QUIZ • 10 MIN

Linear Regression with Multiple Variables

Submit your assignment
DUE Jul 19, 11:59 PM PDT **ATTEMPTS** 3 every 8 hours

Try again

Receive grade
TO PASS 80% or higher

Grade
100%
View Feedback
We keep your highest score

Linear Regression with Multiple Variables

TOTAL POINTS 5

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: — 1 point

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 $x_1^{(3)}$? (Hint: midterm = 94, final = 87 is training example 3.) Please round off your answer to two decimal places and enter in the text box below.

Enter answer here

2. You run gradient descent for 15 iterations — 1 point

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?

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

3. Suppose you have $m = 14$ training examples with $n = 3$ features (excluding the additional all-ones feature for the intercept term, which you should add). The normal equation is $\theta = (X^TX)^{-1}X^Ty$. For the given values of m and n, what are the dimensions of θ, X, and y in this equation? — 1 point

- X is 14×3, y is 14×1, θ is 3×3
- X is 14×3, y is 14×1, θ is 3×1
- X is 14×4, y is 14×1, θ is 4×1
- X is 14×4, y is 14×4, θ is 4×4

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? — 1 point

- Gradient descent, since it will always converge to the optimal θ.
- Gradient descent, since $(X^TX)^{-1}$ will be very slow to compute in the normal equation.
- The normal equation, since it provides an efficient way to directly find the solution.
- The normal equation, since gradient descent might be unable to find the optimal θ.

5. Which of the following are reasons for using feature scaling? — 1 point

- [ ] It speeds up solving for θ using the normal equation.
- [ ] It speeds up gradient descent by making it require fewer iterations to get to a good solution.
- [ ] It prevents the matrix X^TX (used in the normal equation) from being non-invertable (singular/degenerate).
- [ ] It is necessary to prevent gradient descent 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

Save

Submit