×

## Linear Regression with Multiple Variables

5 questions

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) <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  $x_1^{(1)}$ ? (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.

0.30

1 point 2.

You run gradient descent for 15 iterations

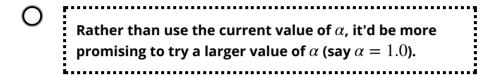
with  $\alpha = 0.3$  and compute

 $J(\theta)$  after each iteration. You find that the

value of  $J(\theta)$  decreases slowly and is still

decreasing after 15 iterations. Based on this, which of the

following conclusions seems most plausible?



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

1 point

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  $\theta=(X^TX)^{-1}X^Ty$ . For the given values of m and n, what are the dimensions of  $\theta$ , X, and y in this equation?

- **O**  $X ext{ is } 28 \times 4, y ext{ is } 28 \times 1, \theta ext{ is } 4 \times 1$
- **O**  $X ext{ is } 28 \times 4, y ext{ is } 28 \times 1, \theta ext{ is } 4 \times 4$
- **O**  $X ext{ is } 28 \times 5, y ext{ is } 28 \times 5, \theta ext{ is } 5 \times 5$
- O X is  $28 \times 5$ , y is  $28 \times 1$ ,  $\theta$  is  $5 \times 1$

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  $\theta$  to our data. Should you prefer gradient descent or the normal equation?

0	Gradient descent, since $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ will be very slow to compute in the normal equation.
0	Gradient descent, since it will always converge to the optimal $ heta.$
0	The normal equation, since gradient descent might be unable to find the optimal $\boldsymbol{\theta}.$
0	The normal equation, since it provides an efficient way to directly find the solution.
1 point	of the following are reasons for using feature scaling?
	It speeds up gradient descent by making it require fewer iterations to get to a good solution.
	It prevents the matrix $\boldsymbol{X}^T\boldsymbol{X}$ (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 solving for $ heta$ using the normal equation.
	Submit Quiz





