

Homework 12 (20 Points)

Submit:

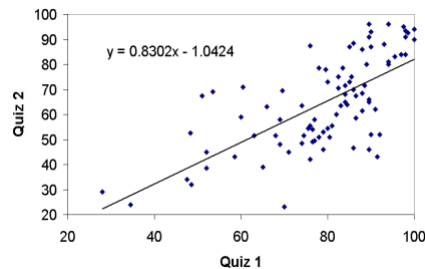
Code (.m file)

A flowchart of the code in your loop function

Short video describing your code

Description:

In this homework, we are going to use linear regression to fit a line to a set of data. Linear regression is used in a variety of different areas to describe the relationship between two variables. For instance, I might want to find the relationship between test scores on the first test and test scores on the second test.



Given the x (quiz 1) and y (quiz 2) data, I can find the line that best describes the relationship. I could then use this to set the curve for the class for example. The equation for the line we would like to fit to the data is:

$$y = mx + b$$

where

$$m = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n (x_i)^2 - \left(\sum_{i=1}^n x_i \right)^2}$$
$$b = \frac{\sum_{i=1}^n y_i - m \sum_{i=1}^n x_i}{n}$$

In this homework assignment you will write a program to create a linear regression equation from a set of data the user enters.

1. First, like the lab, you should use two arrays x and y.
2. Second, create a loop to have the user enter the values of the x and y arrays and to determine the actual number of data points.
3. Using the code from lab, find the mean and standard deviation of x and of y using the formulas. Compare these to what you get out of the built in Matlab functions mean and std (you need to both calculate these with the mean and standard deviation formula and using the built in function).

Mean:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Standard Deviation:

$$stdev = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

- Using a for loop calculate the various sums needed to calculate m and b.
- Calculate m and b.

$$m = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n (x_i)^2 - \left(\sum_{i=1}^n x_i \right)^2}$$

$$b = \frac{\sum_{i=1}^n y_i - m \sum_{i=1}^n x_i}{n}$$

- Display the linear regression equations ($y=mx+b$) to the user and compare it to the results of the polyfit functions ($p=polyfit(x,y,1)$).
- Use a plot to display your data and the line fit

Grading, what are we looking for:

- Code:** The code should get inputs from the user for the array, calculate the means and standard deviations of the two arrays, and then calculate the slope and intercept (m and b) of the linear regression between the two arrays.
- Commenting:** The code should have a header section with a title, your name, and a description. There should also be comments every several lines to describe each step of the code.
- Flow Chart:** We are looking for the steps of the code to be diagrammed, including a correct implementation of the loops for collecting the data and for getting the sums used in the mean, standard deviation and linear regression.
- Video:** You must either submit a video by Thursday or do a demonstration to a GTA or the instructor by Friday to receive any credit for the homework. No video or demonstration is a zero for the assignment.