

# Python プログラミング入門資料 3

中田 昌輝

2021 年 2 月 9 日

## 問題 4

関数について慣れよう。再帰関数について取り扱う。

(1) 再帰関数の漸化式が以下のように与えられるとする。

$$\begin{cases} a_1 = 1, & a_2 = 1 \\ a_{n+2} = a_{n+1} + a_n \end{cases}$$

入力  $N$  に対して  $a_N$  を出力せよ。つまり, 入力形式

$N$

のとき, 出力形式

$a_N$

である。  $N < 20$  という制約のもと計算してよい。ちなみに  $a_{10} = 55$  である。

(2) 書き方を工夫して  $a_{100}$  を求めよ。

## 問題 5

引き続きファイル入出力およびライブラリの使用方法に慣れていこう。ライブラリが自分のパソコンに入っていない場合は `pip install ...` という形式でインストールすることができる。最小二乗法をメインとして扱う。最小二乗法の概要については `Optimisation and Machine Learning` のファイルを参照。

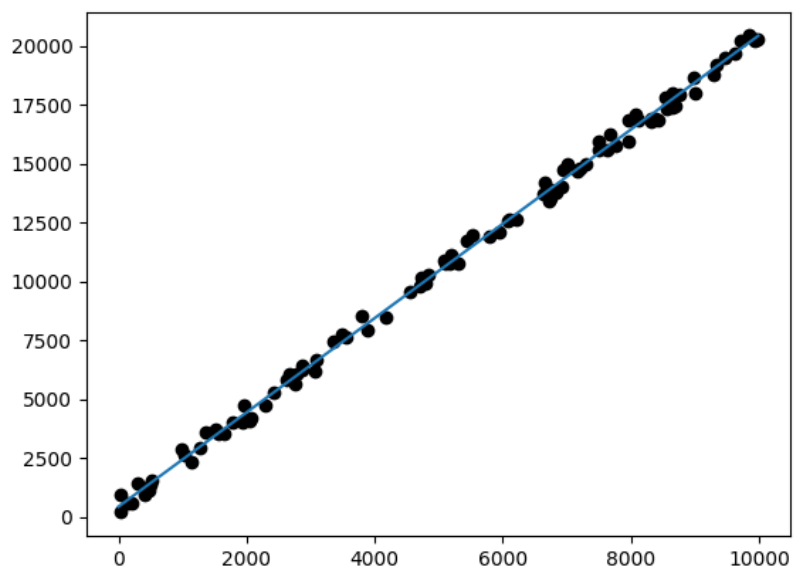
(1) `data2.txt` を読み込んで, 最小二乗法で計算した  $y = ax + b$  についてのグラフを出力せよ。また傾き  $a$  について出力せよ。ここで与えられる  $x$ ,  $y$  は浮動小数点数であることに注意すること。ファイル内部の入力形式は以下のように空白区切り及び改行で与えられる。

$$\begin{array}{ll} x_1 & y_1 \\ x_2 & y_2 \\ & \vdots \\ x_n & y_n \end{array}$$

