**LAB 1 - Simple Linear Regression**

In this lab, we will perform simple linear regression in Python. To do this, we will need the .csv provided in the LAB1 folder.

The .csv file contains information related to an imaginary soccer team (it is the same .csv file we examined in LAB0). We will investigate the “age” and “experience” columns and see how player age affects player experience. Simple linear regression is going to be the method of choice for this task. Please be observant of following instructions:

- Read the .csv file, extract the “age” and “experience” values in two different lists. The age list is going to be our “x”, and experience list is going to be our “y”. Remember to import the csv library.

- Implement two different methods, where one computes and returns the regression coefficients (explained below), and the other plots the results. In the script (or main method), read the .csv file, extract the lists and call the corresponding methods in order. The script and the methods should be on the same file.

- When plotting, do a scatter plot using the actual data points, and draw the regression line on top. You can call the plot() function twice and use the show() function to display both plots on the same graph. Both functions belong to the matplotlib library, so don’t forget to import it.

- When extracting the lists, don’t use standard lists. Use the numpy library to create an array with the array() function, then use the append() function to add items to the array. We will use numpy arrays since it is much easier to do element-wise operations with them, unlike standard lists, so again don’t forget to import numpy.

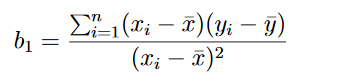
- Any other packages besides “csv”, “matplotlib” and “numpy” are strictly forbidden for this lab. In the future, when the mathematical models get more complicated, we will use packages which make our lives easier.

- Regression coefficient calculation: The eventual aim is to find two coefficients (b0,b1) such that:

h(x) = b1x + b0,

where h(x) (sometimes also denoted by y-hat) is our estimation using x as our independent variable (The “age” data in our case). This line is later going to be plotted.

The calculation of b0 and b1 consists of simple mathematical equations, where:







refers to the *average* values of the *x* and *y* datasets, respectively. After calculating the coefficients, pass them as a parameter to the method which plots the results.

In the plot method, do one scatter plot (you can call the “scatter” function, or the “plot” function with an extra parameter “o”), plotting *y* (the “experience” dataset) against *x* (the “age” dataset), and then plot the regression line by plotting h(x) against *x*. Use the “plot” function for this line. This should be the resulting plot:

