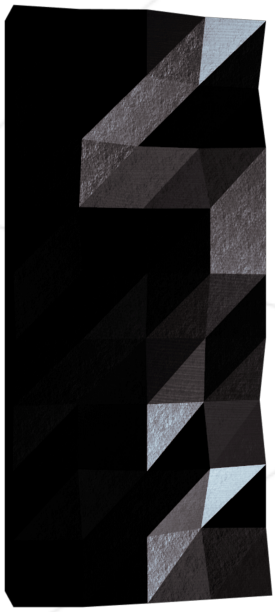


Piscine Pro AI / Machine Learning

Simple Linear Regression 2

Summary: In this Module, you will learn about Simple Linear Regression using the gradient method.

Version: 1.00

Contents

Chapter I

Introduction

Welcome !

If you haven't already done so, read en.toolkit.pdf.

What this Module will show you:

In this module, you will delve into the practical application of simple linear regression using foundational libraries and tools. While our approach will differ from previous offerings, where we utilized the scikit-learn library, this time we will exclusively rely on mathematical computations using numpy. You will gain a hands-on understanding of data manipulation techniques, as well as how to calculate regression parameters and predictions using numpy's capabilities. By embracing this math-centric approach, you'll strengthen your grasp of the underlying principles and broaden your skills in manual implementation, ensuring a comprehensive comprehension of simple linear regression using the gradient method.

Good luck to all.

Chapter II

General instructions

Unless explicitely specified, the following rules will apply every day of this Piscine Pro.

- This subject is the one and only trustable source. Don't trust any rumor.
- This subject can be updated up to one hour before the turn-in deadline.
- The assignments in a subject must be done in the given order. Later assignments won't be rated unless all the previous ones are perfectly executed.
- Be careful about the access rights of your files and folders.
- Your assignments will be evaluated by your peers.
- You must not leave in your turn-in your workspace any file other than the ones explicitly requested By the assignments.
- You have a question? Ask your left neighbor. Otherwise, try your luck with your right neighbor.
- Every technical answer you might need is available in the `man` or on the Internet.
- By Thor, by Odin! Use your brain!!!

Chapter III

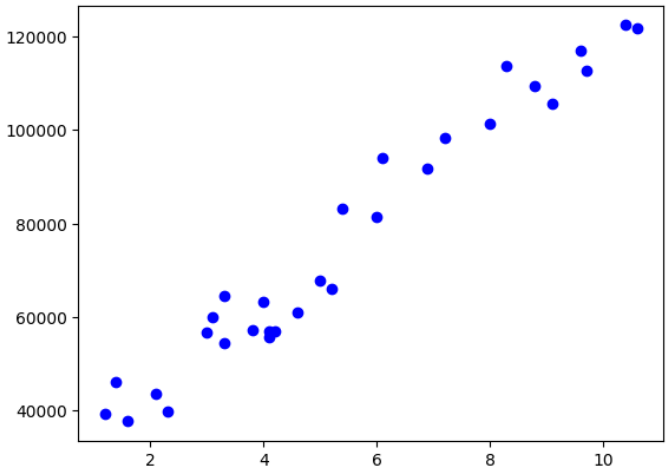
Exercise 00

	Exercise 00
Displaying your data	
Turn-in directory : $ex00/$	
Files to turn in : `Beginner01.ipynb`	
Allowed functions : `All exept pandas and seaborn`	

As in the previous module, in this first exercise you'll be asked to display data using a library other than seaborn.

It's up to you to find out for yourself what exists.

You should have something like this:



Chapter IV

Exercise 01

	Exercise 01
Train and Display	
Turn-in directory : $ex01/$	
Files to turn in : `Beginner01.ipynb`	
Allowed functions : `All exept pandas, seaborn`	

You can consult the mathematical explanations on Wikipedia, or search the Internet to find out how to use numpy for linear regression.
https://en.wikipedia.org/wiki/Linear_regression

The formula for a simple linear regression is:

$$y = \beta + \beta X + \epsilon$$

- y is the predicted value of the dependent variable (y) for any given value of the independent variable (x).
- B0 is the intercept, the predicted value of y when the x is 0.
- B1 is the regression coefficient – how much we expect y to change as x increases.
- x is the independent variable (the variable we expect is influencing y).
- e is the error of the estimate, or how much variation there is in our estimate of the regression coefficient.

