

12주차. 주성분분석

- * 주성분 분석(principal component analysis) : 가장 널리 사용되는 차원 축소 기법 중 하나로, 원 데이터의 분포를 최대한 보존하면서 고차원 공간의 데이터들을 저차원 공간으로 변환한다. 주성분 분석의 목표는 관련 '공분산 행렬로부터 얻은 정보'를 최대한 보존하는 '훨씬 더 적은 개수의 새로운 변수들'을 찾으려는 것이다.
- * 센터링(centering) : 데이터 행렬에서 각 확률변수의 평균을 0으로 조정하는 것이다.
- * 주축(principal axis) : 주성분을 구하기 위하여 찾는 축으로, 원 데이터를 이 축 상에 정사영할 때, 얻어진 (새로운) 데이터들의 분포가 가장 최대가 되도록 하는 축을 말한다.
- * 주성분 점수(PC score) : 원 데이터를 주축에 정사영 한 것이다.