데이터 마이닝

13강군집분석॥

통계·데이터과학과 장영재 교수



❤️ 한극방송통신대학교

01 군집분석 관련 R 함수 02 R 사용 예제



1. 군집분석 관련 R 함수

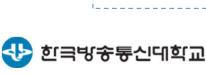


- 1 dist 함수
 - I 함수의 구조 dist(x, method ="euclidean")
 - ▮ 기능

행사이의거리를계산하는 기능을 수행하며, 통상 dist는 hclust 등 다른 명령문에 이용되므로 결과는 비유사성 행렬의 대각선 윗부분이 출력됨. 즉, $n \times p$ 행렬에 대하여 n(n-1)/2개의 원소를 갖는 벡터가 생성

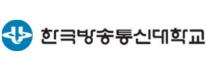
▮ 옵션

- x: 행 사이의 거리를 계산하는 행렬
- method : 거리 계산방법 옵션." euclidean"은 유클리디안 거리를 이용하여 거리를 계산하고, "manhattan"은 맨해튼 거리를 이용하여 거리를 계산. 기본값은 euclidean





- 2 hclust 함수
 - I 함수의 구조hclust(dist,method="complete")
 - ▮ 기능계층적 군집분석 중한 가지 방법인 응집분석을 수행
 - ▮ 옵션
 - dist: 거리 구조 또는 거리 행렬 객체. 보통 dist 함수 적용 결과를 이용
 - method : 응집분석에 따른 계층적 군집화 방법 옵션. single은 단일 연결법, complete는 완전연결법, average는 평균연결법을 이용한 계층적 군집화를 수행. 기본값은 complete

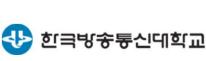




- 3 diana 함수 (cluster 패키지 설치 및 호출 필요)
 - I 함수의 구조 diana 함수 (x, metric = "euclidean")
 - ▮ 기능분할분석에 따른 계층적 군집화를 수행

▮ 옵션

- x: data frame 또는 data 행렬
- metric: 거리 계산방법 옵션. 예를 들어 "euclidean"은 유클리디안 거리를 이용하여 거리를 계산하고, "manhattan"은 맨해튼 거리를 이용하여 거리를 계산. 기본값은 euclidean





- 4 kmeans 함수
 - I 함수의 구조kmeans(x, centers, algorithm="Hartigan-Wong")
 - ▮ 기능 K - 평균 군집분석을 수행
 - ▮ 옵션
 - x: data 행렬
 - centers : K 평균 군집분석을 수행하기 위한 초기값을 가지고 있는 행렬. 각 행은 각 군집의 초기값을 가지고 있어야 하며, K - 평균 군집화 를 위한 군집 수는 centers에서의 행의 수
 - algorithm: K 평균 군집분석에 사용되는 알고리즘. 기본값은 "Hartigan- Wong". 매퀸(1967)에 기초한 알고리즘을 사용하려면 "MacQueen"을 사용



- 5 plot 함수
 - I 함수의 구조plot(object)
 - ▮ 기능일반적으로 object에 대한 그림을 생성
 - ▮ 옵션
 - object : 명령문 hclust나 diana에 의하여 생성된 결과 객체(object)



- 6 cutree 함수
 - I 함수의 구조cutree(tree, k=)
 - 기능 명령문 hclust나 diana에 의하여 생성된 결과 객체(object)를 가지고 주어진 군집수에 대하여 각 개체에 대한 id를 갖는 벡터를 생성

▮ 옵션

- tree : 명령문 hclust 또는 diana의 결과를 가지고 있는 object
- k: 계층적 군집화로부터 얻기를 원하는 군집 수



- 7 table 함수
 - I 함수의 구조table(...)
 - 기능분할표를 생성
 - **▮** 옵션
 - ... : 범주형으로 해석될 수 있는 한 개 이상의 객체(object)



- 8 tapply 함수
 - I 함수의 구조 tapply(x, indices, FUN=)
 - ▮ 기능자료에서 같은 범주에 속한 개체에 대하여 함수의 결과를 산출
 - ▮ 옵션
 - x: 자료행렬
 - indices: 범주를 가지고 있는 리스트
 - FUN= : 함수의 이름을 가지고 있는 문자 string. K 평균을 위해서는 mean 을 사용하면 되고 FUN=은 생략가능



2. R 사용 예제



