Advanced Algorithms

Floriano Zini
Free University of Bozen-Bolzano
Faculty of Computer Science

Academic Year 2013-2014

Lab 12 – Linear regression and gradient descent
Assignment 10

Exercise 1
Consider the problem of predicting how well students do in their 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).

Exercises 1 through 3 will use the training set on the right 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.

For the given training set, what is the value of $m$?

<table>
<thead>
<tr>
<th>$x$</th>
<th>$y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Assignment 10

Exercise 2
For this question, continue to assume that we are using the training set given in the previous slide and let $J(\theta_0, \theta_1)$ be the cost function as defined in the lectures. What is $J(0,1)$?
Assignment 10

Exercise 3
Suppose we set $\theta_0 = -2$, $\theta_1 = 0.5$. What is $h_\theta(6)$?

Exercise 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 $\theta_0$ and $\theta_1$.
Which of the following statements are true? (Check all that apply.)
- Setting the learning rate $\alpha$ to be very small is not harmful, and can only speed up the convergence of gradient descent.
- If $\theta_0$ and $\theta_1$ 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 $\theta_0 = \theta_1$.
- If $\theta_0$ and $\theta_1$ are initialized at the global minimum, then one iteration of gradient descent will not change their values.
- 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 $\alpha$ to too large a value.
Exercise 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 $\theta_0, \theta_1$ such that $J(\theta_0, \theta_1) = 0$. Which of the statements below must then be true? (Check all that apply.)

- 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.)
- 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.
- This is not possible: By the definition of $J(\theta_0, \theta_1)$, it is not possible for there to exist $\theta_0$ and $\theta_1$ so that $J(\theta_0, \theta_1) = 0$