1 point 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 $h_{\theta}(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2$, 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 $x_2^{(2)}$? (Hint: midterm = 72, final = 74 is training example 2.) Please round off your answer to two decimal places and enter in the text box below.

-0.36

1 point 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

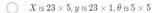
time. Based on this, which of the following conclusions seems



most plausible?

- Rather than use the current value of α , it'd be more promising to try a larger value of α (say $\alpha=1.0$).
- (a) Rather than use the current value of α , it'd be more promising to try a smaller value of α (say $\alpha=0.1$).
- lpha=0.3 is an effective choice of learning rate.

1 point S. Suppose you have m=23 training examples with n=5 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 θ , X, and y in this equation?





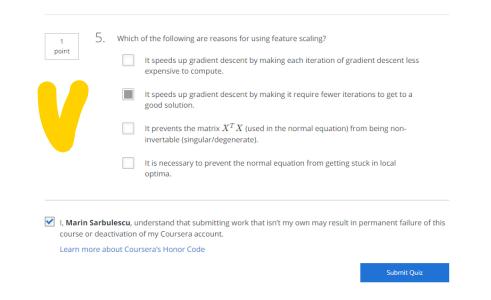
- lacksquare X is 23 imes 6 , y is 23 imes 1 , heta is 6 imes 1
- X is 23×5 , y is 23×1 , θ is 5×1

1 point

- 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 $(X^TX)^{-1}$ will be very slow to compute in the normal equation.



- The normal equation, since gradient descent might be unable to find the optimal $\boldsymbol{\theta}.$
- The normal equation, since it provides an efficient way to directly find the solution.
- Gradient descent, since it will always converge to the optimal heta.



3 P P