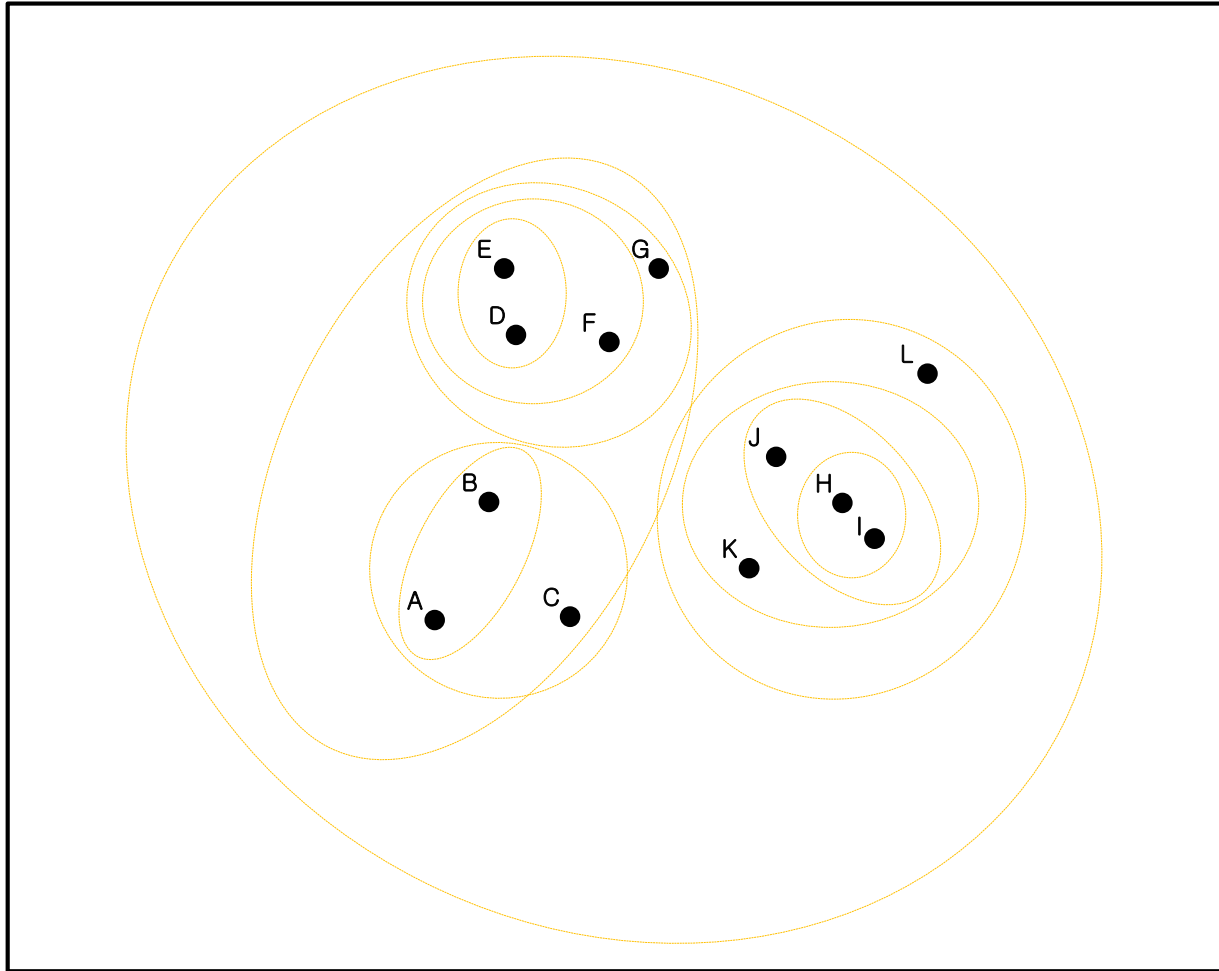




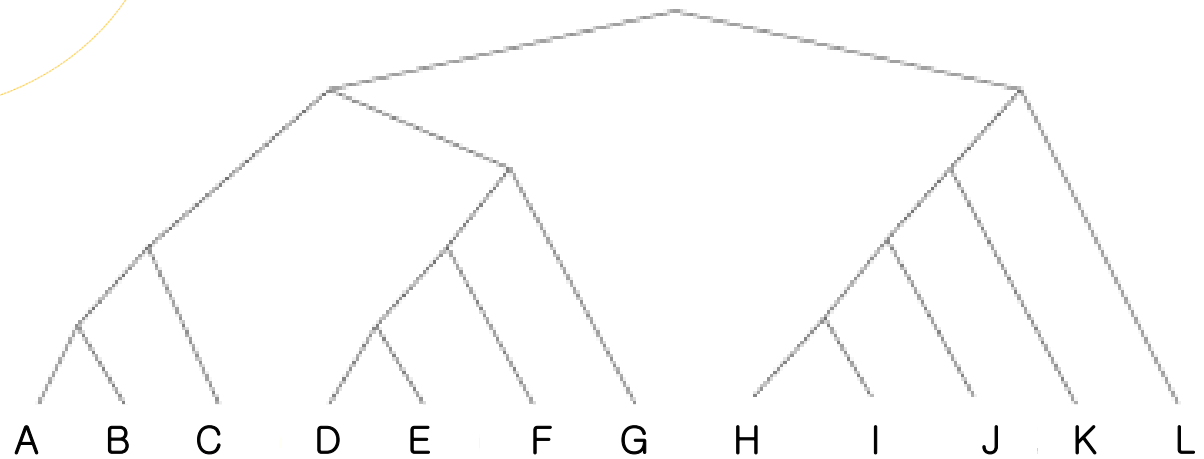
