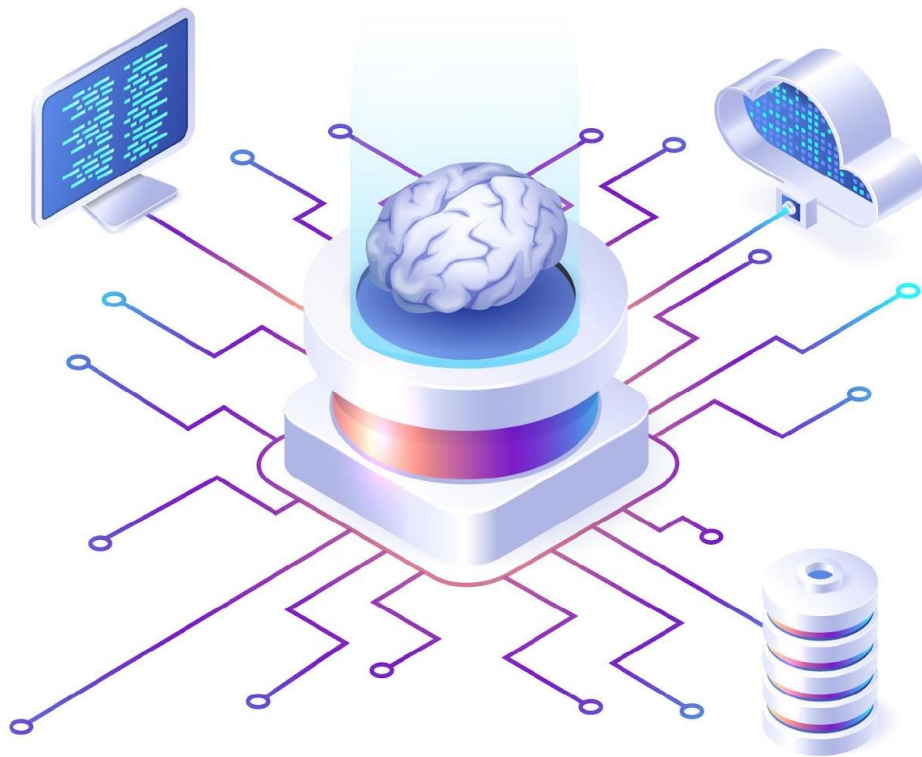


차원 축소

실무형 인공지능 자연어 처리



차원 축소 (Dimension Reduction)

선형대수 기초

1

차원 축소 (Dimension Reduction)

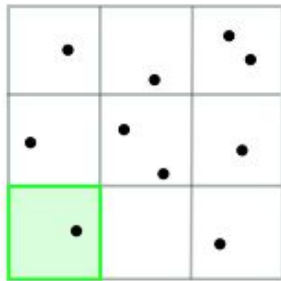


차원의 저주

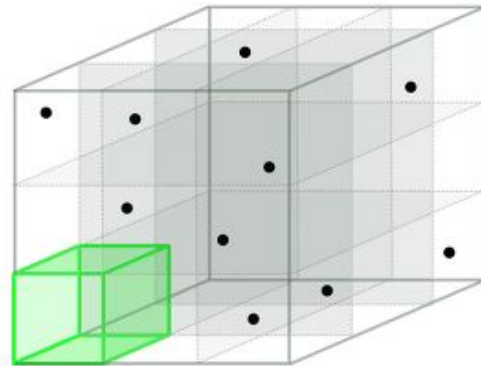
- 차원 증가할 수록 데이터의 표현이 어려움
 - 차원 증가할 수록 연산(분석) 어려움
- => 차원 축소가 필요함



