

1. "PCAex.csv" 데이터 셋은 4개의 연속형 특성변수로 구성된 자료이다. 이 자료에 주성분분석(PCA)를 적용하고자 한다. 다음 물음에 답하여라. 단, PCA 분석을 위한 연산은 sklearn에서 제공하는 PCA 클래스를 이용하지 말고, numpy 등에서 제공하는 행렬 연산함수를 이용하여 산출할 것.
 - (1) 주어진 자료에 대한 표본상관행렬을 구하여라.
 - (2) (1)의 표본상관행렬에 기초하여 주어진 자료에 대한 주성분을 구하고, 각 주성분 별 변동성을 구하여라.
 - (3) 첫 두개의 주성분은 특성변수의 전체 변동 중 어느정도를 설명하는가?
 - (4) 첫 두 개의 주성분 점수를 산점도를 이용하여 시각화 하여라.
 - (5) 두번째 주성분 점수가 높은 관찰치는 어떤 특징이 있다고 볼 수 있는가? 각 특성변수와 해당 주성분의 연관성을 이용하여 설명하여라.
2. "mutfunds.csv" 데이터 셋은 25개월 간 연속적으로 측정한 15개 뮤추얼 펀드의 가격(M1,...,M15)과 해당 월말의 이자율(IRATE) 및 주식시장지수(MINDEX)가 포함되어 있다. 유사한 자산 포트폴리오를 보유한 펀드들의 가격은 비슷한 방식으로 달라질 것이다.
 - (1) 주어진 자료에서 단 하나의 잠재요인으로 관찰 된 뮤추얼 펀드 가격의 변동을 설명하고자 한다면, 최대한 어느 정도가 설명되는가?
 - (2) (1)에서의 잠재요인은 무엇을 반영하는 변수로 볼 수 있겠는가?
 - (3) (1)에서의 잠재요인 이외에 15개의 뮤추얼 펀드 가격 변동을 설명하는 데 도움이 되는 추가적인 잠재요인이 있는지 파악하여라. 추가적인 요인은 무엇을 반영하는 변수로 볼 수 있겠는가?
3. "ldadata.csv" 자료는 어느 시점을 기준으로 도산한 기업과 정상운영한 기업을 대상으로 지난 2년 동안 발생된 재무상황의 변동을 조사한 결과이다. 변수 Y는 도산여부(1:도산, 2:정상운영), 나머지 X1,...,X5는 재무상황 변동에 관한 변수이다.
 - (1) Fisher의 판별함수를 구하여라.
 - (2) (1)의 결과를 이용하여 X1,...,X5 변수의 정보를 1차원인 newX 변수로 압축할 수 있다. newX의 값들의 분포를 Y의 그룹별로 시각화하여라. (예. Y가 1인 경우와 2인경우로 구분해서 newX 값들의 히스토그램, 상자그림 등을 그릴 것)