

Required to pass: 80% or higher

You can retake this quiz up to 3 times every 8 hours.

Back to Week 2

Retake



0/1 points

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) <sup>2</sup>	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  $h_{\theta}(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2$ , 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  $x_1^{(3)}$ ? (Hint: midterm = 94, final = 87 is training example 3.) Please round off your answer to two decimal places and enter in the text box below.



1/1 points

2

You run gradient descent for 15 iterations

with lpha=0.3 and compute J( heta) after each

iteration. You find that the value of  $J(\theta)$  **increases** over

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

most plausible?



1/1 points

3.

Suppose you have m=14 training examples with n=3 features (excluding the additional all-ones feature for the intercept term, which you should add). The normal equation is  $\theta=(X^TX)^{-1}X^Ty$ . For the given values of m and n, what are the dimensions of  $\theta$ , X, and y in this equation?



1/1 points

4.

Suppose you have a dataset with m=50 examples and n=200000 features for each example. You want to use multivariate linear regression to fit the parameters  $\theta$  to our data. Should you prefer gradient descent or the normal equation?



0/1 points

5.

Which of the following are reasons for using feature scaling?