3/10



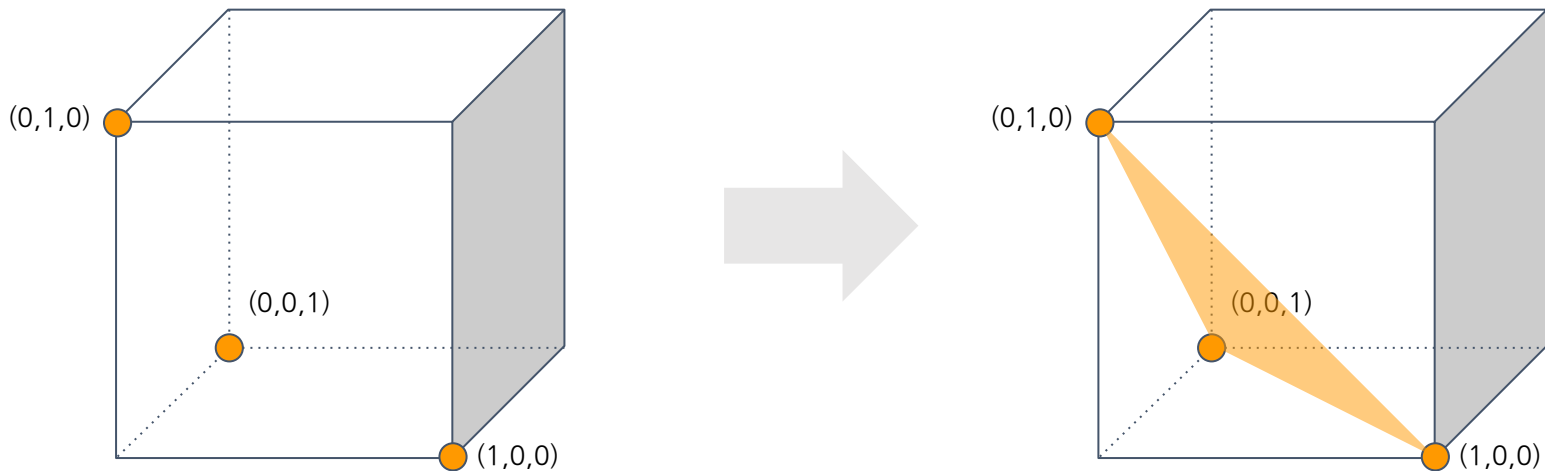
1/10



0/10

차원 축소 개요

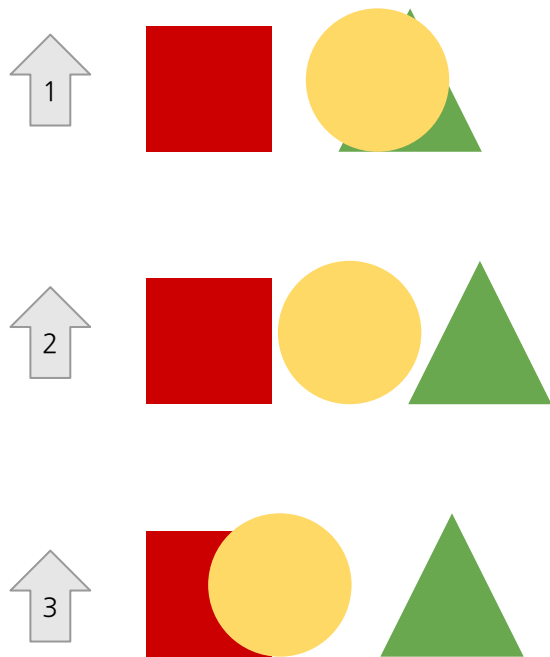
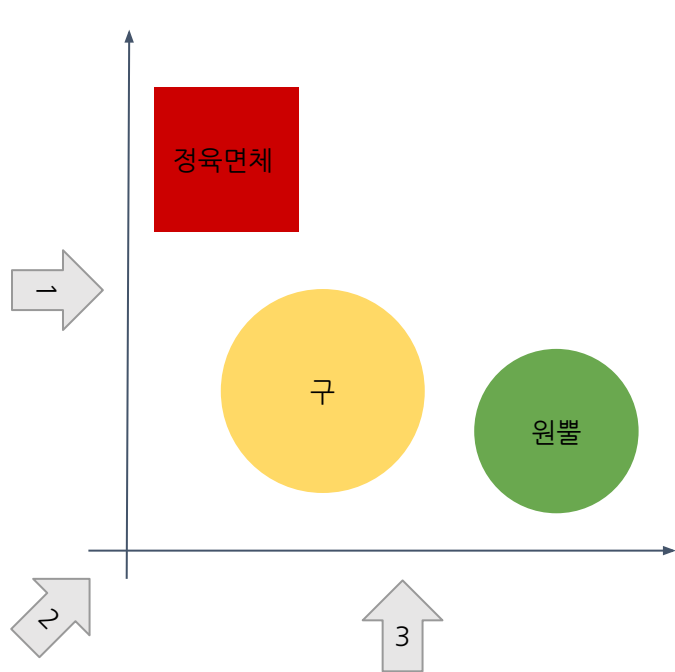
- 입력 데이터의 차원이 클 경우, 차원의 저주와 학습 속도가 저하됨
- 같은 정보를 표현하는데 낮은 차원을 사용하여 정보를 표현하는 것을 차원 축소라고 함



