K-means라는 알고리즘, 방법이 있다. 이 사상의 의미는, 공간 중에 k개의 점이 중심에 모여있고.

그 점들과 가장 가까이 붙어있는 대상으로 분류한다

반복법을 통해, 각 모여있는 것들의 중심의 값을 갱신해라

가장 좋은 모여있는 것들의 결과가 될 때까지 진행하라

가설, 양본 집합을 k개의 종류로 나누고. 알고리즘 묘사는 다음과 같다.

1. K개 종류의 시작 중심값을 지정.
2. I번 째 반복 중에, 임의의 양본에서 k개의 중심과의 거리를 계산, 양본을 분류한다, 중심과의 거리가 제일 짧은 것에 속한 클래스로>
3. 균일값 방식을 이용해 이 클래스의 중심값을 바꾼다（n개 전부 더한 후 n으로 나누기）
4. 모든 k개의 모여있는 것 들의 중심에서 2번과 3번의 반복법으로 중심값 갱신 후, 값이 변하지 않았다면 반복 종료, 값이 변했다면 게속 반복

Java를 이용해 K-means 알고리즘을 작성하라(k는 자신이 정의)

花의 속성에 근거해 Iris.data로 작업 시작, 작업결과를 cluster.txt 문서에 출력해서 문서화 저장.

Iris.data 데이터는 5개의 속성을 가지고 있다.

Sepal length – 꽃받침 세로 길이

Sepal width – 꽃받침 가로 길이

Petal length – 꽃잎 세로 길이

Petal Width – 꽃잎 가로 넓이

Cluster laber – 양본이 어느 클래스에 속해있는지 알려주는 값?

글자판으로 data 개봉 가능? 먼소리야

\*거리는 벡터의 유클리드 거리로 계산하라(일반적인 2개의 점의 거리 계산)

일단 구현해야 하는 기능

1. 데이터 문서 개봉 및 데이터 탐색
2. Txt 문서 생성 및 결과값을 문서에 저장(Iris.data 의 형식처럼 저장할 수 있어야 함)
3. k개의 임의의 초기 시작점 설정(이 실험에서는 4차원 공간의 점)
4. 모든 데이터들에 대해 k개의 점들과의 거리 비교
5. 4번 과정 후에 k개의 점들 중 가장 거리가 짧은 점에 속해지는 과정 필요
6. k개들의 점에 대해 각 점에 속해있는 데이터들의 평균값 계산 과정
7. 5번 후의 평균값을 각 k개들의 점에 대입, k개점의 위치 갱신
8. 4-7번 과정 반복 후, 모든 k개들의 점의 위치가 변하지 않을 때 반복 탈출

예상 문제점

1. 입력은 어떻게 받는가?
2. k개의 점들에 속해지는 과정은 어떻게 처리할 것인가?