Уравнение Курамото-Сивашински:

$$\frac{\partial u}{\partial t} + \frac{\partial^4 u}{\partial x^4} + \frac{\partial^2 u}{\partial x^2} + u\frac{\partial u}{\partial x} = 0 \text{ , где } u = u(x,t) \text{ (1)}$$

Это PDE(part differential equation) т.е. уравнение содержащее саму функцию и ее частные производные.

Уравнение (1) содержит как линейную, так и нелинейную часть, что осложняет использование явных или неявных методов. Такие задачи решаются методом IMEX(Implicit-Explicit Method):

$$u_t = Lu + N(u) \text{ (2), где L – линейная часть, а N – нелинейная.}$$

Пусть $Lu = -u_{xx} - u_{xxxx}$, а $N(u) = -uu_x$, тогда в соответствии с разностной схемой CNAB2 (CrankNicolson (Trapezoidal rule) Adams-Bashforth 2):

$$u^{n+1} = u^n + \Delta t\left[\frac{3}{2}N(u^n) - \frac{1}{2}N(u^{n-1})\right] + \frac{\Delta t}{2}[Lu^{n+1} + Lu^n]$$

$$u^{n+1} - \frac{\Delta t}{2}Lu^{n+1} = u^n + \Delta t\left[\frac{3}{2}N(u^n) - \frac{1}{2}N(u^{n-1})\right] + \frac{\Delta t}{2}Lu^n \text{ (2)}$$

Но как теперь дискретизировать u по x?

Граничное условие: $u(x,0) = cos\left(\frac{x}{16}\right)\left(1 + sin\frac{x}{16}\right)$, т.к. оно периодическое для дискретизации будем использовать спектральный метод Фурье:

$$f(x) = \int_{-\infty}^{\infty} \hat{f}(k) \cdot e^{2\pi ikx}\, dk,$$

внеся 2π в k для дискретного случая:

$$f(x) = \sum_{k=-m}^{m-1} \hat{f}(k) \cdot e^{ikx}$$

тогда $(L\hat{u})(k) = (k^2 - k^4)\hat{u}(k)$, $N\left(\widehat{u^2}\right) = -\frac{1}{2}\frac{d}{dx}\widehat{u^2} = -\frac{ik}{2}\left(F\left(\left(F^{-1}(\hat{u})\right)^2\right)\right)$, где F – дискретное преобразование Фурье

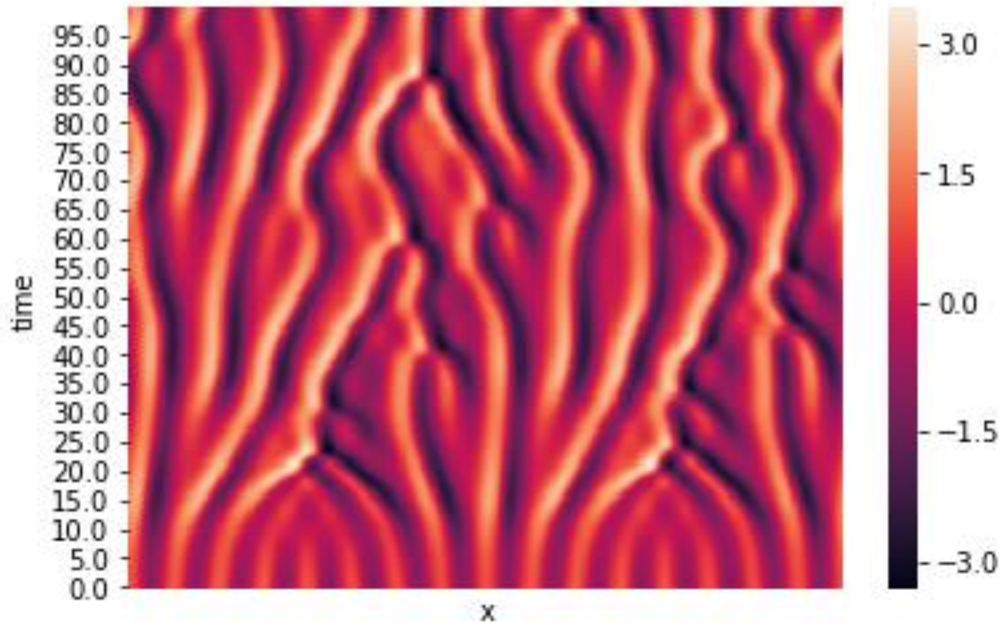
В матричном виде уравнение (2):

$$\left(I - \frac{\Delta t}{2}L\right)\hat{u}^{n+1} = \left(I + \frac{\Delta t}{2}L\right)\hat{u}^n + \frac{3\Delta t}{2}N^n - \frac{\Delta t}{2}N^{n-1},$$

