

# 생물정보학 및 실습 2 실습 보고서 8

2021-20471 박명규

<https://github.com/PMKYU98/BioInfo2>

이번 주차 문제들은 지난 주차까지의 tree 문제처럼 자료 구조를 복잡하게 떠올리지 않아도 되어서 전체적으로 난이도가 어렵지 않았던 것 같다.

## Q1 (ba8a)

FarthestFirstTraversal algorithm

이번 학기 생물정보학 및 실습 2 수업에서 처음으로 패키지를 사용하였다. 제곱근을 계산하기 위해서 `math.sqrt` 함수가 필요했기 때문인데,  $1/2$  승을 계산하는 것보다 훨씬 효율적이라고 해서 채택하였다. [<https://stackoverflow.com/questions/327002/which-is-faster-in-python-x-5-or-math-sqrtx>] 또한 이때까지 input dataset 때문에 데이터가 길어지는 것이 마음에 걸려 input을 받아오는 것도 file parsing으로 대체하였다.

그 외에는 주어진 algorithm instruction과 pseudocode를 구현하였다. 그러나 임의로 첫번째 포인트를 하나의 center로 정하고, 그저 멀리 있는 다른 점을 고를 뿐인데 이 경우 '그냥 전체 dataset에서 양 끝 점을 찾는 것이 아닌가?' 하는 생각이 들었다. 이렇게 나온 답을 center의 optimal solution이 되는지 의문이 들었다.

## Q2 (ba8b)

FarthestFirstTraversal algorithm에서 squared error distortion 구하기

Center도 주어져 있고 포인트의 위치도 모두 주어져 있어서 거리 계산만 하면 되는 간단한 문제였다.

그러나 이 distortion이라는 개념이 어떤 의미를 가지는지 궁금했다.

## Q3 (ba8c)

Lloyd algorithm for k-means clustering

처음에는 임의의 점으로 시작하지만 점점 solution을 향해 수렴해간다는 점이 Q1의 FarthestFirstTraversal algorithm 보다 훨씬 합리적으로 느껴졌다.