

## Homework 7

Due: 12월 12일 수요일 5시 (uos.stat.hw@gmail.com)

1. K-means 알고리즘 코드를 위 메일로 제출하시오.

- 파일 이름을 반드시 HW7\_고유번호.R로 하시오. 고유번호가 1023 인 경우, 파일이름은 HW7\_1023.R
- 함수 이름은 HW7로 하시오.
- K-means 알고리즘에 들어가는 거리 함수를 유클리드 거리

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_p - y_p)^2}$$

대신에, 다음 함수를 이용하시오.

$$d(x, y) = |x_1 - y_1| + |x_2 - y_2| + \dots + |x_p - y_p|$$

- 반드시 다음 사항을 따르시오.
  - Input: data, centers, max.iter.
  - data : matrix or dataframe.
  - K : scalar(the number of clusters).
  - iter.max : maximum number of iterations.
  - Output: cluster
  - 예) n = 10인 데이터를 K = 2라두고 클러스터링 했을 때, 홀수번째 obs가 1 클러스터로, 짝수번째 obs가 2클러스터로 배정된 경우 1 2 1 2 1 2 1 2 1 2 의 벡터를 Output으로 줘야 함.
  - 알고리즘을 실행중 plot 생성되면, 복사/붙여넣기로 간주하고 무조건 0점.

```
HW7 = function( data, K, iter.max ){
  ...
  return( cluster )
}
```

2. HW7을 이용하여 iris 데이터의 첫 2열의  $K = 2, 3, 4$ 에 대한 클러스터링 결과를 출력하여 미래관 7층 숙제함에 제출하시오. 또한, 각각에 대한 3개의 Scatter Plot 들을 제출하여라. 클러스터마다 색 또는 모양을 다르게 하여 꼭 구분 할 수 있게 할 것, 구분 안 될 시 0점.

```
result = HW7( iris[,1:2], ...)
par(mfrow = c(1,3))
plot(iris[,1:2], col = ...)
```