×

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)^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.

1.17

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 $J(\theta)$ **increases** over

time. Based on this, which of the following conclusions seems

most plausible?

- $\alpha = 0.3$ is an effective choice of learning rate.
- Rather than use the current value of α, it'd be more promising to try a larger value of α (say $\alpha = 1.0$).
- Rather than use the current value of α, it'd be more promising to try a smaller value of α (say $\alpha = 0.1$).

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

- X is 28×5, y is 28×1, θ is 5×1
- X is 28×5, y is 28×5, θ is 5×5
- X is 28×4, y is 28×1, θ is 4×4
- X is 28×4, y is 28×1, θ is4×1

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 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.
- ○ The normal equation, since gradient descent might be unable to find the optimal θ.
- ○ Gradient descent, since it will always converge to the optimal θ.
- ○ The normal equation, since it provides an efficient way to directly find the solution.

1
point

5.

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.
- ☐ It prevents the matrix $X^T X$ (used in the normal equation) from being non-invertable (singular/degenerate).
- ☐ It speeds up gradient descent by making each iteration of gradient descent less expensive to compute.
- ☐ It is necessary to prevent the normal equation from getting stuck in local optima.

☑ I understand that submitting work that isn't my own may result in permanent failure of this course or deactivation of my Coursera account. Learn more about Coursera's Honor Code

Anderson Banihirwe

Submit Quiz