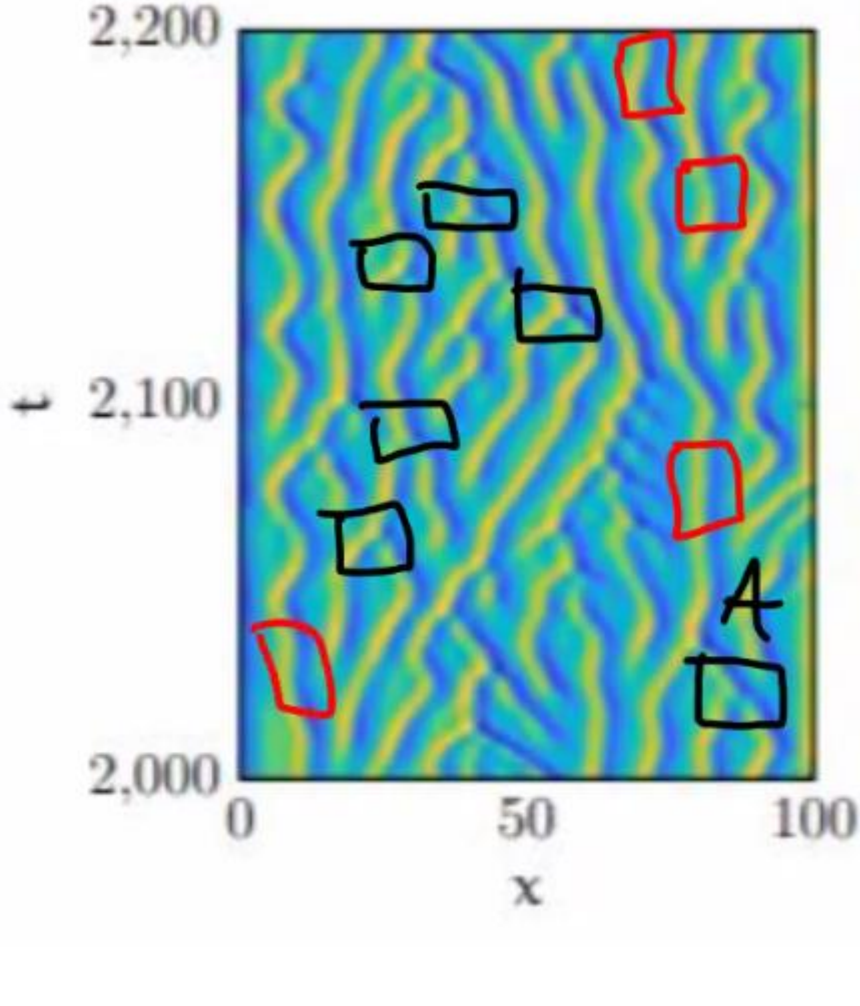
$$\hat{u}^{n+1} = B\left(A\hat{u}^n + \frac{3\Delta t}{2}N^n - \frac{\Delta t}{2}N^{n-1}\right) \text{ (3)}$$

где $A = I - \frac{\Delta t}{2}L$, $B = \left(I - \frac{\Delta t}{2}L\right)^{-1}$. Затем просто находим $F^{-1}(\hat{u})$.

В результате получаем:



Теперь что мы хотим: выявлять “интересные места” динамических систем на примере уравнения Курамото-Сивашинского, примерно так:



В чем проблема: мы хотим, чтобы алгоритм сам в качестве кластеров выбрал черные и красные метки, но не понятно, как этого добиться. В итоге определим эту задачу как задача кластеризации. На примере Рис.1 попытаемся подобрать оптимальный метод кластеризации для динамических систем.

Немного о кластеризации.

В задаче кластеризации обучающая выборка x1, ..., xl состоит только из объектов, но не содержит ответы на них, а также одновременно является и тестовой выборкой. Требуется расставить метки y1, ..., yl таким образом, чтобы похожие друг на друга объекты имели одинаковую метку, то есть разбить все объекты на некоторое количество групп.