차원 축소의 예

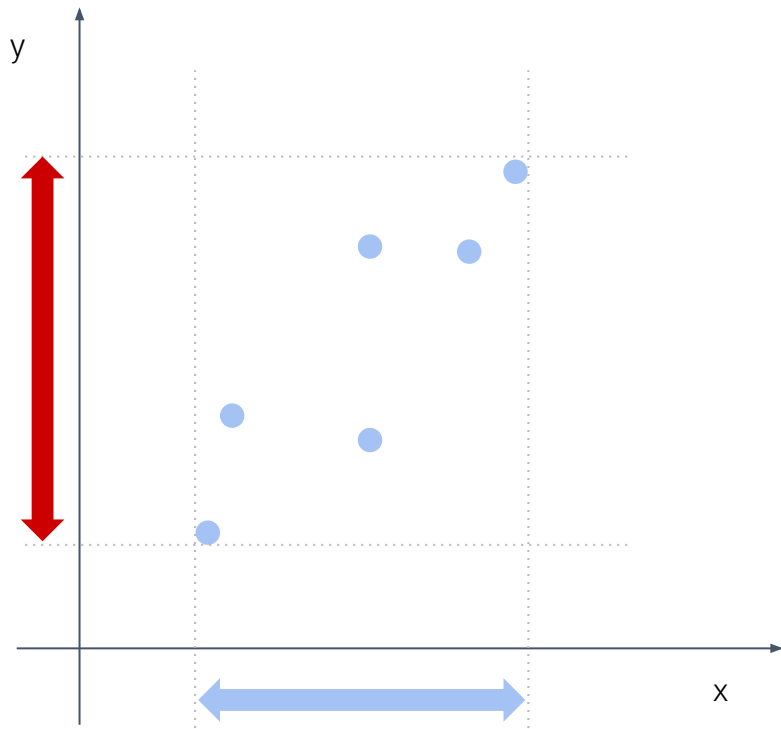
- 예시 1)
 - 국어 : 80점, 영어 : 60점, 수학 : 90점, 과학 : 80점
 - 평균 = $(80 + 60 + 90 + 80)/4 = 77.5$ 점
 - 4개의 차원을 "평균"이라는 1개의 차원으로 축소
- 예시 2)
 - 체질량지수(BMI) = $\text{몸무게} / \text{키}^2$
 - 몸무게, 키를 1개 차원으로 축소
- 차원축소시 정보의 손실이 발생한다. 차원 축소에서 가장 중요한 것은 정보 손실을 최소화하는 것

차원 축소의 기하학적 의미 (1)

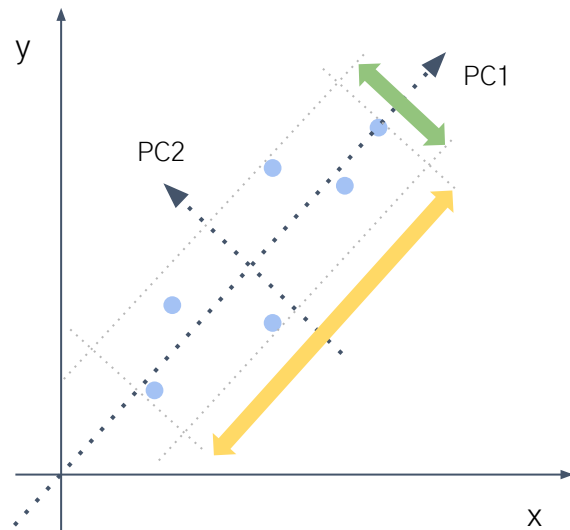
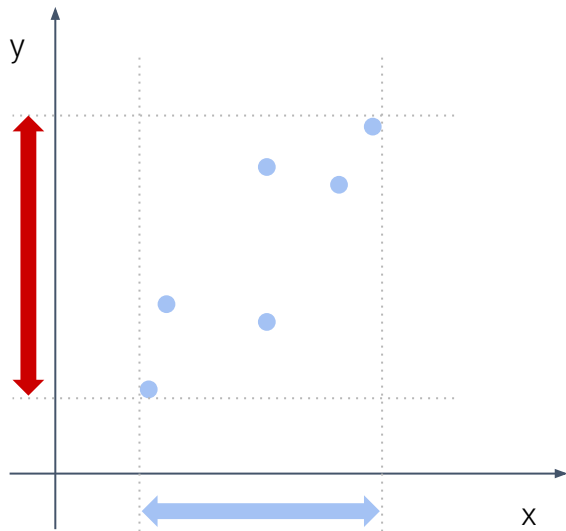


