Linear Regression with One Variable

TOTAL POINTS 5

---

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).

   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 (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).

   > 4

2. Consider the following training set of $m = 4$ training examples: (1 point)

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.)

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

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

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.)

- [ ] 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 $\theta_0 = \theta_1$.

- [ ] Even if the learning rate α is very large, every iteration of

      gradient descent will decrease the value of $f(\theta_0, \theta_1)$.

- [x] If the learning rate is too small, then gradient descent may take a very long

      time to converge.

- [x] If θ_0 and θ_1 are initialized at

      a local minimum, then one iteration will not change their values.

5. Suppose that for some linear regression problem (say, predicting housing prices as in the lecture), we have some training   1 point
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.)

- [ ] 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 $h_\theta(x) = 0$

- [x] Our training set can be fit perfectly by a straight line,

      i.e., all of our training examples lie perfectly on some straight line.

- [ ] For this to be true, we must have $y^{(i)} = 0$ for every value of $i = 1, 2, \ldots, m$.