Linear Least-Squares Regression

There are many cases in science and engineering where there are **noisy sets of data**, and we wish to estimate the **straight line** which **"best fits"** the data.

This problem is called the **linear regression problem**.

Given a set of measurements (*x*, *y*) that appear to fall along a straight line, how can we find the equation of the line $y = mx + b$ which **"best fits" the measurements**?

If we can determine the **regression coefficients** *m* and *b*, then we can use this equation to **predict** the value of *y* at any given *x* by evaluating the equation $y = mx + b$ for that value of *x*.

A standard method for finding the regression coefficients *m* and *b* is the method of **least squares**.

This method is named **"least squares"** because it produces the line $y = mx + b$ for which the **sum of the squares of the differences** between the **observed** *y* values and the **predicted** *y* values is as **small as possible**.

The **slope** of the least squares line is given by

$$m = \frac{\left(\sum xy\right) - \left(\sum x\right)\bar{y}}{\left(\sum x^2\right) - \left(\sum x\right)\bar{x}}$$

And the **intercept** of the least squares line is given by

$$b = \bar{y} - m\bar{x}$$

Where

- $\sum x$ is the sum of the *x* values
- $\sum x^2$ is the sum of the squares of the *x* values
- $\sum xy$ is the sum of the products of the corresponding *x* and *y* values
- $\bar{x}$ is the mean (average) of the *x* values
- $\bar{y}$ is the mean (average) of the *y* values

Link: https://www.desmos.com/calculator/jwquvmikhr

Design a class named **LinearRegression** that has the following data attributes:

- **m** (least-squares slope)
- **b** (*y*-axis intercept)

The class should also have the following methods:

- **fit**

  The fit method takes a set of measured data points (*x*, *y*), and calculates the values of the least-squares **slope** *m* and *y*-axis **intercept** *b*.

- **predict**

  The predict method calculates the value of *y* at a given point *x* using the equation $y = mx + b$

- **slope**

  The slope method should return the value of the least-squares slope *m*.

- **intercept**

  The intercept should return the value of the least-squares *y*-axis intercept *b*.

i. **Write** a program which will calculate the least-squares **slope** *m* and *y*-axis **intercept** *b* for a given set of measured data points (*x*, *y*).

ii. **Use** the following **dataset** to find a straight line that best fits the data.

iii. **Predict** the value of *y* at *x* = 6.

x	1	2	3	4	5
y	5	7	9	11	14

Optional Enhancement: create a module named **regression** and put your code inside the module.