

Universidade Evangélica de Anápolis

Sistemas Inteligentes: Regressão Linear Simples

Professor: Alexandre Tannus

Felipe Franco

Camila Souza

Anápolis 2022

Dados Analisados: <https://www.kaggle.com/datasets/karthickveerakumar/salary-data-simple-linear-regression>

GitHub: https://github.com/FelpsFranco/Regressao_Linear

Código

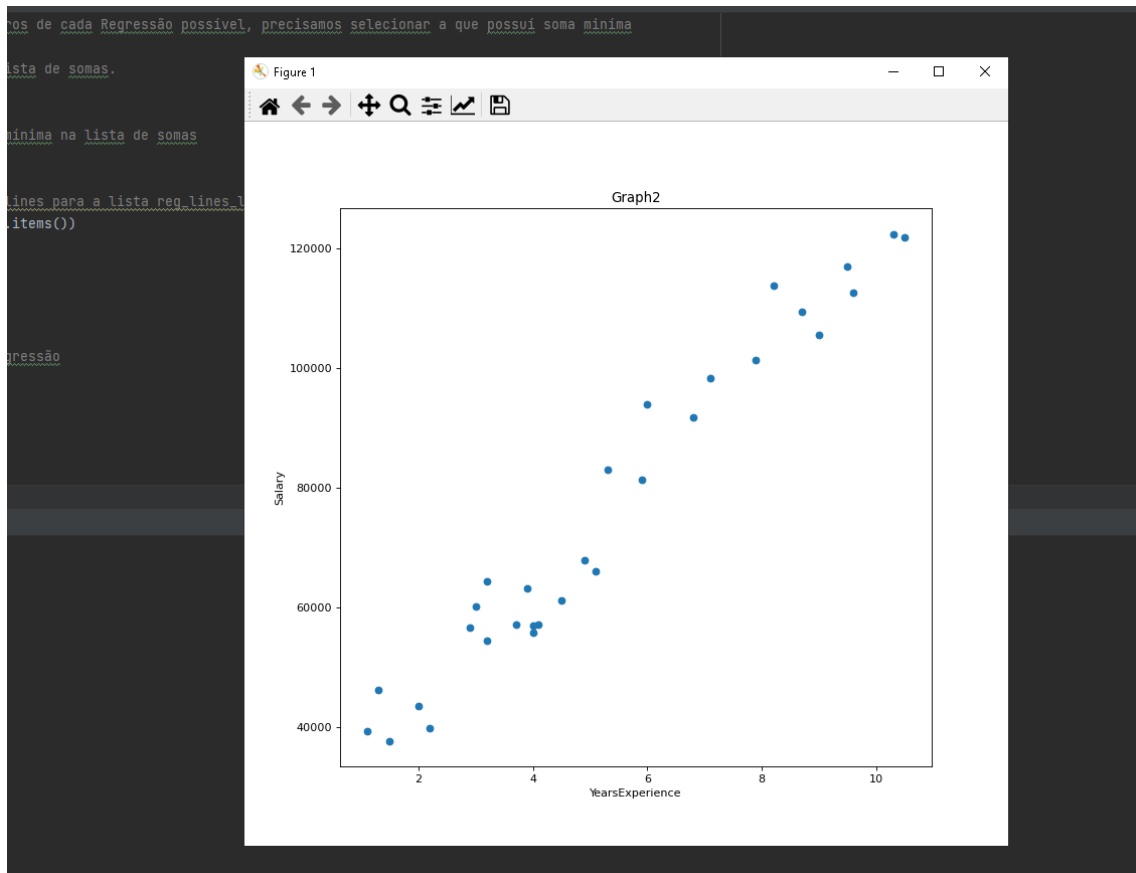
Primeiramente foi realizado a importação das bibliotecas utilizadas para análise dos dados (Pandas e Matplotlib.pyplot), foi criado uma variável que assume o valor da nossa planilha de dados “Dados”, dessa forma conseguimos chamar de uma maneira fácil e resumida essas informações.

Com nossa variável dados conseguimos apresentar dentro da IDLE utilizada “PYCHARM” os dados que estão no arquivo CSV:

	YearsExperience	Salary
0	1.1	39343.0
1	1.3	46205.0
2	1.5	37731.0
3	2.0	43525.0
4	2.2	39891.0
5	2.9	56642.0
6	3.0	60150.0
7	3.2	54445.0
8	3.2	64445.0
9	3.7	57189.0
10	3.9	63218.0
11	4.0	55794.0
12	4.0	56957.0
13	4.1	57081.0
14	4.5	61111.0
15	4.9	67938.0
16	5.1	66029.0
17	5.3	83088.0
18	5.9	81363.0
19	6.0	93940.0
20	6.8	91738.0
21	7.1	98273.0
22	7.9	101302.0
23	8.2	113812.0
24	8.7	109431.0
25	9.0	105582.0
26	9.5	116969.0
27	9.6	112635.0
28	10.3	122391.0
29	10.5	121872.0

