Week 2

Quiz 1-Linear Regression with Multiple Variables

QUES1.

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 $x_2^{(4)}$? (Hint: midterm = 69, final = 78 is training example 4.) Please round off your answer to two decimal places and enter in the text box below.

Answer:
mean = (7921+5184+8836+4761)/4 = 6675.5
deviation = 8836-4761 = 4075
normalized x2(4) = (4761-6675.5) / 4075 = -0.50

Ques2. You run gradient descent for 15 iterations with α=0.3 and compute J(θ) after each iteration. You find that the value of J(θ) 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.

Ques3. 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^T X)^{-1} X^T yθ. For the given values of m and n, what are the dimensions of θ, X, and y in this equation?
Answer: X is 28×5, y is 28×1, θ is 5×1

Ques4. Suppose you have a dataset with m = 1000000 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^T X)^{-1} will be very slow to compute in the normal equation.

Ques5. 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.