`For the time being, we'll ignore the error of the estimate`

For this second exercise we're going to get down to business: you'll have to train your first model.
You have then display your linear regression bar on your first graph with an increasing number of iterations.

You should have something like this:

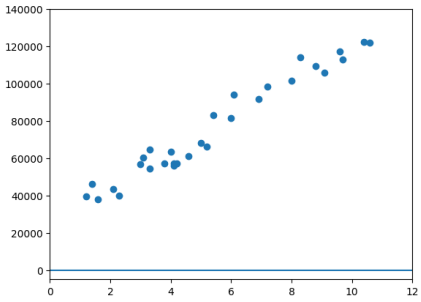


Figure IV.1: num_iters = 0

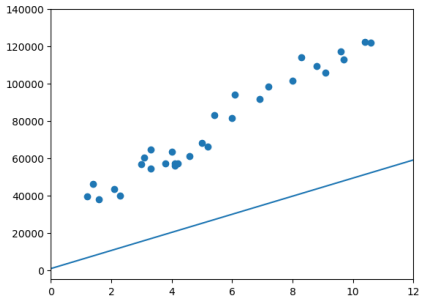


Figure IV.2: num_iters = 1

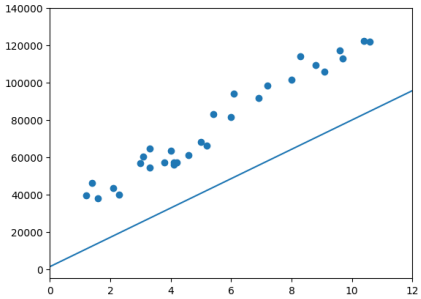


Figure IV.3: num_iters = 2

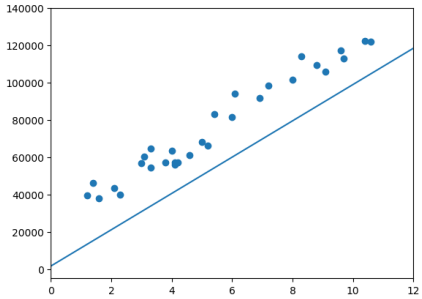


Figure IV.4: num_iters = 3

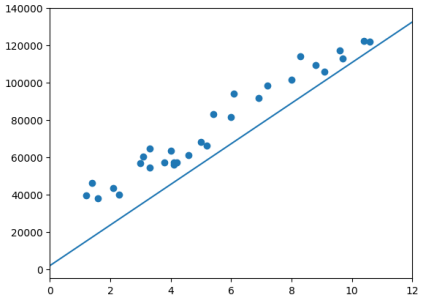


Figure IV.5: num_iters = 4

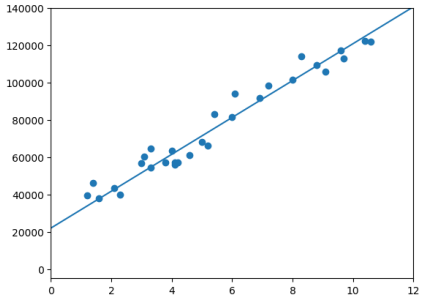


Figure IV.6: num_iters = 1000

`What happens as the iterations progress?`

Chapter V

Exercise 02

	Exercise 02
Predict	
Turn-in directory : $ex02/$	
Files to turn in : `Beginner01.ipynb`	
Allowed functions : `All exept pandas, seaborn`	

Great, now you're going to provide a salary based on the number of years of experience.

You should find a result close to this one with 10 years of experience:

```
Predicted salary for 10 years of experience 127134.91360616997
Predicted salary for 15 years of experience 189547.72611796323
```

`Can you explain the difference in results with your first module?`

Chapter VI

Submission and peer-evaluation

- Create a `professional_training_beginner` folder at the root of your home, and move around in it.
- Create a new `module01` folder and navigate to it.

`Please note, during your defense anything that is not present in the folder for the day will not be checked.`