Question 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.

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 $h_\theta(x)=\theta_0+\theta_1 x$, and we use m to denote the number of training examples.

x	y
5	4
3	4
0	1
4	3

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 m? In the box below, please enter your answer (which should be a number between 0 and 10).

Answer:

4

Question 2

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

x	y
1	0.5
2	1

x	y
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.)

- $\theta_0=0.5,\theta_1=0$
- $\theta_0=0.5,\theta_1=0.5$
- $\theta_0=1,\theta_1=0.5$
- $\theta_0=0,\theta_1=0.5$
- $\theta_0=1,\theta_1=1$
-

Answer:

$\theta_0=0,\theta_1=0.5$

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

Suppose we set $\theta_0=-1,\theta_1=0.5$. What is $h_\theta(4)$?

Answer:

Setting x = 4, we have $h_\theta(x)=\theta_0+\theta_1 x$ = -1 + (0.5)(4) = **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.)

- Even if the learning rate α is very large, every iteration of gradient descent will decrease the value of $f(\theta_0,\theta_1)$.
- If the learning rate is too small, then gradient descent may take a very long time to converge.
- If θ_0 and θ_1 are initialized at a local minimum, then one iteration will not change their values.
- 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 $\theta_0=\theta_1$.

Answers:

True or False	Statement	Explanation
True	If the learning rate is too small, then gradient descent may take a very long time to converge.	If the learning rate is small, gradient descent ends up taking an extremely small step on each iteration, and therefor can take a long time to converge
True	If θ_0 and θ_1 are initialized at a local minimum, then one iteration will not change their values.	At a local minimum, the derivative (gradient) is zero, so gradient descent will not change the parameters.

True or False	Statement	Explanation
False	Even if the learning rate α is very large, every iteration of gradient descent will decrease the value of $f(\theta_0,\theta_1)$.	If the learning rate is too large, one step of gradient descent can actually vastly "overshoot" and actually increase the value of $f(\theta_0,\theta_1)$.
False	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 $\theta_0=\theta_1$.	The updates to θ_0 and θ_1 are different (even though we're doing simulaneous updates), so there's no particular reason to update them to be same after one iteration of gradient descent.

Other Options:

True or False	Statement	Explanation
True	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 to too large a value	if alpha were small enough, then gradient descent should always successfully take a tiny small downhill and decrease $f(\theta_0,\theta_1)$ at least a little bit. If gradient descent instead increases the objective value, that means alpha is too large (or you have a bug in your code!).
False	No matter how θ_0 and θ_1 are initialized, so long as learning rate is sufficiently small, we can safely expect gradient descent to converge to the same solution	This is not true, depending on the initial condition, gradient descent may end up at different local optima.

True or False	Statement	Explanation
False	Setting the learning rate to be very small is not harmful, and can only speed up the convergence of gradient descent.	If the learning rate is small, gradient descent ends up taking an extremely small step on each iteration, so this would actually slow down (rather than speed up) the convergence of the algorithm.

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.)

- For this to be true, we must have $y^{(i)}=0$ for every value of $i=1,2,\ldots,m$.
- Gradient descent is likely to get stuck at a local minimum and fail to find the global minimum.
- 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.

True or False	Statement	Explanation
False	For this to be true, we must have $y^{(i)}=0$ for every value of $i=1,2,\ldots,m$.	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 $y^{(i)}$ for all our examples.

True or False	Statement	Explanation
False	Gradient descent is likely to get stuck at a local minimum and fail to find the global minimum.	none
False	For this to be true, we must have $\theta_0=0$ and $\theta_1=0$ so that $h_\theta(x)=0$	If $J(\theta_0,\theta_1)=0$ that means the line defined by the equation "y = θ_0 + θ_1x" 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).
True	Our training set can be fit perfectly by a straight line, i.e., all of our training examples lie perfectly on some straight line.	None

Other Options:

True or False	Statement	Explanation
False	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.)	None
False	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 $J(\theta_0,\theta_1)=0$	None
True	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 $(x^{(i)},y^{(i)})$	None