

클러스터 분석의 목적!

- 다변량 데이터가 주어진다. n 개의 관찰 값이 q 개의 변수들에 대해서 주어지는 것이다.

목적: 우리는 클러스터라고 하는 그룹이 존재하는지 안하는지에 대해서 찾고 싶은 것이다.

: 클러스터 내부의 원소들은 서로 매우 비슷하다.

: 다른 클러스터 사이의 원소들은 매우 다르다.

* 클러스터링의 시각화:

↖ 변수 개수
(q)가 2에 대해서 사람의 눈으로 충분히 클러스터링이 가능하다.

- q 가 3개 이상이면, 보통 시각적으로 클러스터들을 찾는 것은 가능하지 않다 -> 클러스터 분석

* 클러스터 분석의 사용:

- 클러스터 분석은 유용한데 아래에 대해서,

: 데이터를 요약 할 때,

: 새로운 통찰력을 얻을 때, 예를 들면,

= 마켓 부분: 다양한 마케팅 목적을 위해서 손님들에 대한 그룹을 찾는 것

= 적합한 치료법을 찾기 위해서 부분 유형의 질병을 찾으려고 환자들을 클러스터링 하는 것