바라보는 시선에 따라 분류의 정도가 달라질 수 있다.

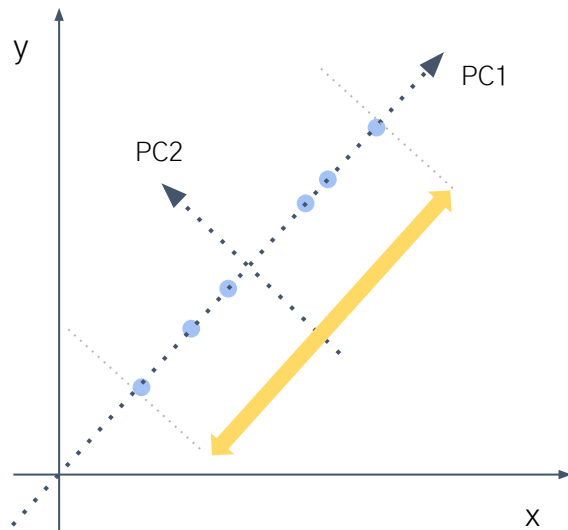
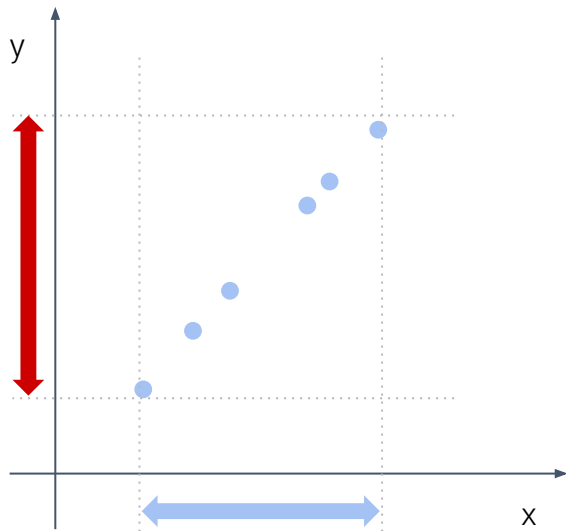
차원 축소의 기하학적 의미 (2)



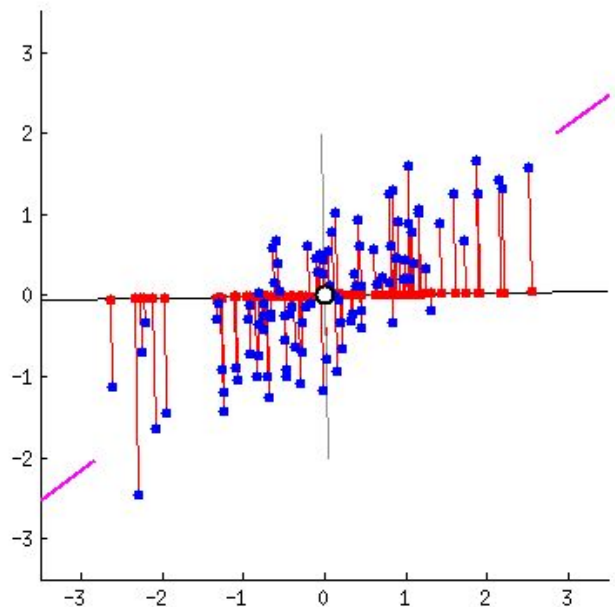
차원 축소의 기하학적 의미 (3)