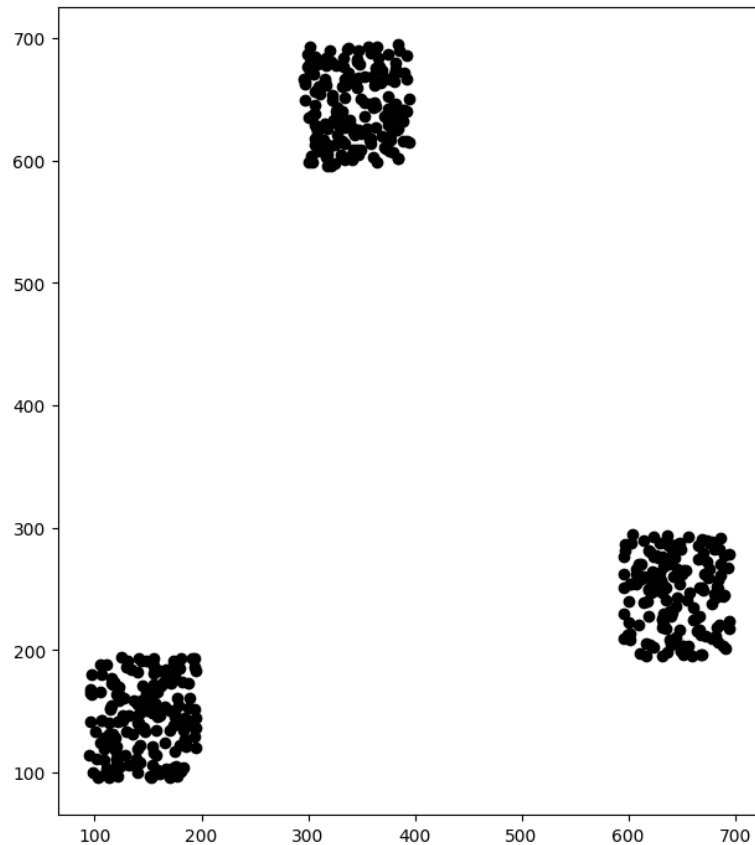
出力形式は

$a$

とする. このときグラフについても出力を行え. ちなみに `sample2.txt` を読み込んだ際は  $a = 2.002767527987209$  となる. また, グラフ出力は以下.



(2) 以下のグラフは `data3.txt` を読み込んだ際のプロットになる.



グラフをみてわかるとおり、3つの集団となっている。このとき、座標をすべて保持していくよりはそれぞれの集団の中心となる座標を保持している方が遥かにデータ容量が少なくて効率が良い。したがって、ここから代表的な1点を選ぶプログラムを書きたい。今回は K-means 法を題材にした。各々の代表値 (集団の中心) を以下のようにして決める。

- (0,0) との距離が小さい順に並べる。
- $n$  個あった際に小さい順に  $n/4$ ,  $2n/4$ ,  $3n/4$  番目となる位置の点を集団の中心とする。

このとき集団の中心の座標を求めよ。

(3) 続いて精度よく中心となる座標を得たい。以下の処理を行った場合の各集団の個数を求めよ。

- ある点は、3 集団の中心からの距離を求めて一番近い点の集団に属する。

(4) 続いて各集団において以下の処理を行う。

- 集団の座標の平均座標を求める
- 平均座標と集団の各点の座標との距離を求める
- 平均座標と最も近い点を新たな集団の中心とする

この処理を行った際に選ばれる各集団の中心の座標を求めよ。

(5) 一連の流れを繰り返すと集団の中心の座標が変わらなくなる.

1.  $(0,0)$  との距離が小さい順に並べる.
2.  $n$  個あった際に小さい順に  $n/4, 2n/4, 3n/4$  番目となる位置の点を集団の中心とする.
3. ある点は, 3 集団の中心からの距離を求めて一番近い点の集団に属する.
4. 集団の座標の平均座標を求める
5. 平均座標と集団の各点の座標との距離を求める
6. 平均座標と最も近い点を新たな集団の中心とする
7. すべての集団において中心の座標が変わらない場合はそこで終了
8. 3 に戻る.

流れとしては  $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow \dots \rightarrow 8 \rightarrow 3 \rightarrow 4 \rightarrow \dots \rightarrow 8 \rightarrow 3 \rightarrow \dots \rightarrow 6 \rightarrow 7$  という形になる. このとき,  $3 \rightarrow 4 \rightarrow 5 \rightarrow 6$  の流れは何回したか出力せよ. また, 各集団の中心となる座標を出力せよ.

(6) 3 つの点に圧縮した後, 解凍することで精度良く結果を得たい. つまりすべてのデータを情報として使うのは重すぎる処理であるが, 最終的な結果としてはすべてのデータを反映させたいということである. そのために各集団の個数を (3) に習って求めよ.

(7) 各集団において

- ある点  $p$  は中心からの座標が一番遠く, その距離は  $r$  である.

とする. このとき (6) で求めた集団の個数が中心から距離  $r$  以内に一様分布しているとしてランダムに座標を決定するとする. このときランダムに決定した座標をもとにしてデータを反映させよ. つまり (2) の `data3.txt` を読み込んだ際のプロットと同じようなグラフを出力させよ.