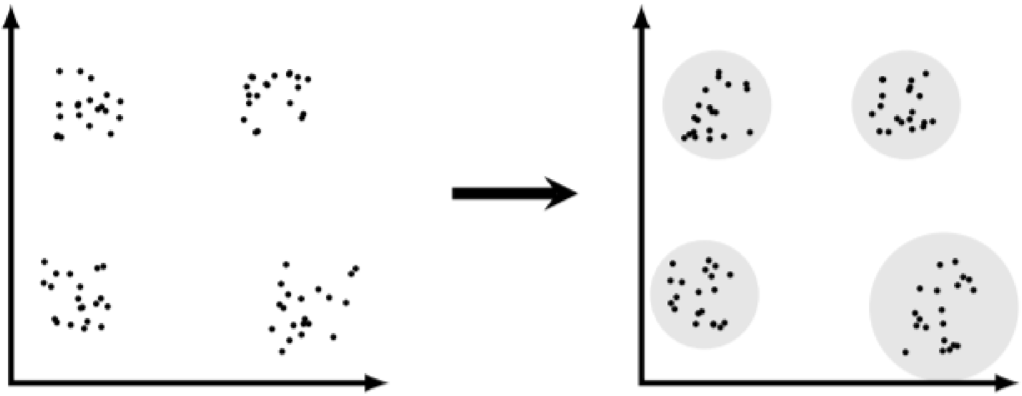


Рис. 1.1: Пример задачи кластеризации

Для начало узнаем наличие кластерной структуры в наших данных с помощью статистики Хопкинса:

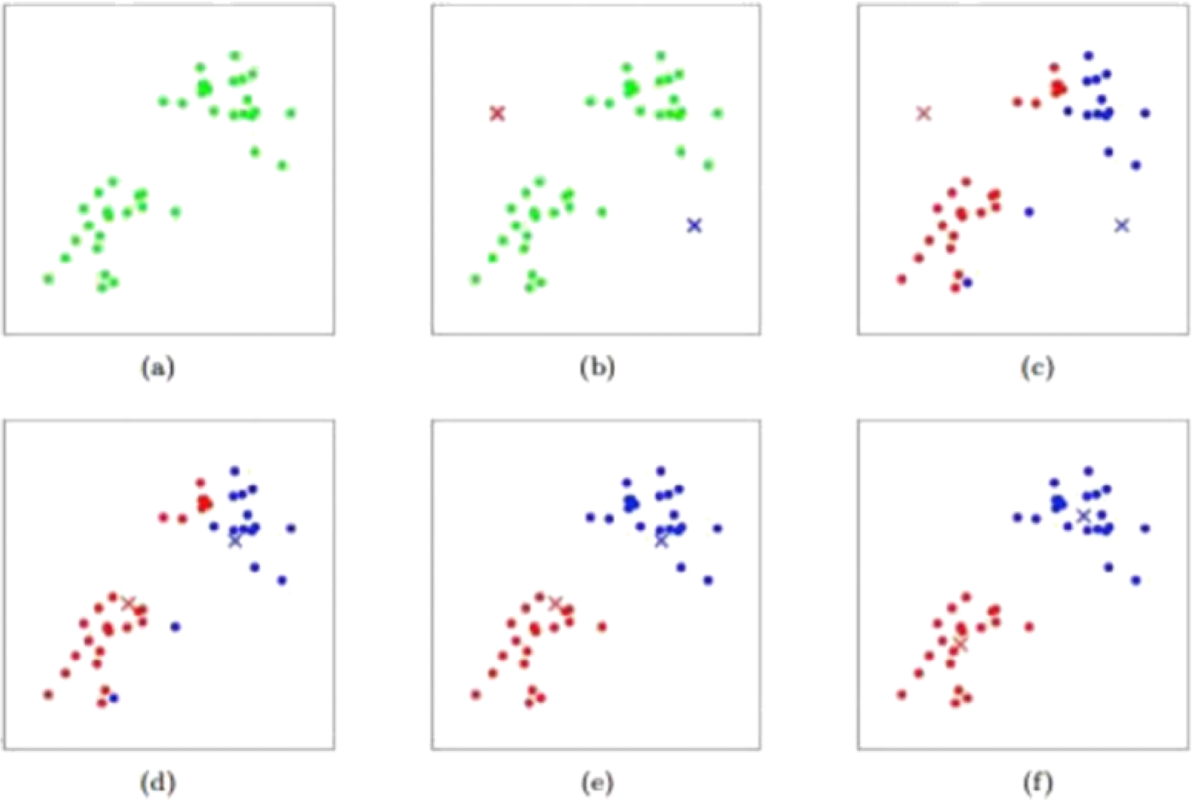
$$H = \frac{\sum_{i=1}^{p} u_i}{\sum_{i=1}^{p} u_i + \sum_{i=1}^{p} \omega_i},$$

Где ω_i – расстояние от i-ой случайной точки до ближайшей случайно, u_i – расстояние от i-ой точки из выборки до другой ближайшей точки из выборки.

