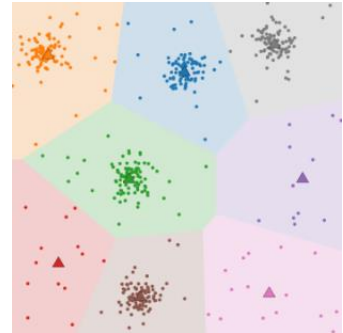


## 문제 1 (학부/대학원 공통): K-평균 알고리즘 병렬화

(배점: 학부 30 점/대학원 20 점)

### 문제 개요

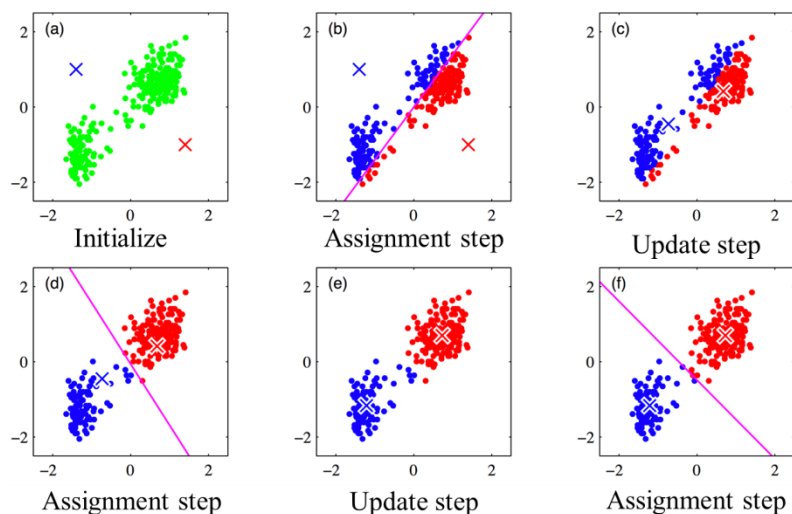
K-평균 알고리즘(K-means algorithm)은 주어진 데이터를 k개의 클러스터로 묶는 방법으로, 각 클러스터와의 거리 차이 분산을 최소화하는 방식으로 동작한다. 본 알고리즘은 자율학습(Unsupervised Learning)의 일종으로, 레이블이 없는 입력 데이터에 레이블을 달아주는 역할을 수행한다. K-평균 알고리즘은 공간데이터 분석(물류센터 위치 선정), 웹문서 분류 등 다양한 분야에서 이용되고 있다.



- 출처 : 위키피디아 -

### 문제 설명

K-평균 알고리즘과 관련한 순차 프로그램이 주어졌다. 크게 assignment step 과 update step 으로 구분되어 있으며, assignment step 에서는 점(point)으로부터 각 클러스터의 중심점까지의 직선 거리를 계산, 그 점에서 가장 가까운 클러스터를 찾아 점을 배당한다. Update step 에서는 각 클러스터에 있는 점들의 무게중심 값으로 해당 클러스터의 중심위치를 재설정해준다. 만약 클러스터의 중심위치가 변하지 않는다면 반복을 중지한다.



## 참고 및 유의 사항

1. 포트란 또는 씨 둘 중 하나의 언어를 선택해서 병렬화한다.
2. util.c 파일과 Makefile 은 변화시키지 않는다.
3. kmeans.c 파일은 수정 가능하다. 단, assignment\_step, update\_step 함수의 경우에는 매개변수 추가는 가능하나 주어진 매개변수는 변경하지 않는다.
4. main.c 파일 2. Execute k-means clustering 부분만 수정 가능하다. 단, do not modify 로 표시된 부분의 순서변경 또는 수정은 허용하지 않는다.
5. 대회 당일 제공하는 입력 파일(input.dat)은 실행파일과 동일한 폴더에 복사하여 이용한다.

## 평가 방법

1. 병렬코드가 32 core 에서 수행되는 시간을 time command 로 측정한다.
2. 실행 결과(각 클러스터의 중심점)가 core 수와 상관없이 동일한지의 여부로 코드의 정확도를 측정하며, 결과가 오차범위( $1e-8$ )내에서 일치하지 않는 경우 정확성 검증 실패로 결격처리한다.