차원 축소의 기하학적 의미 (4)



차원 축소의 기하학적 의미 (5)



<https://stats.stackexchange.com/questions/2691/making-sense-of-principal-component-analysis-eigenvectors-eigenvalues>

차원 축소 (Dimension Reduction)

선형대수 기초

2

PCA (주성분 분석)

Principal Component Analysis



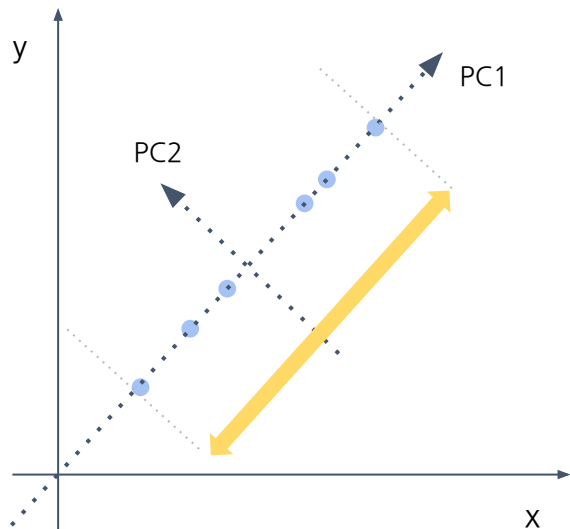
주성분 분석(PCA)

주성분 분석(主成分分析, Principal component analysis; PCA)은 고차원의 데이터를 저차원의 데이터로 환원시키는 기법이다. 서로 연관 가능성이 있는 고차원 공간의 표본들을 선형 연관성이 없는 저차원 공간(주성분)의 표본으로 변환하기 위해 직교 변환을 사용한다. 주성분의 차원수는 원래 표본의 차원수보다 작거나 같다. 주성분 분석은 데이터를 한개의 축으로 사상시켰을 때 그 분산이 가장 커지는 축을 첫 번째 주성분, 두 번째로 커지는 축을 두 번째 주성분으로 놓이도록 새로운 좌표계로 데이터를 선형 변환한다. 이와 같이 표본의 차이를 가장 잘 나타내는 성분들로 분해함으로써 여러가지 응용이 가능하다. 이 변환은 첫째 주성분이 가장 큰 분산을 가지고, 이후의 주성분들은 이전의 주성분들과 직교한다는 제약 아래에 가장 큰 분산을 갖고 있다는 식으로 정의되어있다. 중요한 성분들은 공분산 행렬의 고유 벡터이기 때문에 직교하게 된다.