Для нашей выборки H = 0.84 – т.е. точки как-то группируются.

Основные методы кластеризации:

K-Means



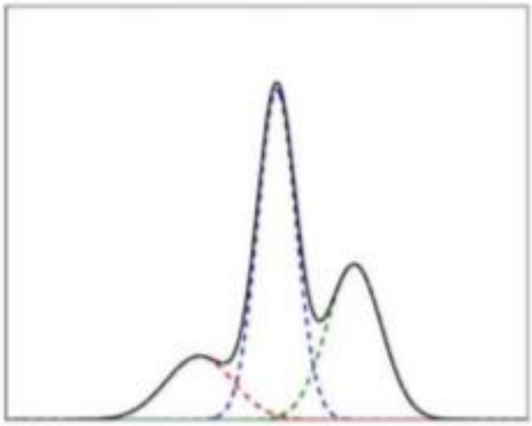
EM - алгоритм

Пусть $\omega_1, \dots, \omega_k$— априорные вероятности кластеров, $p_1(x), \dots, p_k(x)$ — плотности распределения кластеров, тогда плотность распределения вектора признаков x сразу по всем кластерам равна:

$$p(x) = \sum_{j=1}^{k} \omega_j p_j(x)$$

Необходимо на основе выборки оценить параметры модели $\omega_1, \dots, \omega_k$, $p_1(x), \dots, p_k(x)$. Это позволит оценивать вероятность принадлежности к кластеру и, таким образом, решить задачу кластеризации. Такая задача называется задачей разделения смеси распределений:

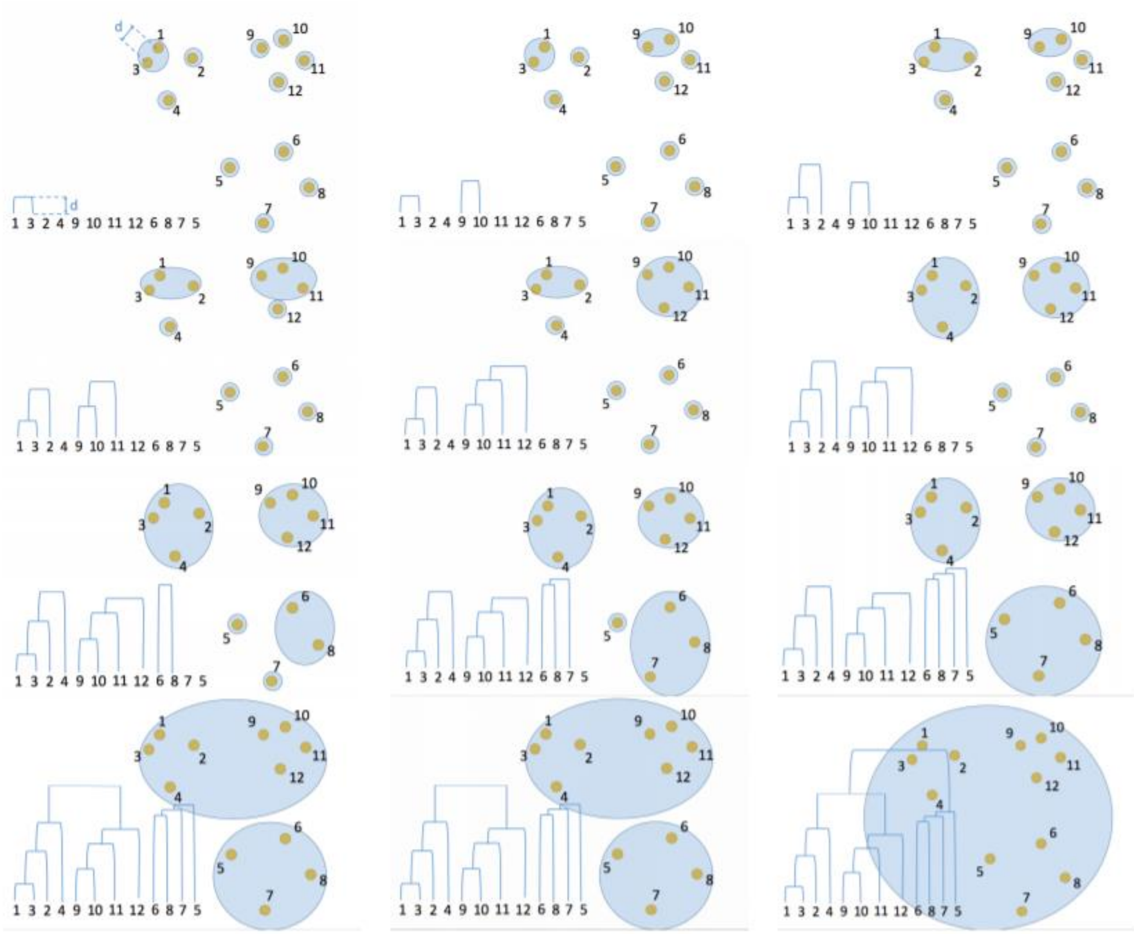
$p_j(x) = \varphi(\theta_j; x)$, где θ_j – параметр распределения $p_j(x)$



