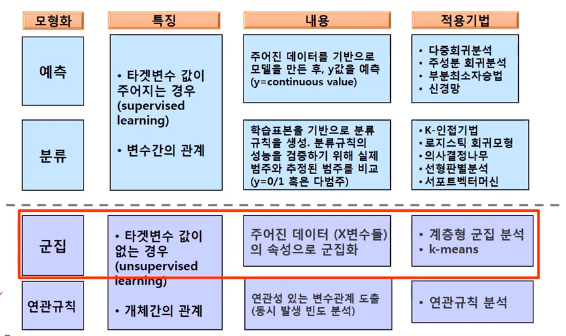
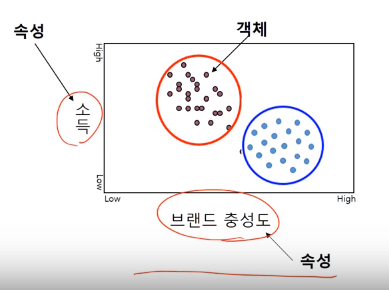
군집분석은 비지도학습(Unsupervised Learning) : 속성 변수들의 특징으로 그룹화



1. 군집분석(cluster analysis)이란, 유사한 속성을 가진 객체들을 군집으로 나누는 데이터마이닝 기법



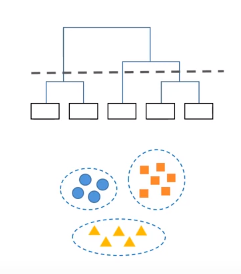
예제: 고객들의 구매패턴을 반영하는 속성들에 관한 데이터가 수집된다고 할 때

* 군집분석을 통해 유사한 구매패턴을 보이고 고객들을 군집화하고 판매전략을 도출

2. 군집분석의 종류

군집분석의 방법은 (1) 계층적 방법과 (2) 비계측정 방법으로 구분

1. 계층적 군집 : 사전에 군집 수 k를 정하지 않고 단계적으로 군집 트리 제공
2. 비계층적 군집 : 사전에 군집 수 k를 정한 후 각 객체를 k개 중 군집에 배정



3. 유사성 척도

객체 간의 유사성 정도를 정량적으로 나타내기 위해서 척도가 필요

1. 거리 척도 : 거리가 가까울수록 유사성이 크다. 거리가 멀수록 유사성이 적어짐
2. 상관관계척도 : 객체간 상관관계가 클수록 두 객체의 유사성이 커짐

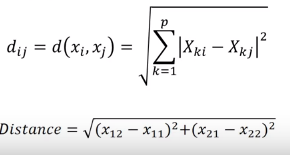
3-1 거리척도

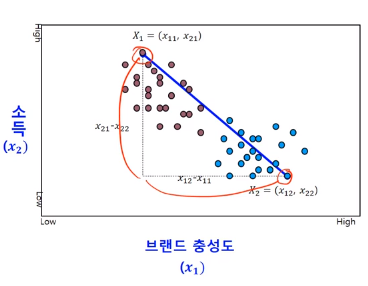
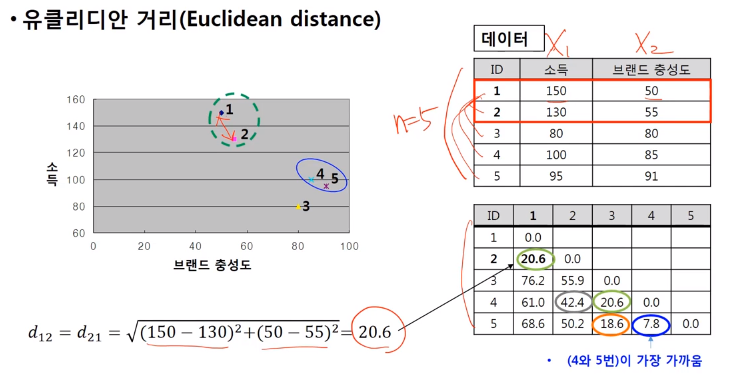
(1) 객체 i의 p차원 공간에서의 좌표는 다음과 같은 열벡터로 표현

