



2016년도 제3회 통계적 방법론 공개강좌

과목 소개

잠재구조의 탐색 : 변수 군집화와 PLS 회귀

이 강의는 다변량 자료에 숨어 있는 잠재구조의 탐색(exploration of latent structures)을 주제로 다음의 2개 부분으로 구성된다.

제1부는 p 개 X 변수들의 군집화를 통해 m 개의 잠재변인(latent factors)을 찾아내는 잠재적 군집화를 다룬다. 이때, 관측변수 수 p 에 비해 잠재변인 수 m 은 훨씬 작다. 이러한 변수 군집화는 고차원 데이터를 이해하는 실마리가 된다.

제2부는 설명변수 수 p 가 관측개체의 수 n 보다 큰 다변량 데이터에 적용 가능한 선형회귀 방법인 PLS 회귀를 다룬다. PLS 회귀(partial least squares regression)는 p 개 X 변수들을 적은 수의 잠재변인으로 축약하여 단변량 및 다변량 Y 변수들에 대한 예측을 해낸다.

본 강의에서는 방법론 설명과 함께 R의 “ClustOfVar” 패키지와 “pls” 패키지가 활용된 다수의 수치적 사례들을 제시할 것이다.

고려대학교 통계연구소