Ch6. 차원 축소

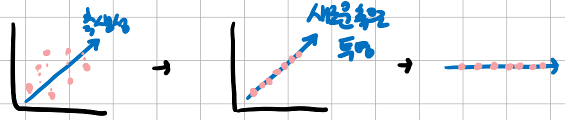
01 차원축소 개요

- 차원축소 → 직관적 해석가능, 데이터크기↓
  - 피처선택 : 불필요한 피처 아예 제거, 주요피처만 선택
  - 피처추출 : 함축적 요약 특성 추출 → 잠재적인 요소
  - 이미지 변환·압축, 텍스트문서 ~ 숨겨진 의미추출

02 PCA

- PCA

: 여러 변수간 상관관계 이용 → 이를 대표하는 주성분 추출 → 차원축소



① 입력 데이터세트의 공분산 행렬 생성

② 공분산 행렬의 고유값, 고유벡터 계산

③ 고유값 큰순으로 K개(PCA 변환 차수 만큼)의 고유벡터 추출

④ 추출된 고유벡터를 이용해 입력 데이터 변환

03 LDA

- LDA : 선형판별분석법

- PCA와 유사
  - PCA: 입력데이터의 변동성 가장 큰 축
  - LDA: 입력데이터의 결정값 클래스 분리 최대화하는 축

① 클래스내부분산 행렬 / 클래스간 분산 행렬 구하기
   S_W (within) / S_B (between)

(입력 데이터의 결정 값 클래스별 개별피처의 평균벡터 기반)

② $$S_W^T S_B = \begin{bmatrix} e_1 & \cdots & e_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} \begin{bmatrix} e_1^T \\ \vdots \\ e_n^T \end{bmatrix}$$

③ 고유값이 가장 큰순으로 K개(LDA 변환차수만큼) 추출

④ 추출된 고유벡터를 이용해 새롭게 입력 데이터 변환

04 SVD

- SVD : 특이값 분해 (정방행렬 아니어도 적용가능)

$$\underset{m\times n}{A} = \underset{m\times m}{U}\ \underset{m\times n}{\Sigma}\ \underset{n\times n}{V^T}$$

(Σ: 대각행렬)

- Σ의 대각원소 中 0인 부분 (+ 그에 대응하는 U, V의 원소) 제거

$$\underset{m\times n}{A} = \underset{m\times p}{U}\ \underset{p\times p}{\Sigma}\ \underset{p\times n}{V^T}$$

- Truncated SVD :
  Σ의 대각원소 中 상위 몇개만 추출 → 차원을 더 줄인형태

05 NMF

- NMF

: Truncated SVD 같은 낮은 행렬 근사방식의 변형

$$\underset{4\times 6}{V} \approx \underset{4\times 2}{W} \times \underset{2\times 6}{H}$$

- W : 길고 가는 행렬
  원본 행에 대해 잠재요소의 값이 얼마나 되는지

- H : 작고 넓은 행렬
  잠재 요소가 원본 속성으로 어떻게 구성됐는지