https://ko.wikipedia.org/wiki/%EC%A3%BC%EC%84%B1%EB%B6%84_%EB%B6%84%EC%84%9D

다차원 데이터의 정보를 가능한 한 손실 없이 저차원 공간에 압축하는 것

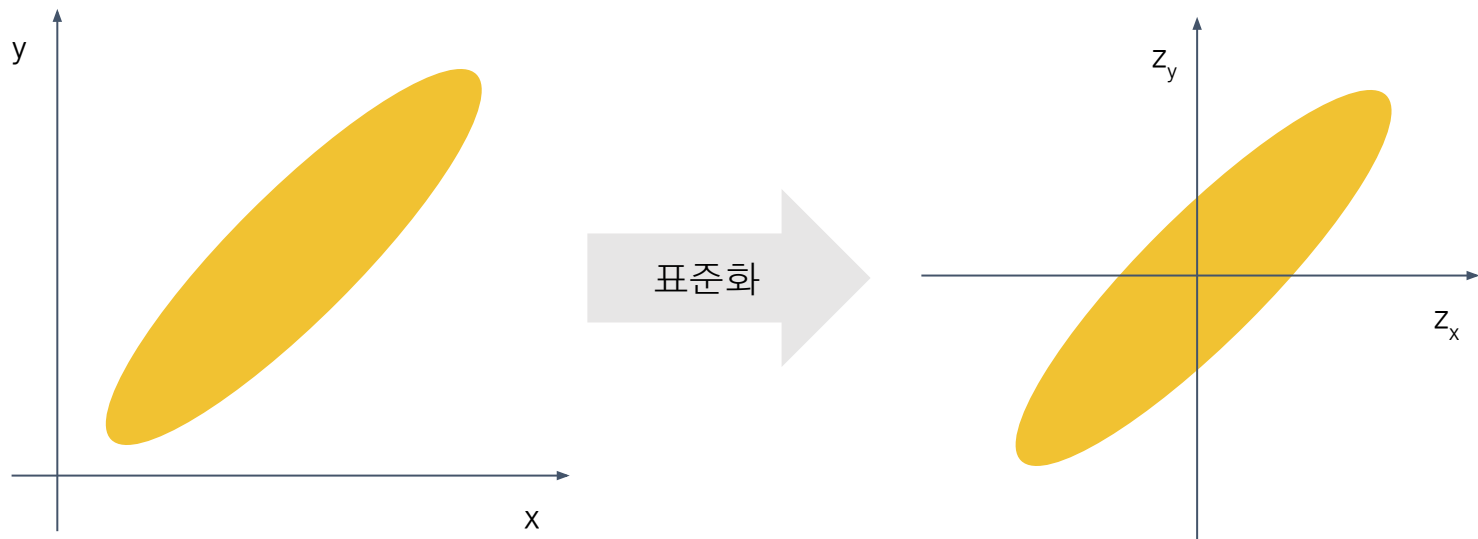
주성분 분석 개요



- 첫번째 주성분 축생성
 - 분산이 가장 큰(데이터가 가장 넓게 퍼져있는) 방향을 구한다
 - 그 방향으로 첫번째 축을 만든다
- 두번째 주성분 축생성
 - 첫번째 축과 90도 직교하며, 분산이 두번째로 큰 방향을 구한다.
 - 그 방향으로 두번째 축을 만든다
- 세번째 주성분 축생성
 - ...

