Linear Regression with Multiple Variables

Quiz, 5 questions

5/5 points (100%)

Congratulations! You passed!

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

Answer: 0.32

✔ 1 / 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** quickly then levels off. Based on this, which of the following conclusions seems most plausible?

Answer: α = 0.3 is an effective choice of learning rate.

✔ 1 / 1 point

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

Answer: X=23*6 ; y=23*1 ; θ=6*1

✔ 1 / 1 point

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 θ to our data. Should you prefer gradient descent or the normal equation?

Answer: Gradient descent, since $(X^TX)^{-1}$ will be very slow to compute in the normal equation.

✔ 1 / 1 point

5. Which of the following are reasons for using feature scaling?

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