Иерархическая кластеризация

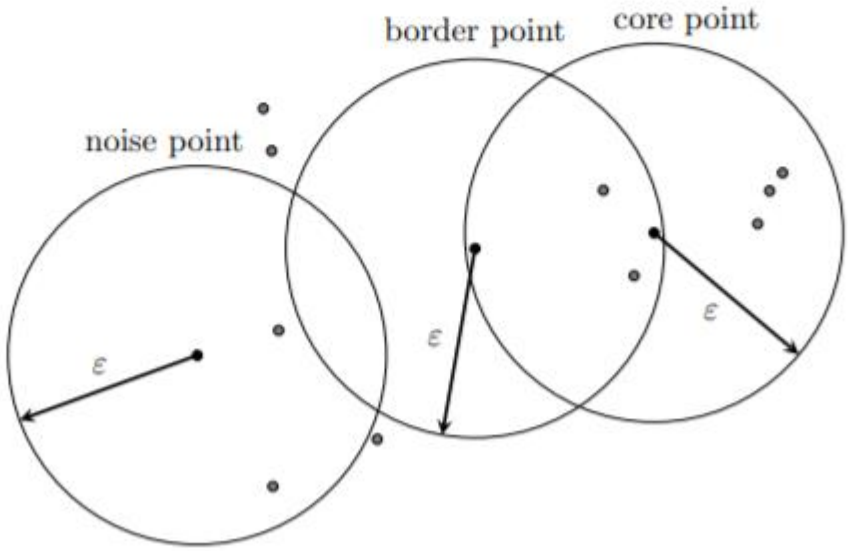
- Агломеративный подход: каждый объект помещается в свой собственный кластер, которые постепенно объединяются.
- Дивизионный подход: сначала все объекты помещаются в один кластер, который затем разбивается на более мелкие

Агломеративный подход:



Метод основанные на плотности

Идея density-based методов заключается в том, чтобы рассматривать плотность точек в окрестности каждого объекта выборки. Если в окрестности радиуса R с центром в некоторой точке выборки находится N или более других точек выборки, то такая точка считается основной. Здесь R и N — параметры алгоритма. Если точек меньше, чем N, но в окрестности рассматриваемой точки содержится основная точка, то такая точка называется граничной. В ином случае точка считается шумовой.



DBSCAN — это один из density-based методов, который состоит из следующих шагов:

1. Разделить точки на основные, пограничные и шумовые.

