10월1일~10월8일

1. K-means :

k-means 클러스터링을 위한 알고리즘 단계

먼저, X = { x1, x2, x3, ………, xn}을 데이터 점의 집합이라하고, V = {v1 , v2, ……., vc }는 중심 집합이다.

∙ 무작위로 ‘c’ 클러스터 센터를 선택한다.

∙ 각 data point와 클러스터 센터 간의 거리를 계산한다.

∙ 클러스터 중심으로부터의 거리가 모든 클러스터 중심 중에서 최소 인 클러스터 중심까지의 데이터 지점을 분석한다.

∙ 다음을 사용하여 새 클러스터 센터를 다시 계산한다.

‘’는 클러스터의 데이터 포인트 수를 나타낸다.

∙ 각 데이터 포인트와 새로 획득 한 클러스터 센터 사이의 거리를 다시 계산한다.

∙ 데이터 포인트가 재 할당되지 않은 경우 중지하고, 그렇지 않으면 3단계에서 반복한다.

2. K-nearest neighbors을 계산하는 방법에 대한 알고리즘 :

∙ 매개 변수 K= number of nearest neighbors 를 미리 결정한다. 이 값은 우리가 임의로 정한다.

∙ query-instance와 모든 training samples 간의 거리를 계산한다. Distance 알고리즘을 사용할 수 있다.

∙ 모든 training samples의 거리를 정렬하고 K번째 최소 거리를 기준으로 가장 가까운 neighbor을 결정한다.

∙ supervised learning 이므로, K에 해당하는 정렬된 값에 대한 training data의 모든 범주를 가져온다.

∙ 대부분의 가장 가까운 이웃을 예측 값으로 사용한다.