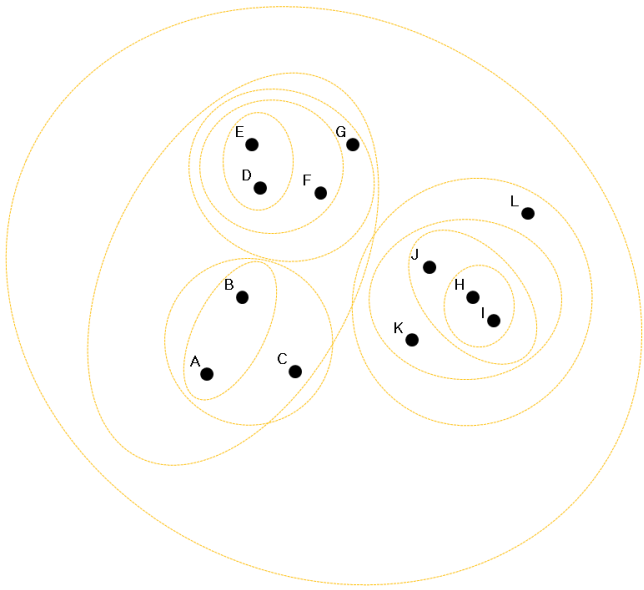
# Lecture 3

