TOBB ETÜ, BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ

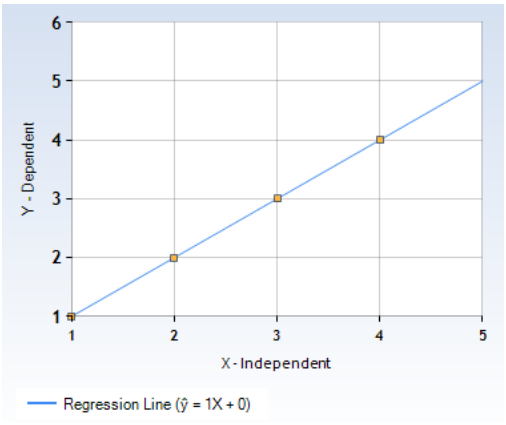
BİL 113/012 Computer Programming I

HOMEWORK 1 (15 Points)

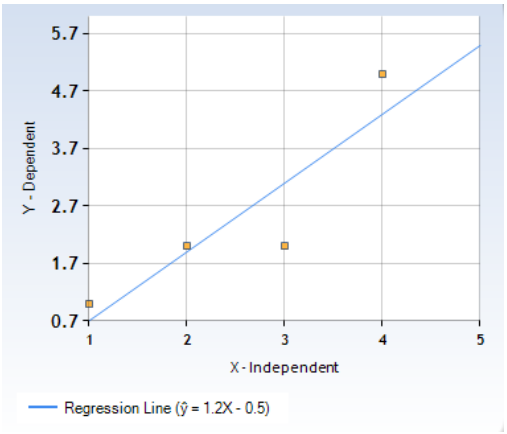
Oct 19, 2020

1 [15 POINTS] REGRESSION ANALYSIS

Regression analysis is a set of statistical processes for estimating the relationships between a dependent variable and one or more independent variables. The most common form of regression analysis is linear regression, in which one finds the line that most closely fits the data according to a specific mathematical criterion. For example, the method of least squares computes the unique line that minimizes the sum of squared distances between the true data and that line. Consider the examples below.



(a) Example I



(b) Example II

Figure 1.1: Least Squares Examples

In the first example, there are 4 points $(1,1),(2,2),(3,3),(4,4)$. As you can see, the line $y = x$ perfectly fits these points and the sum of squared distances equals to 0.

In the second example, there are 4 points $(1,1),(2,2),(3,2),(4,5)$. This time, there are no line that perfectly fits these 4 points and it is not easy to find the best line by looking. But do not worry, as it turns out, we can still find the line that minimizes the sum of squared distance using a closed-form solution that can be easily derived using calculus.

Let (x_i, y_i) represent the i^{th} point. The line that minimizes the sum of squared distance between the line and the points $(x_1, y_1),(x_2, y_2),(x_3, y_3),(x_4, y_4)$ is $y = ax + b$, where

$$a = \frac{4\sum_{i=1}^{4} x_i y_i - \sum_{i=1}^{4} x_i \sum_{i=1}^{4} y_i}{4\sum_{i=1}^{4} x_i^2 - (\sum_{i=1}^{4} x_i)^2} \; and \; b = \frac{\sum_{i=1}^{4} y_i - a\sum_{i=1}^{4} x_i}{4}$$

If you calculate formula above for points from the second example, you will find that $a = 1.2$ and $b = -0.5$. So the line $y = 1.2x - 0.5$ minimizes the sum of squared distances.

In this homework, you will write a Java program which takes 4 points as input and outputs the line that minimizes the sum of squared distances between the line and the points. Inputs will consist of 4 lines and each line will contain 2 numbers which are separated by single space. Your program should find the line using the closed-form formula described above and output the coefficients a and b in the format that complies with the examples below.

EXAMPLE 1

```
Please enter 4 points:
1 1
2 2
3 3
4 4
a:1.0 b:0.0
```

EXAMPLE 2

```
Please enter 4 points:
1 1
2 2
3 2
4 5
a:1.2 b:-0.5
```

EXAMPLE 3

```
Please enter 4 points:
5.3 5.2
17.2 18
3 4
5 4
a:1.0470018146648854 b:-0.18338883681975116
```

Make sure that your programs outputs match the ones above.

There are least square line calculators in the Internet, you can (and should) test your code on other examples. You can calculate the correct answers of these examples using these calculators.

Note: Some of you may encounter an error (specifically java.util.InputMismatchException) on the last example. In this case you can use "," instead of "." for decimal numbers in the input.