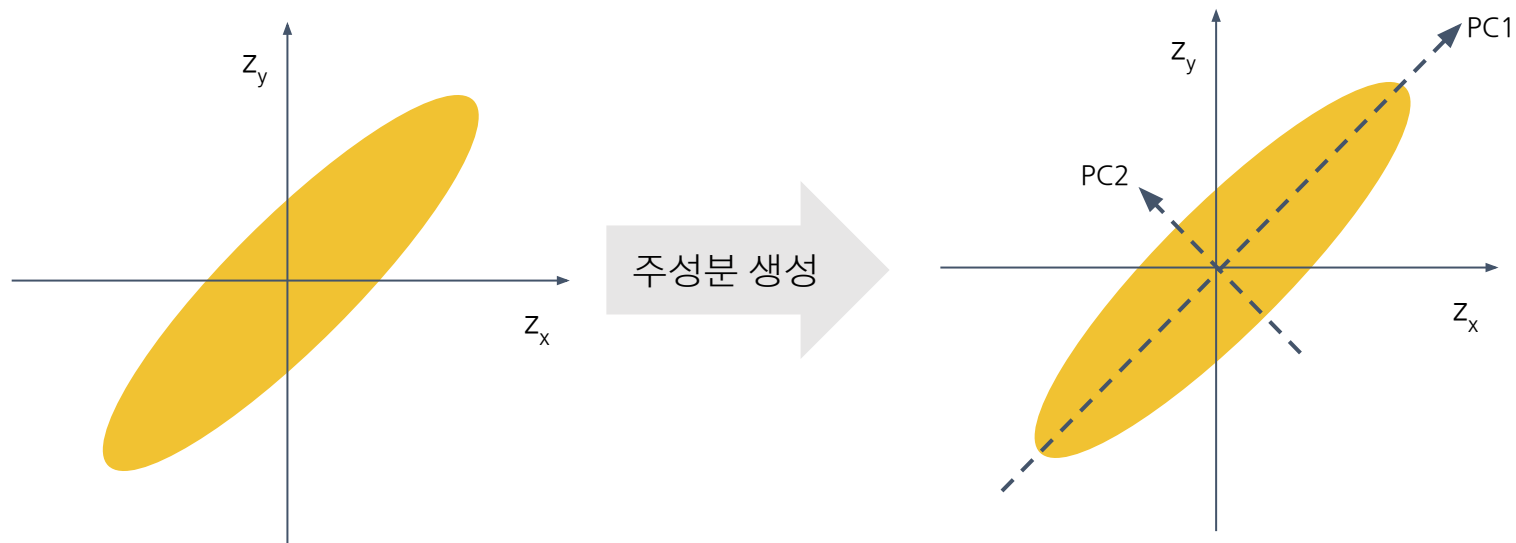
다차원 데이터의 정보를 가능한 한 손실 없이 저차원 공간에 압축하는 것

주성분 분석 절차(1)



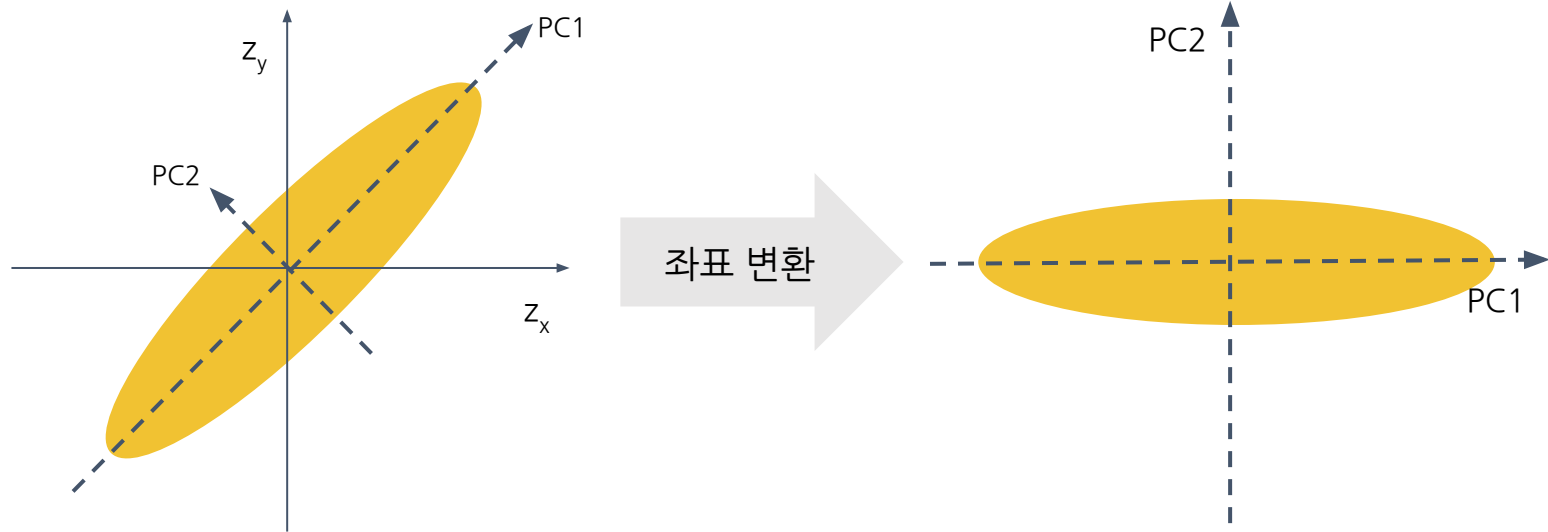
데이터를 표준화(평균 0, 표준편차 1)

주성분 분석 절차(2)



분산이 가장 큰(데이터가 가장 많이 퍼져있는) 방향을 찾아서 PC1을 생성 (반복)

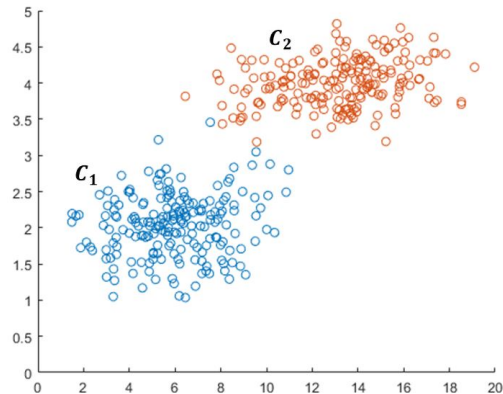
주성분 분석 절차(3)



