

matplotlib LPA

Fire risk ¶

Population ¶

The economy of a country can be characterized by the number of inhabitants and the average salary of these inhabitants. But are these two variables linearly related? To analyze the relationship of the two variables, it is necessary to find the parameters of the line, m and b , such that $y = mx + b$.

For the estimation of the parameters of the line, consider the function `np.polyfit` and to calculate the values of the line, the function `np.polyval` or `np.poly1d`.

For example, given sequences of records with population and salary:

6.1101	17.592
5.5277	9.1302
8.5186	13.662
7.0032	11.854
5.8598	6.8233
8.3829	11.886
7.4764	4.3483
8.5781	12
6.4862	6.5987
5.0546	3.8166
5.7107	3.2522
14.164	15.505
5.734	3.1551
8.4084	7.2258
5.6407	0.71618
5.3794	3.5129
6.3654	5.3048
5.1301	0.56077
6.4296	3.6518
7.0708	5.3893
6.1891	3.1386
20.27	21.767
5.4901	4.263
6.3261	5.1875
5.5649	3.0825

The result of the parameters of the line, m and b respectively, should be as follows:

1.252078 -1.821804

Display the scatter plot and the line with the estimated parameters. for the test case

This is actually fitting a linear regression into your data by using numpy.

Centroids ¶

Three points, designated centroids, are located at the following positions: $[5.0, 0.0]$, $[45.0, 70.0]$, $[50.0, 90.0]$. Centroids can be identified by : 0,1,2 .

From a disorganized list of points, it is intended to classify each point in the respective centroid. To choose the correct centroid you should choose the nearest centroid to your point.

For example, given a sequence of points with *x,y coordinates* :

```
15 16
16 18 18.5
17 20.2
16.4 17.12
17.23 18.12
43 43
44.43 45.212
45.8 54.23
46.313 43.123
50.21 46.3
99.22
100.32 98.123
100.32 97.423
102 93.23
102.23 94.23
```

The ranking result should be as follows:

```
[0 0 0 0 0 1 1 1 1 1 2 2 2 2 2]
```

display the scatter plot by giving one different color to each group of points.

CCT graphic ¶

Recall the CCT data used earlier.

Let's try to do a simple analysis of the data.

Show in a graph the evolution of closing prices over 10 years. In the same graph show the evolution of SMA 30 (sma9 was calculated in numpy practical) .

Avocado graphic ¶

Recall the avocado data used earlier.

Filter the data to get only Atlanta data. Add the expensive, cheap label as done in the lectures. Use seaborn to plot the pairplots using the attribute cheap, expensive.