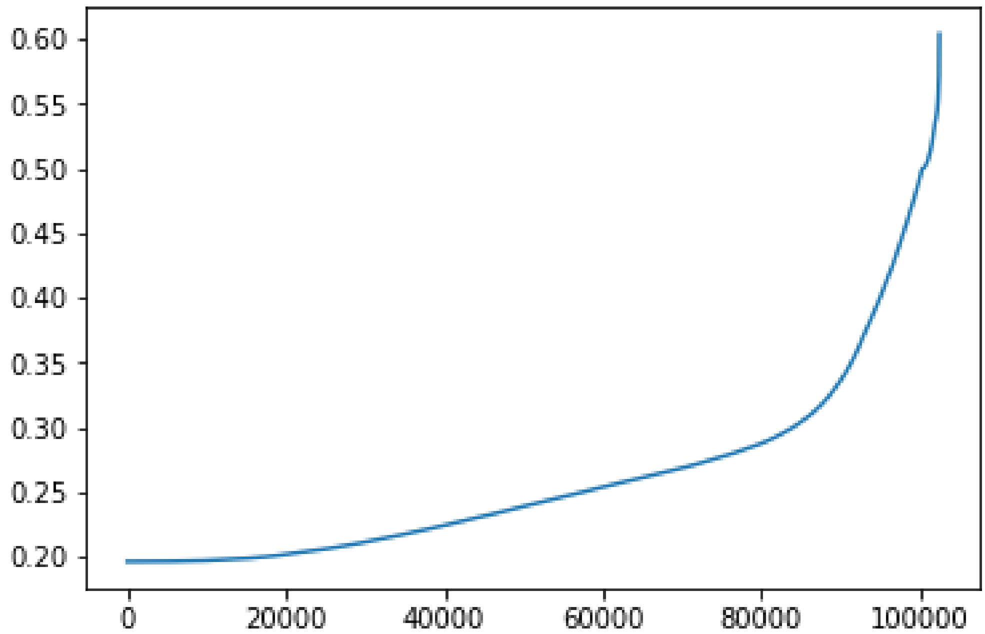
2. Отбросить шумовые точки.

3. Соединить основные точки, которые находятся на расстоянии ε друг от друга. В результате получается граф.

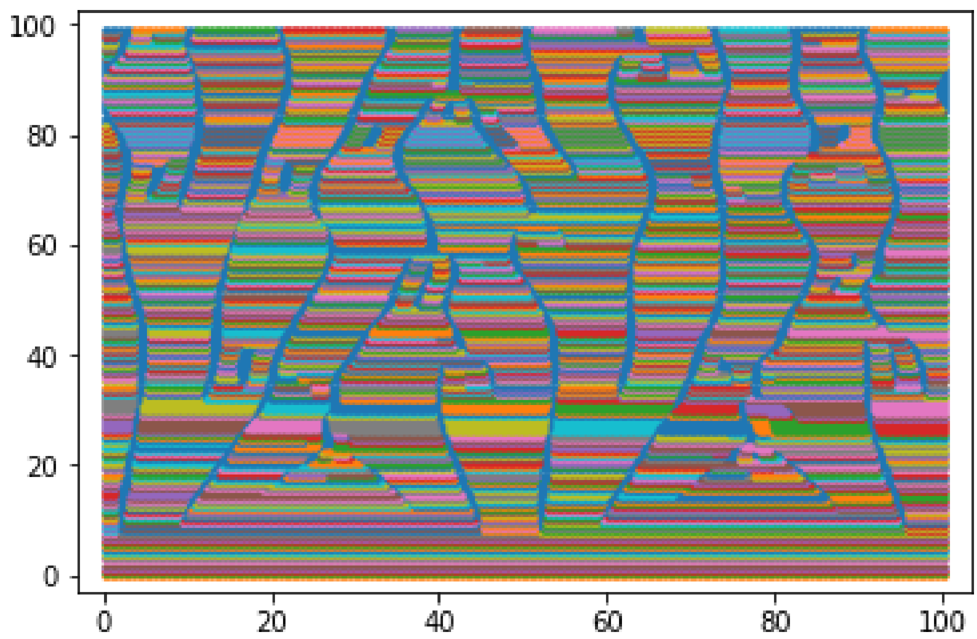
4. Каждую группу соединенных основных точек объединить в свой кластер (то есть выделить связные компоненты в получившемся графе).

5. Отнести пограничные точки к соответствующим им кластерам.

