

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).

Refer to the following training set of a small sample of different students' performances (note that this training set will 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).

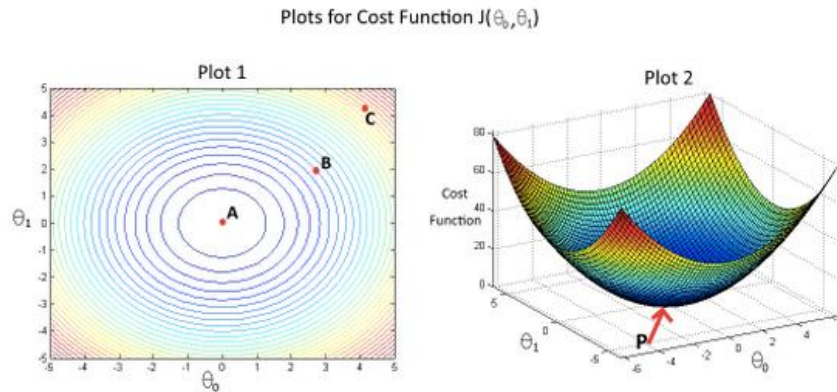
2. For this question, continue to assume that we are using the training set given above. Recall our definition of the cost function was $J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$.

What is $J(0, 1)$? In the box below,

please enter your answer (use decimals instead of fractions if necessary, e.g., 1.5).

3. Suppose we set $\theta_0 = 0, \theta_1 = 1.5$. What is $h_{\theta}(2)$?

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).



- ☐ 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.
 - ☐ 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.
 - ☒ Point P (the global minimum of plot 2) corresponds to point A of Plot 1.
 - ☐ 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.
 - ☐ Point P (The global minimum of plot 2) corresponds to point C of Plot 1.
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.)

- ☐ For this to be true, we must have $y^{(i)} = 0$ for every value of $i = 1, 2, \dots, m$.
- ☐ For this to be true, we must have $\theta_0 = 0$ and $\theta_1 = 0$
so that $h_{\theta}(x) = 0$
- ☒ Our training set can be fit perfectly by a straight line,
i.e., all of our training examples lie perfectly on some straight line.
- ☐ Gradient descent is likely to get stuck at a local minimum and fail to find the global minimum.