

11.1 11.2 유사성 측도

2019년 6월 16일 일요일 오전 3:12

판별분석 후의 그룹 분류는 이미 분류되어 있는 알려진 군집으로부터 판별함수를 구하고 새로운 관측치가 있을 때 판별함수를 통해 어느집단에 속하는지를 판별하여 분류함.

군집분석에서는 군집의 개수나 구조에 대한 가정없이 다변량 데이터로부터 거리 기준에 의해 자발적인 군집화를 유도함.

군집분석의 목적 1) 적절한 군집으로 나누는 것 2) 각 군집의 특성, 군집간의 차이 등에 대한 탐색적 연구

각 관측벡에 대해 어느 관측벡터와 같은 군집으로 묶일 수 있는지 판단해야함.

관측벡터간의 유사성 또는 근접성을 측정해야함.

유사성에 대한 편리한 측도로 관측벡터간의 거리를 이용함.

거리공식들 p301

11.3 11.4 계층적 & 비계층적 군집 방법

2019년 6월 16일 일요일 오전 3:13

<계층적 군집 방법>

처음에 n 개의 군집으로부터 시작하여 점차 군집의 개수를 줄여나가는 방법

11.3.1 최단연결법 (p302)

두 군집간의 최단거리를 군집간 거리로 정의함.

거리행렬로부터 최단거리의 쌍 U, V 를 찾아 한 군집으로 병합하는것.

거리행렬을 구할때 최단거리를 이용함

11.3.2 최장연결법 (p304)

두 군집간의 최장거리를 군집간 거리로 정의함.

거리행렬로부터 최단거리의 쌍 U, V 를 찾아 한 군집으로 병합함.

거리행렬을 구할때 최장거리를 이용함

11.3.3 평균연결법 (p306)

두 군집에 속한 모든 개체들 간의 거리의 평균을 군집간 거리로 정의함.

11.3.4 ward의 계층적 군집 방법 (p309)

군집간 정보 (편차제곱합 ESS) 의 손실을 최소화하도록 군집화함.

<비계층적 군집 방법>

n개의 개체를 g개의 군집으로 나눌 수 있는 모든 가능한 방법을 점진하여 최적의 군집을 형성하는것이 전략.

K-means method

K-평균법은 전체 개체를 K개의 군집으로 나누는 방법.

한 개체가 속해있던 군집에서 다른 군집으로 이동하는 재배치가 가능함.

더 이상 개체의 군집간 이동이 없을때 최적의 군집으로 결정함.

계층적 군집 방법과 같이 쓰일 수 있음.

11.5, 6, 7 군집개수 결정 & 군집 타당성 & 변수들에 대한 군집

2019년 6월 16일 일요일 오전 3:13

<군집 개수>

계층적 군집 방법에서는 나무구조도에서 수평선을 이용하여 가지를 잘라냄으로써 g 개의 군집을 결정할 수 있음.

이때 병합되는 과정에서 군집간의 거리 차이에 큰 변화를 보이는 경우를 고려하여 군집의 개수를 택함.

ward 방법의 경우 군집개수에 대한 ESS의 증분을 고려하여 급격한 변화가 일어나는 위치에 해당하는 부분에서 군집의 개수를 결정함.

<군집 타당성>

교차타당성을 이용한 방법

데이터를 두개의 부분으로 분류하고 각 부분에서 따로 군집분석을 한 후 합쳐서 군집분석한 결과와 비교하여 비슷하면 결과에 대한 안정성을 볼 수 있다고 판단함.

<변수들에 대한 군집>

n 개의 개체들에 대한 군집문제 외에 p 개 변수들에 대한 군집을 구하고자 할 경우.

변수들 간의 유사성의 척도로 상관계수가 이용되어 변수들의 유사성, 비유사성을 나타내주는 상관계수행렬을 이용해 변수들을 군집으로 나눌 수 있음.