

머신러닝(Machine Learning)

비지도학습

목 차

01 차원 축소?

02 차원 축소의 종류

03 차원 축소의 유의한점

04 PCA(주성분 분석)

01 차원 축소(Dimension Reduction)

▶ 차원 축소(Dimension Reduction)

고차원 데이터로부터, 저차원의 데이터로 변환하는 방법

▶ WHY?

- A. 피처가 많을 경우, 개별 피처 간에 상관관계가 높을 가능성이 크다.
- B. 다중 공선성 문제로 인해 모델의 예측 성능이 저하될 가능성이 크다.

02 차원 축소(Dimension Reduction) 종류

▶ 대표적인 차원 축소 알고리즘

PCA(Principal Component Analysis) - 주성분 분석

LDA(Latent Dirichlet allocation) - 잠재 디리클레 할당

SVD(Singular Value Decomposition) - 특이값 분해

NMF(Non-negative matrix factorization) - 음수 미포함 행렬 분해

03 차원 축소(Dimension Reduction) 유의한점

▶ 일반화 성능 향상 (이미지의 경우)

예를 들면 차원이 축소된 이미지는 원본 이미지보다 훨씬 적은 차원이기 때문에 이미지 분류 등의 분류 작업 수행시에 과적합(overfitting)영향력이 작아지게 된다.

▶ 텍스트 문서의 숨겨진 의미를 추출

차원 축소 알고리즘을 통해 문서 내 단어들의 구성에서 숨겨진 시맨틱(Semantic)의미나 토픽(Topic) 잠재 요소를 간주하고 이를 찾아낼 수 있음.

04 PCA(주성분 분석)

▶ PCA(Principal Component Analysis)

- (A) 고차원의 데이터를 저차원으로 변환할 수 있는 알고리즘
- (B) 서로 연관성이 있는 고차원 변환을 직교 변환을 사용하여 저 차원 공간(주성분)의 표본으로 변환
- (C) 데이터를 한개의 축으로 투영했을 때, 분산이 가장 커지는 축을 첫번째 주성분
- (D) 두번째로 커지는 축을 두번째 주성분에 놓이도록 새로운 좌표계로 데이터를 선형 변환
- (E) 주성분들은 직교한다는 제약아래에 가장 큰 분산을 갖고 있다.