Linear Regression with Multiple Variables

Quiz, 5 questions

5/5 points (100%)

Congratulations! You passed!

Next Item

Question Responses

- Question 1
- Question 2
- Question 3
- Question 4
- Question 5

Review Materials

- **Multiple Features**
- **Normal Equation**
- **Gradient Descent in Practice I - Feature Scaling**
- **Normal Equation Noninvertibility**
- **Gradient Descent Intuition**

1 / 1
points

Concepts

Use mean normalization and scaling for feature scaling

Gradient Descent in Practice I - Feature Scaling (05:42)

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 / 1
points

Concepts

Explain the different behaviors of gradient descent when the learning rate is too small or too big

Gradient Descent Intuition (05:31)

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?

1 / 1
points

Concepts

- Define notation for a dataset with multiple features
  - **Multiple Features (00:28)**
- Apply the normal equations to fit linear regression on a dataset
  - **Normal Equation (03:50)**

3.
Suppose you have $m = 23$ training examples with $n = 5$ 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?

1 / 1
points

Concepts

- Choose when to use the normal equations or gradient descent
  - **Normal Equation (11:28)**

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 / 1
points

Concepts

- Explain why features on a similar scale can help gradient descent converge faster
  - **Gradient Descent in Practice I - Feature Scaling (00:13)**
- Identify that redundant or too many features can lead to non-invertibility when using the normal equations

Normal Equation Noninvertibility (02:02)

5.

Which of the following are reasons for using feature scaling?