***WEEK 2 QUIZ***

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 2 exams, which is as follows:

midterm exam (midterm exam)2 final exam

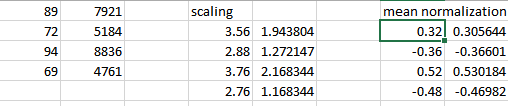
89 7921 96

72 5184 74

94 8836 87

69 4761 78

Use polynomial regression to predict a student's final score from their midterm score. Suppose you want to fit a model of form **hθ(x) = θ0 + θ1x1 + θ2x2**, where x1 = midterm score + x2 = (midterm score)^2. Further, we use *both* feature scaling (dividing by 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.

* 
* **0.32**

1. You run gradient descent for 15 iterations w/ α = 0.3 + compute J(θ) after each + find that the value of J(θ) *increases over time*. Based on this, which of the following conclusions seems most plausible?

* **Rather than use current value of α, it'd be more promising to try a smaller value (say α=0.1)**

1. You run gradient descent for 15 iterations w/ α=0.3 and compute J(θ) after each iteration. You find the value of J(θ) *decreases quickly then levels off*. Based on this, which of the following conclusions seems most plausible?

* **α=0.3 is an effective choice of learning rate.**

1. Suppose you have m=23 training examples w/ n = 5 features (excluding additional x1, all-1’s feature for the intercept term, which you should add). The normal equation is θ = (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?

* **X is 23×6, y is 23×1, θ is 6×1 🡪** X = m\*n+1, y = m\*1, θ = n\*1

1. Suppose you have m = 50 examples + 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?

* **The normal equation, since it provides an efficient way to directly find the solution.**

1. Suppose you have m = 1000000 examples + 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?

* **Gradient descent, since (X(t)\*X)^−1 will be very slow to compute in the normal equation 🡪** *n is too high*

1. Which 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.**