## 계층적 군집분석

# 계층적 군집분석의 개념



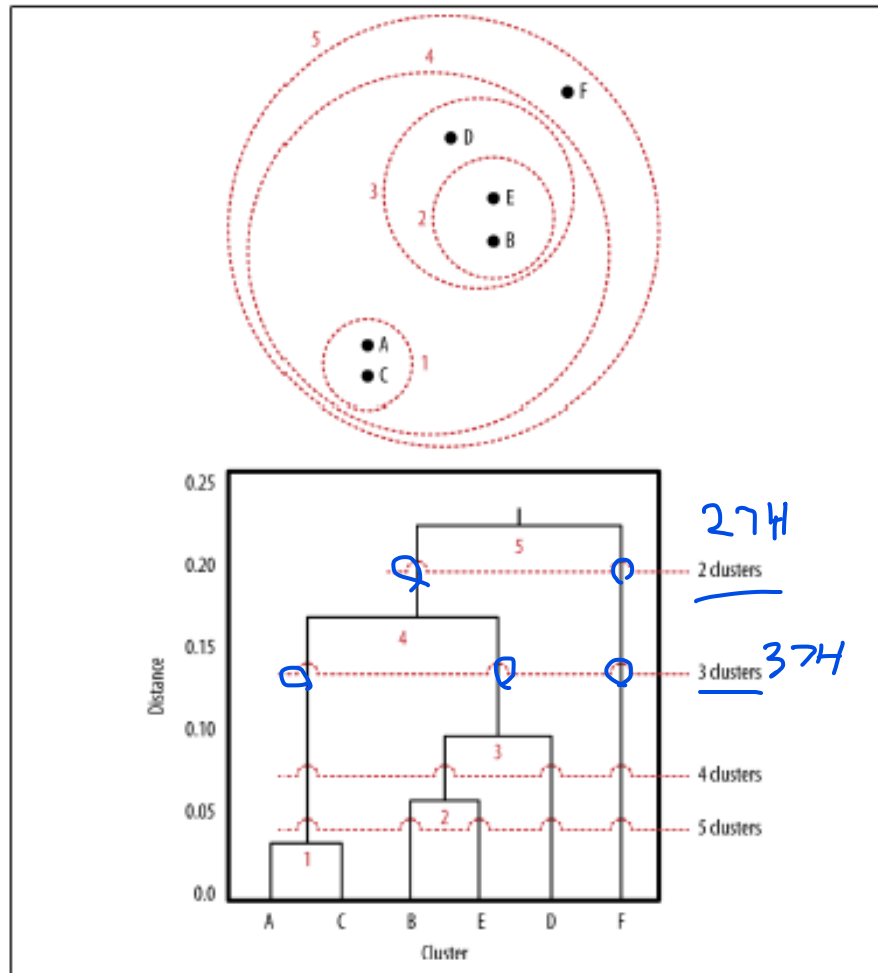
# 계층적 군집분석의 결과



Dendrogram



# 계층적 군집분석의 활용

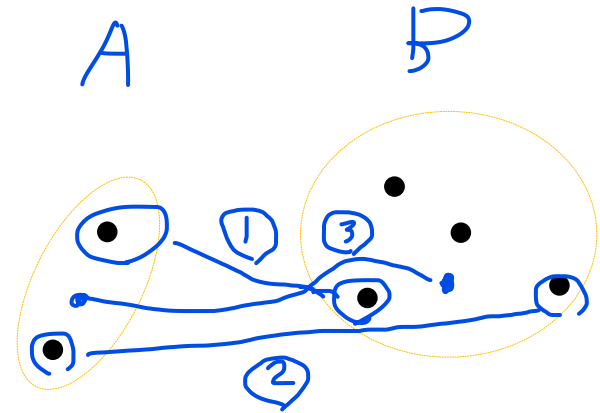
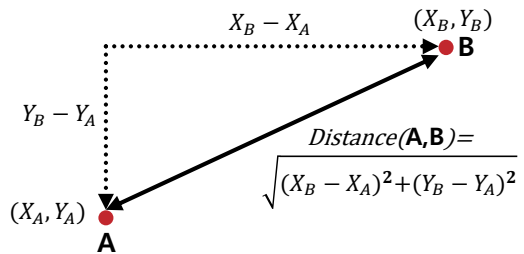


Shmueli et al., "Data Mining for Business Analytics, R Edition"