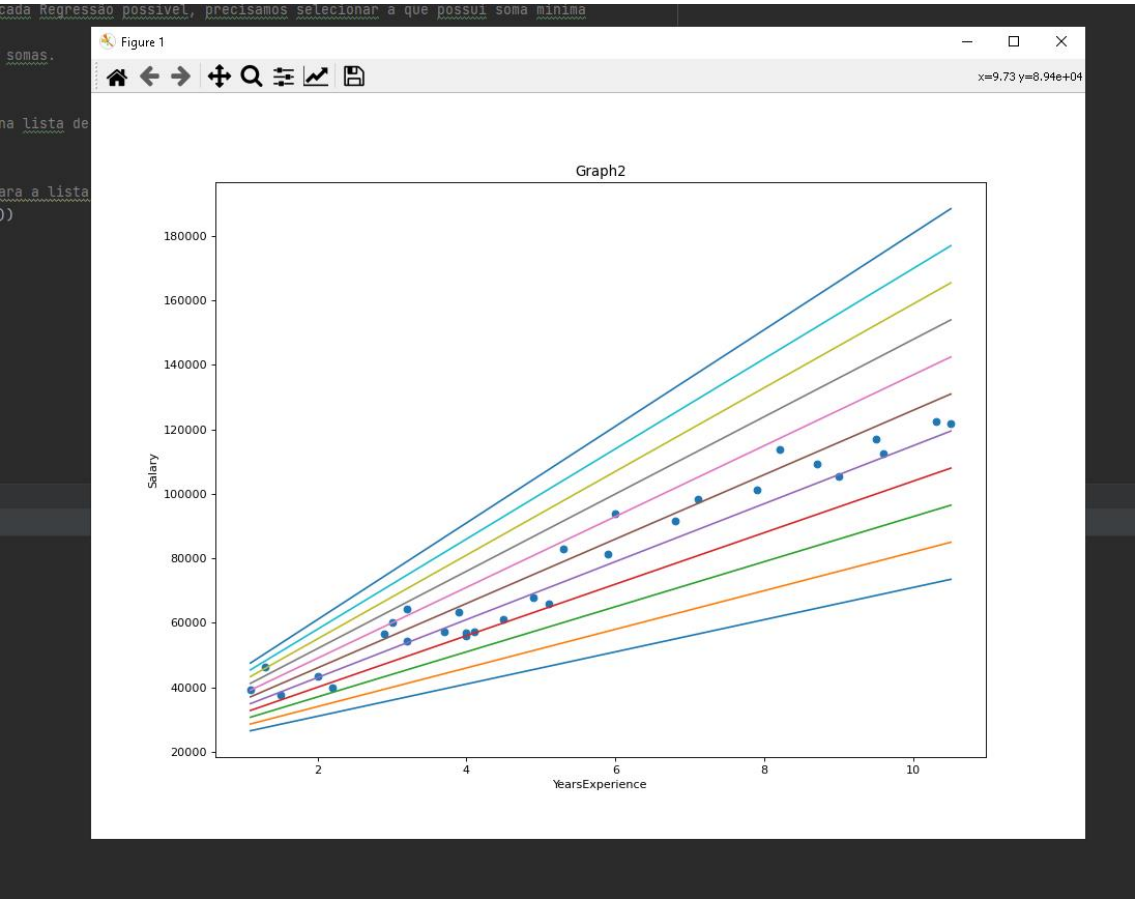
É possível verificar que se tratam dados simples divididos em duas colunas: Anos de Experiência e Salário, porém as informações tratadas não seguem um padrão definido e Fixo. Após isso utilizamos a biblioteca panda a partir do comando **print(pd.insta(dados))** que irá apresentar quaisquer valores nulos como “True” em tela, porém a planilha não conta com nenhum dado dessa forma, não sendo necessário alterações.

Realizado a apresentação dos dados, plotamos os dados em um Gráfico simples, entre os Anos de Experiência e o Salário recebido:



Conseguimos visualizar que os pontos plotados não seguem um padrão, tendo uma grande disparidade de valores, trazendo uma projeção de erros, dessa forma é necessário que seja gerado um novo Plot porém que traga todas as Regressões Lineares possíveis.

Assim conseguimos apresentar como os valores podem se portar durante os anos de experiência, trazendo uma possível margem de erro.



Com essas informações de possíveis Regressões, conseguimos calcular a quantidade de erros de cada uma, afim de utilizar a regressão com menos erros Mínimos, chegando ao Gráfico:

