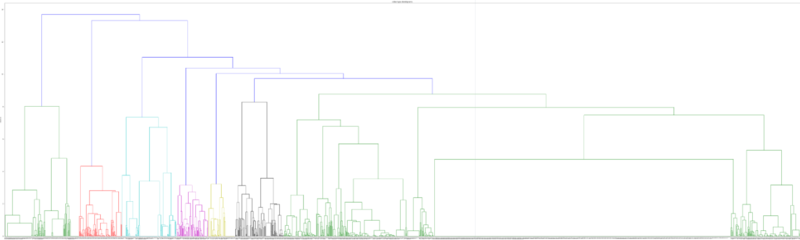
군집분석

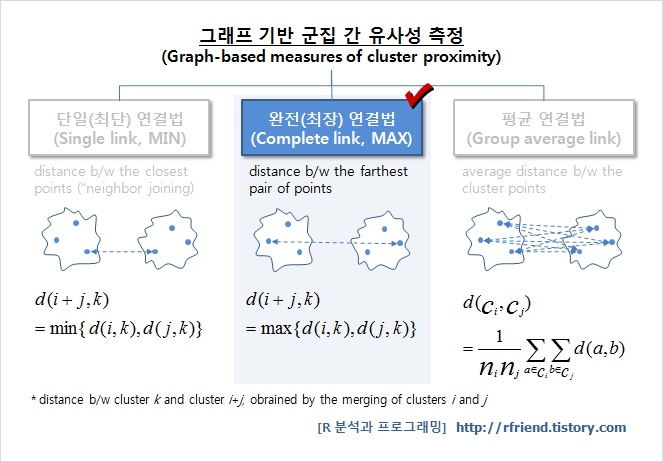
계층적 군집분석 (Hierarchical clustering) : 데이터를 계층적으로 연결해가며 유사도가 높은 군집을 구성해 가는 알고리즘

특성 : k-means(비계층적 방식)과 가장 큰 차이는 최초 클러스터 개수(k) 를 가정할 필요가 없다.

계층도(Dendrogram)으로 적절한 군집 수를 파악할 수 있다. – 가로선을 그어서 원하는 만큼의 군집을 형성할 수 있음



군집 간 거리 측정 방식 (Single, Complete, Average, Centroid Ward Linkage)

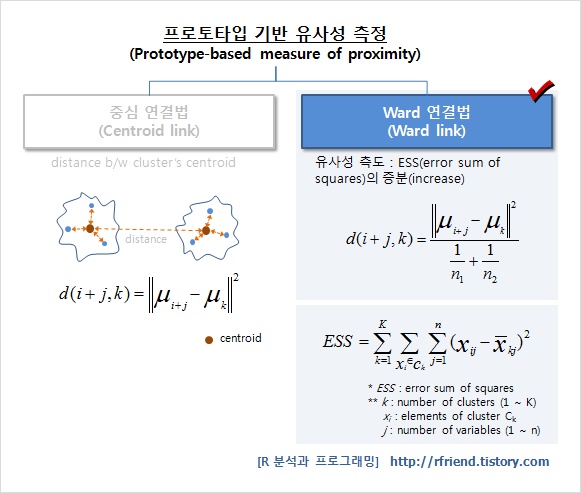


위 3개의 방법은 새로운 점을 그룹에 넣을 때 새로운 점과 기존 그룹의 거리를 재는 방식임.

(예(최장 연결법) : 새로운 점을 그룹에 넣을 때 1그룹에 있는 점과 최장 거리 3, 2그룹의 점과 최장거리 4일 경우 1그룹에 배치

Centroid : 두 군집의(or 1군집과 1점) 중심점을 정의한 다음 두 중심점의 거리를 군집간의 거리로 측정

Ward : 연결될 수 있는 군집 조합을 만들고, 군집 내 편차들의 제곱합을 기준으로 최소 제곱합을 가지게 되는 그룹들과 연결



https://rfriend.tistory.com/203?category=706119