Close

Linear Regression with Multiple Variables

5 questions

1.

Suppose *m*=4 students have taken some class, and the class had a midterm exam and a final exam. You have collected a dataset of their scores on the two exams, which is as follows:

|  |  |  |
| --- | --- | --- |
| midterm exam | (midterm exam)2 | final exam |
| 89 | 7921 | 96 |
| 72 | 5184 | 74 |
| 94 | 8836 | 87 |
| 69 | 4761 | 78 |

You'd like to use polynomial regression to predict a student's final exam score from their midterm exam score. Concretely, suppose you want to fit a model of the form *hθ*(*x*)=*θ*0+*θ*1*x*1+*θ*2*x*2, where *x*1 is the midterm score and *x*2 is (midterm score)2. Further, you plan to use both feature scaling (dividing by the "max-min", or range, of a feature) and mean normalization.

What is the normalized feature *x*(1)1? (Hint: midterm = 89, final = 96 is training example 1.) Please round off your answer to two decimal places and enter in the text box below.



2.

You run gradient descent for 15 iterations

with *α*=0.3 and compute *J*(*θ*) after each

iteration. You find that the value of *J*(*θ*) **increases** over

time. Based on this, which of the following conclusions seems

most plausible?



Rather than use the current value of *α*, it'd be more promising to try a larger value of *α*(say *α*=1.0).



*α*=0.3 is an effective choice of learning rate.



Rather than use the current value of *α*, it'd be more promising to try a smaller value of *α*(say *α*=0.1).

3.

Suppose you have *m*=28 training examples with *n*=4 features (excluding the additional all-ones feature for the intercept term, which you should add). The normal equation is *θ*=(*XTX*)−1*XTy*. For the given values of *m* and *n*, what are the dimensions of *θ*, *X*, and *y* in this equation?



*X* is 28×4, *y* is 28×1, *θ* is 4×4



*X* is 28×5, *y* is 28×1, *θ* is 5×1



*X* is 28×5, *y* is 28×5, *θ* is 5×5



*X* is 28×4, *y* is 28×1, *θ* is4×1

4.

Suppose you have a dataset with *m*=50 examples and *n*=15 features for each example. You want to use multivariate linear regression to fit the parameters *θ* to our data. Should you prefer gradient descent or the normal equation?



Gradient descent, since (*XTX*)−1 will be very slow to compute in the normal equation.



Gradient descent, since it will always converge to the optimal *θ*.



The normal equation, since gradient descent might be unable to find the optimal *θ*.



The normal equation, since it provides an efficient way to directly find the solution.

5.

Which of the following are reasons for using feature scaling?



It speeds up solving for *θ* using the normal equation.



It prevents the matrix *XTX* (used in the normal equation) from being non-invertable (singular/degenerate).



It is necessary to prevent gradient descent from getting stuck in local optima.



It speeds up gradient descent by making it require fewer iterations to get to a good solution.

Submit Quiz

[Help Center](https://accounts.coursera.org/i/zendesk/courserahelp?return_to=https://learner.coursera.help/hc" \t "_blank)