# Linear Regression with Multiple Variables

Latest Submission Grade
40%
1.
Question 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)2^22 | 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)=θ0+θ1x1+θ2x2h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2hθ(x)=θ0+θ1x1+θ2x2, where x1x_1x1 is the midterm score and x2x_2x2 is (midterm score)2^22. 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 x1(1)x_1^{(1)}x1(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 / 1 point**
Incorrect
2.
Question 2

You run gradient descent for 15 iterations

with α=0.3\alpha = 0.3α=0.3 and compute

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

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

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

following conclusions seems most plausible?

**0 / 1 point**
Incorrect

3.
Question 3

Suppose you have m=28m = 28m=28 training examples with n=4n = 4n=4 features (excluding the additional all-ones feature for the intercept term, which you should add). The normal equation is θ=(XTX)−1XTy\theta = (X^TX)^{-1}X^Tyθ=(XTX)−1XTy. For the given values of mmm and nnn, what are the dimensions of θ\thetaθ, XXX, and yyy in this equation?

**1 / 1 point**
Correct
4.
Question 4

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

**1 / 1 point**
Correct
5.
Question 5

Which of the following are reasons for using feature scaling?

**0 / 1 point**
Incorrect