 x$. Using any two values in the table, solve for θ_0, θ_1.

Question 3

Consider the training set below with only $m = 3$ training examples:

x	y
1	1
2	2
3	3

Recall the gradient descent algorithm used to update θ_0 and θ_1:

$\theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})$

$\theta_1 := \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) \cdot x^{(i)}$

Now let's assume we choose θ_0 = 0 and θ_1 = 0.5 as our starting point, and we choose α to be 0.1. After you perform one iteration of gradient descent, what will be the new values for θ_0 and θ_1?

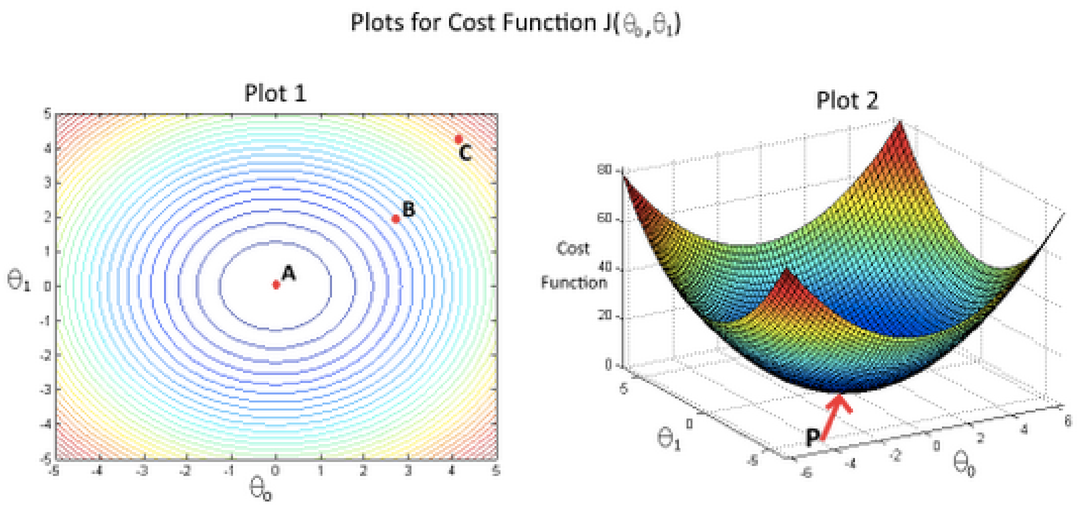
Your Answer		Score	Explanation
◯ θ_0 = 0 and θ_1 = 1			
◯ θ_0 = 0.1 and θ_1 = 0.713			
◯ θ_0 = 0.15 and θ_1 = 0.632			
◉ θ_0 = 0.1 and θ_1 = 0.733	✔	1.00	
Total		1.00 / 1.00	

Question Explanation

Remember to update both θ_0 and θ_1 simultaneously. So,
$\theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})$ = 0 - 0.1 x (1/3) x ((0.5 x 1 - 1) + (0.5 x 2 - 2) + (0.5 x 3 - 3)) = 0.1 $\theta_1 := \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) \cdot x^{(i)}$ = 0.5 - 0.1 x (1/3) x ((0.5 x 1 - 1) x 1 + (0.5 x 2 - 2) x 2 + (0.5 x 3 - 3)) x 3) = 0.733

Question 4

In the given figure, the cost function $J(\theta_0, \theta_1)$ has been plotted against θ_0 and θ_1, as shown in 'Plot 2'. The contour plot for the same cost function is given in 'Plot 1'. Based on the figure, choose the correct options (check all that apply).



Your Answer		Score	Explanation
☑ Point P (the global minimum of plot 2) corresponds to point A of Plot 1.	✔	0.20	Correct. Plot 2 is a 3-D surface plot for cost function $J(\theta_0, \theta_1)$ against θ_0 and θ_1, whereas Plot 1 is the 2-D contour plot for the same cost function. Hence, the correspondence of the two plots can be understood.
☐ If we start from point B, gradient descent with a well-chosen learning rate will eventually help us reach at or near point C, as the value of cost function $J(\theta_0, \theta_1)$ is minimum at point C.	✔	0.20	
☐ Point P (The global minimum of plot 2) corresponds to point C of Plot 1.	✔	0.20	
☐ If we start from point B, gradient descent with a well-chosen learning rate will eventually help us reach at or near point A, as the value of cost function $J(\theta_0, \theta_1)$ is maximum at point A.	✔	0.20	
☑ If we start from point B, gradient descent with a well-chosen learning rate will eventually help us reach at or near point A, as the value of cost function $J(\theta_0, \theta_1)$ is minimum at A.	✔	0.20	Correct. Correct implementation of Gradient Descent Algorithm will help us minimizing the cost function $J(\theta_0, \theta_1)$. Since point A represents the global minimum of the cost function, gradient descent should lead us to reach at or near point A.
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
☐ 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.)	✔	0.25	Even though we can fit our training set perfectly, this does not mean that we'll always make perfect predictions on houses in the future/on houses that we have not yet seen.
☑ Our training set can be fit perfectly by a straight line, i.e., all of our training examples lie perfectly on some straight line.	✔	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.
☐ For this to be true, we must have $y^{(i)} = 0$ for every value of $i = 1, 2, \ldots, m$.	✔	0.25	So long as all of our training examples lie on a straight line, we will be able to find θ_0 and θ_1 so that $J(\theta_0, \theta_1) = 0$. It is not necessary that $y^{(i)} = 0$ for all of our examples.
☐ 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).
Total		1.00 / 1.00	