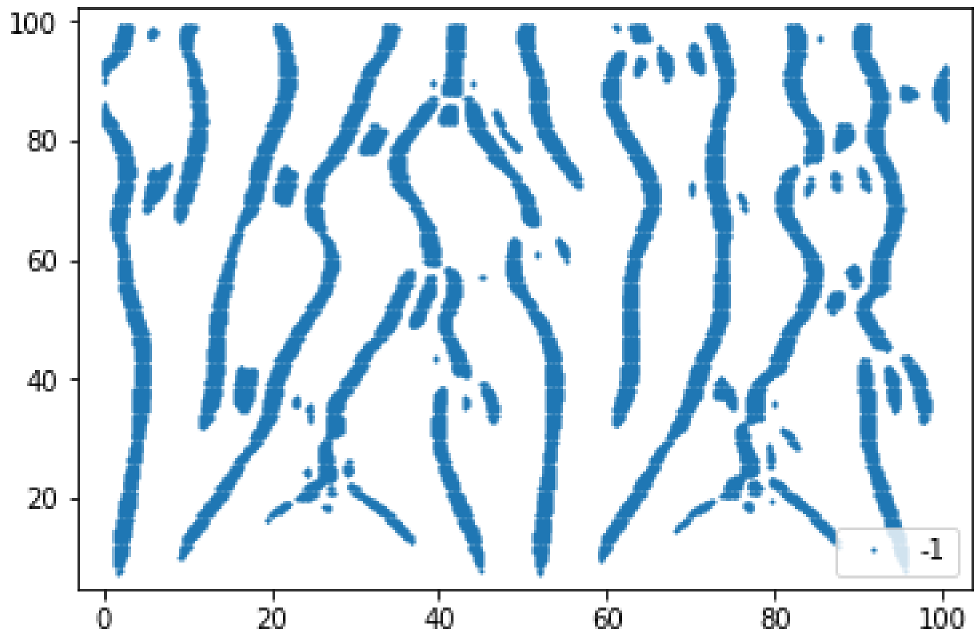
Т.к. DBSCAN может находить кластеры сложной формы, то используем его, чтобы выбрать гиперпараметр R. Для этого построим график, по оси y у которого отложено расстояние до k-го соседа, а по оси x — количество точек, расстояние до k-го соседа соседа у которых меньше. И найдем точку максимальной кривизны:



В результате работы DBSCAN получена кластеризация:

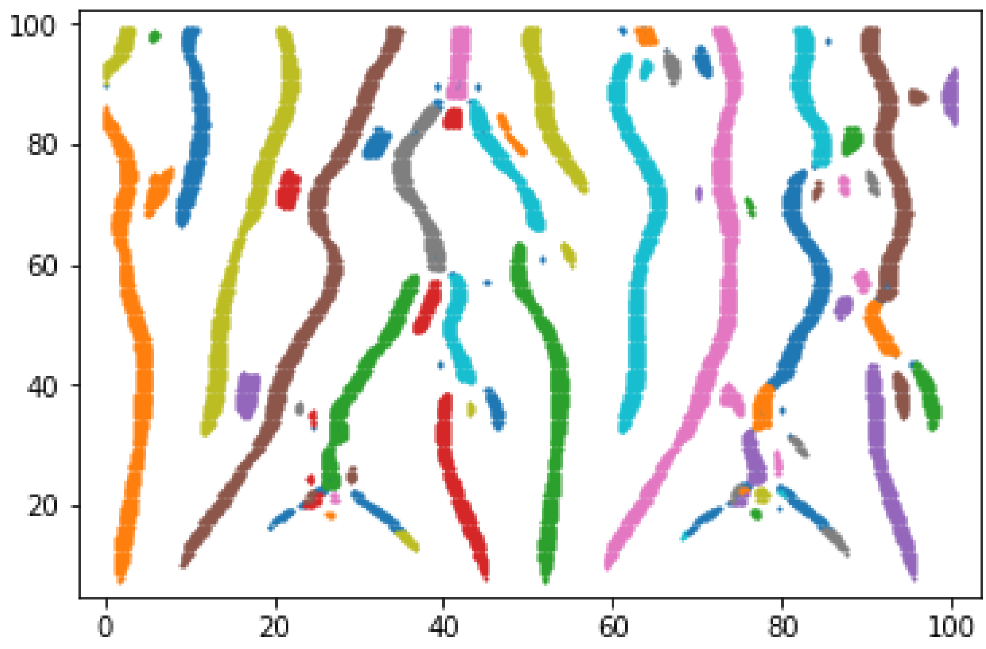


Выведем класс, который помечен, как шумовой:

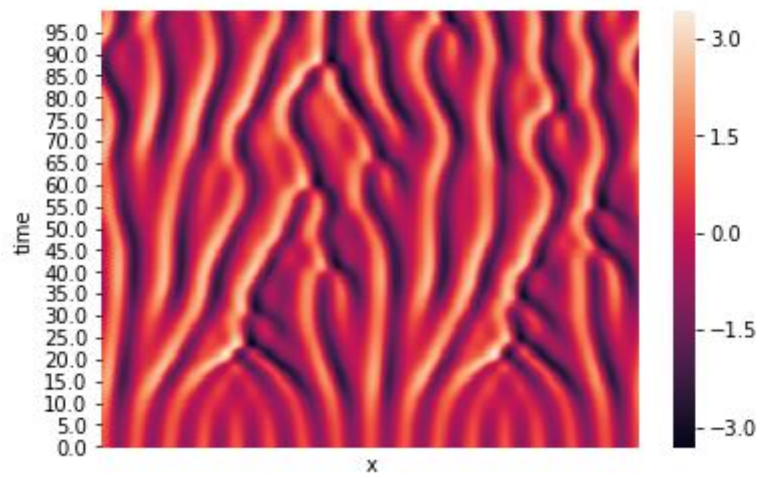


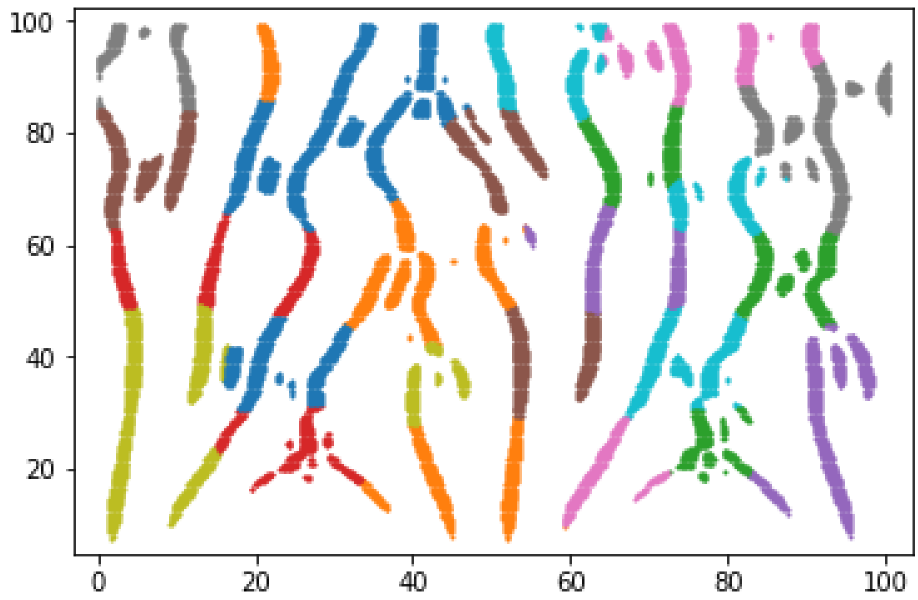
Теперь попробуем провести кластеризацию этих данных.

H = 0.85 – т.е. точки как-то группируются.

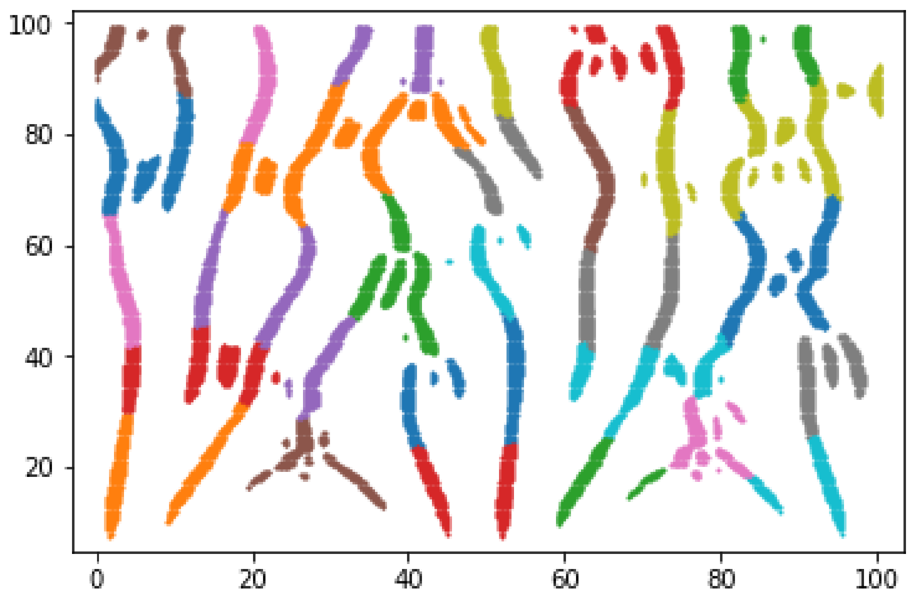


DBSCAN





EM



Агломеративый метод