

1. 강의노트 20 페이지 k-means clustering을 구현하시오. (시연가능한 형태로 준비)

- 강의노트에 있는 예시코드를 확장해서 구현 가능 (R)
- 본인의 style로 바꿔서 구현 가능
- Python으로 구현하는 것도 가능

2. 강의노트 38페이지의 k 값에 따라 달라지는 CH index와 WSS를 계산하여 그래프로 표현하여라.

(protein 데이터 사용)

3. Euclidean Distance 외에 다른 종류의 다양한 distance measure (4개 이상)을 조사하시오.

- 어떻게 계산하는지 설명하여보자. (예시를 통해서)
- 장/단점을 비교하여 설명하고, 각 distance measure가 유리하게 적용될 수 있는 사례를 설명해 보자.
- Euclidean distance 외에 2가지 이상의 서로 다른 distance measure를 이용하여 k-means clustering을 구현하여보고, clustering 결과를 비교 분석하여보자. (직접 구현하지 않고, 제공되는 함수나 패키지를 활용하는 것도 가능함)

- 참고: <https://www.datanovia.com/en/lessons/clustering-distance-measures/>

4. DB scan algorithm을 구현하여라 (Optional – 시연가능한 형태 extra credit)