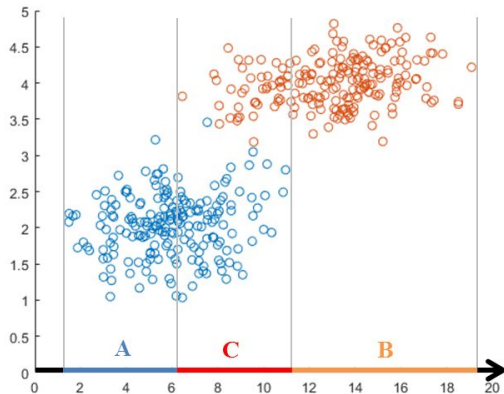
구해진 주성분 축으로 좌표를 변환

주성분 분석 예시 (1)

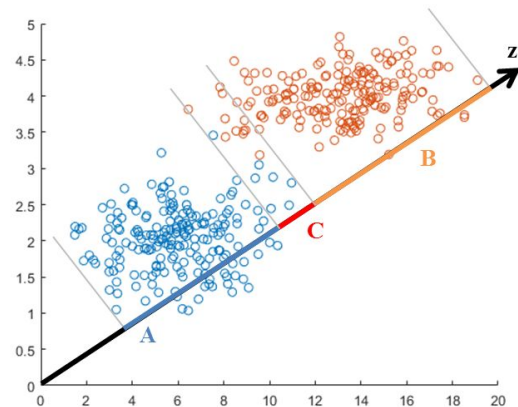
- 2차원의 분류(Classification) 결과를 1차원으로 축소



분류(Classification) 결과



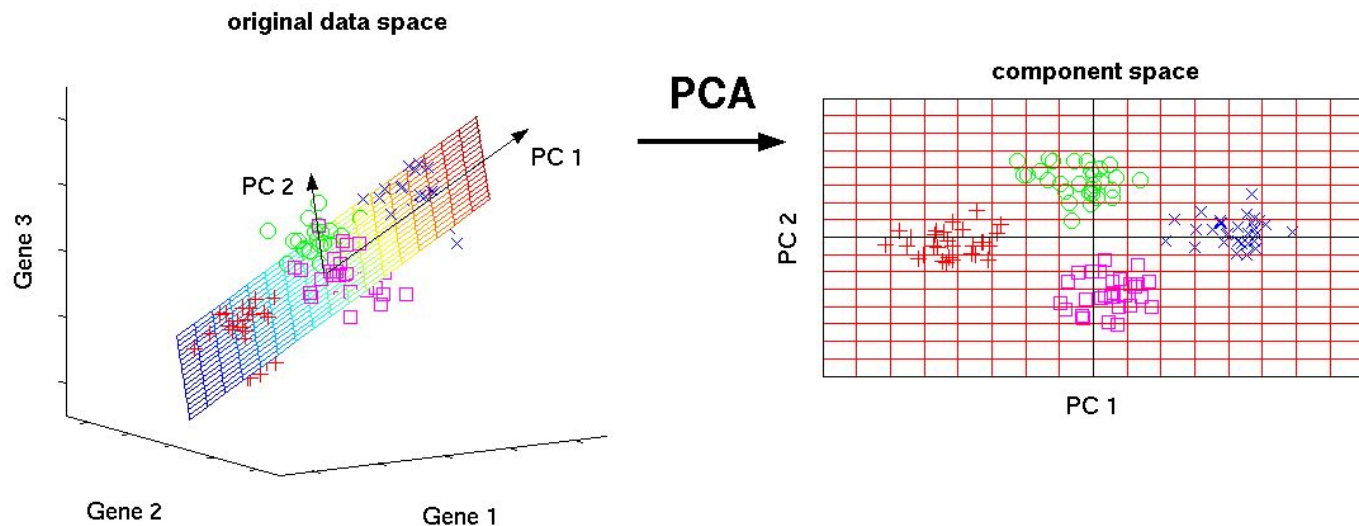
x축에 사상



z축에 사상

주성분 분석 예시 (2)

- 3차원의 분류(Classification) 결과를 2차원으로 축소



감사합니다.

Insight⁺campus