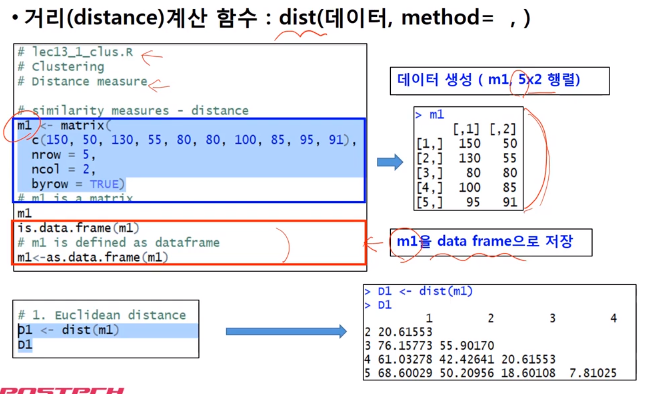
-p개의 속성을 가진 객체 i에 대하여, j번째 속성은 으로 표현

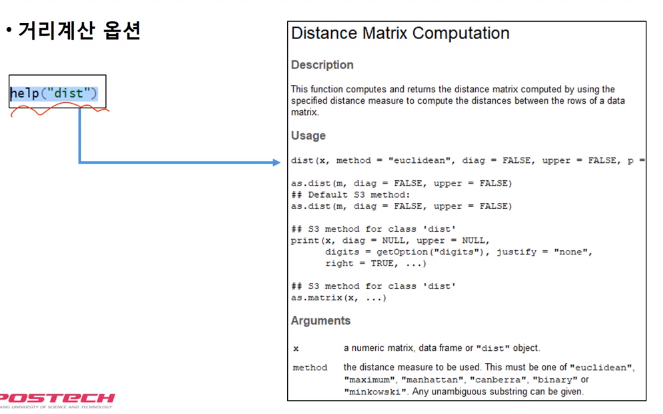


(2) 유클리디안 거리



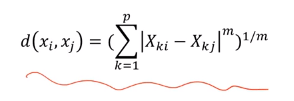
 

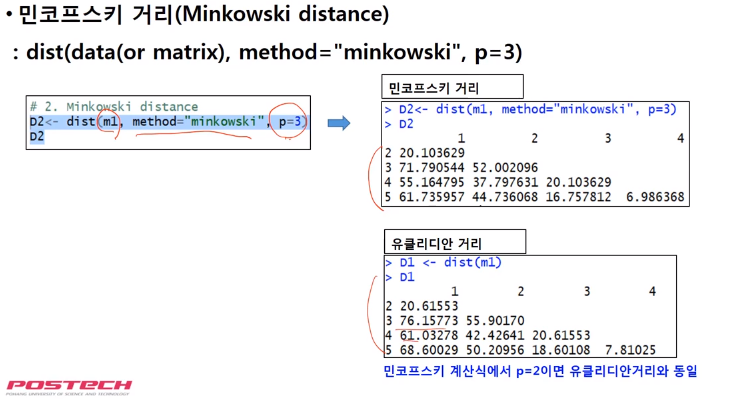




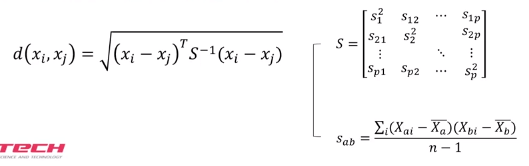
4. 그 외 거리 척도

(1) 민코프디스키 거리 : 유클리디안 거리의 일반화된 방법(m=2 일 때는 유클리디안 거리와 동일)





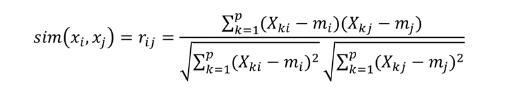
(2) 마할라노비스 거리 : 변수 간의 상관 관계가 존재할 때 사용

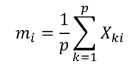


5. 상관관계를 척도로 사용

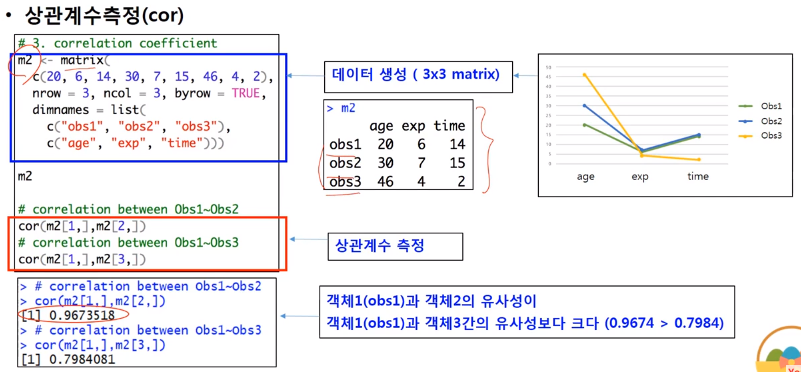
또 다른 유사성 척도로 객체 간의 상관계수를 사용

1. 상관관계가 클수록 두 객체릐 유사성이 크다고 추정

* 객체 i와 객체 j간의 표본상관계수는 다음과 같이 정의
* 
* 이때 는 객체 I의 평균값으로 다음과 같음



6. 상관계수



소감 및 각오

팀원들과 함께 찾아보았던 여러 정보를 제 스스로 정리해보면서 비모수통계의 전반적인 내용에 대해서 이해해볼 수 있었습니다. 팀원들과 공부할 때 보다 혼자 실습해보는 과정은 쉽지만은 않았지만 스스로 공부한 만큼 실습한 내용에 대해서는 잘 잊히지 않을 것 같다는 생각이 들었습니다. 이를 계기로 R언어에 대해 자신감이 더 생긴 것 같습니다. 저는 이런 자신감을 통해 R언어를 더욱더 열심히 공부할 것입니다. 이런 기회를 주신 심재창 교수님 감사합니다.