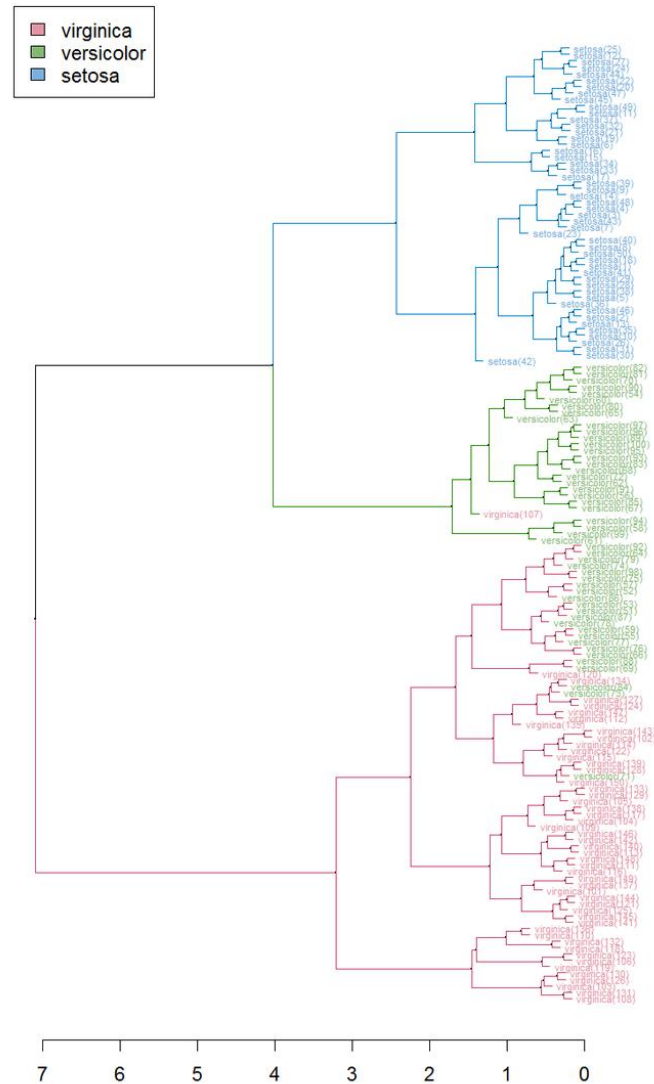


# 군집 거리 계산

- ❖ 단일연결법 (single linkage) ① 가가분
- ❖ 완전연결법 (complete linkage) ② 먼
- ❖ 평균연결법 (average linkage) 나 계산 ] 안정성 ↓
- ❖ 중심연결법 (centroid linkage) ③



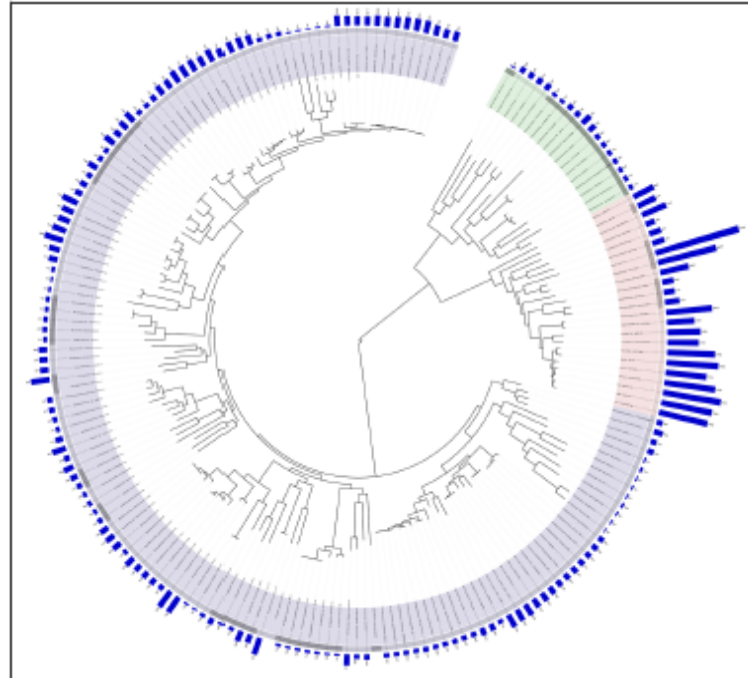
# 실제 덴드로그램



[https://en.wikipedia.org/wiki/Hierarchical\\_clustering#/media/File:Iris\\_dendrogram.png](https://en.wikipedia.org/wiki/Hierarchical_clustering#/media/File:Iris_dendrogram.png)



# 실제 덴드로그램



데이터 ↑  
ex) 품목종

Shmueli et al., "Data Mining for Business Analytics, R Edition"



# 계층적 군집분석의 장단점

## ❖ 장단점

### ● 장점

- 사전에 군집의 개수를 정할 필요 없이, 분석 결과를 바탕으로 적당한 군집 수를 결정할 수 있음
- 비교적 분석과정의 이해가 용이함

### ● 단점

- 매우 큰 데이터 셋의 경우 계산속도가 느림
- 초기에 레코드가 잘못된 군집에 배정되면 다시 수정할 기회가 없음
- 안정성이 낮음
- 평균연결법의 경우, 다른 방법보다 안정성이 낮음
- 이상치에 민감함

+ 완전

outlier

