

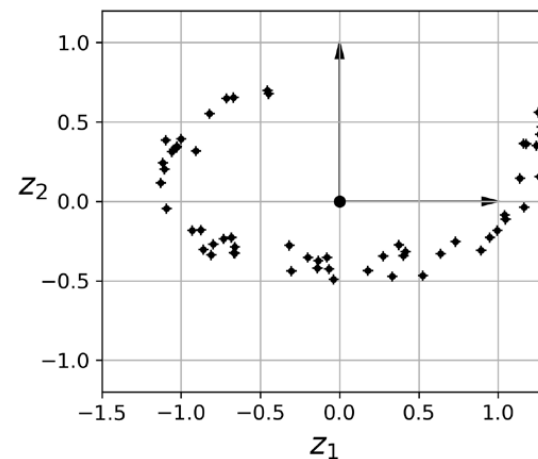
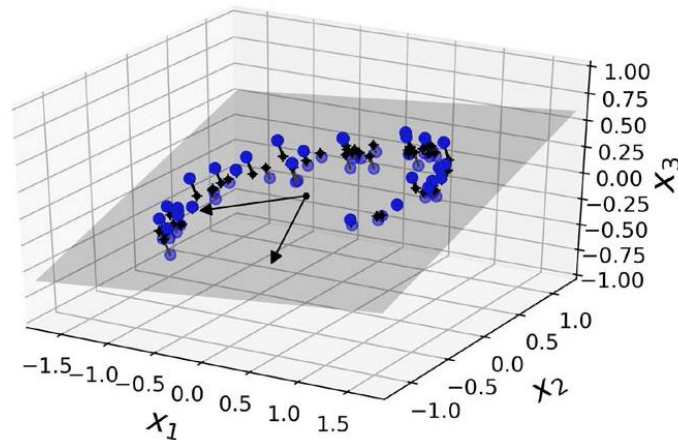


의료인공지능 머신러닝 - PCA

고려대학교 의료빅데이터연구소
채민수(minsuchae@korea.ac.kr)

1. 주성분 분석이란

- 주성분 분석이란(Principal component analysis)
 - 다양한 특성을 가진 항목을 정해진 특성 수 만큼으로 표현하는 방안
 - 사용 목적
 - 특성 수 감소
 - 연산 속도 감소
 - 메모리 절약
 - 시각화



2. 주성분 분석 실습

- PCA 실습
 - 물고기 분류를 이용한 PCA 실습

2. 주성분 분석 실습

- PCA를 통한 압축 실습
 - MNIST를 활용한 압축 실습

3. 다양한 PCA

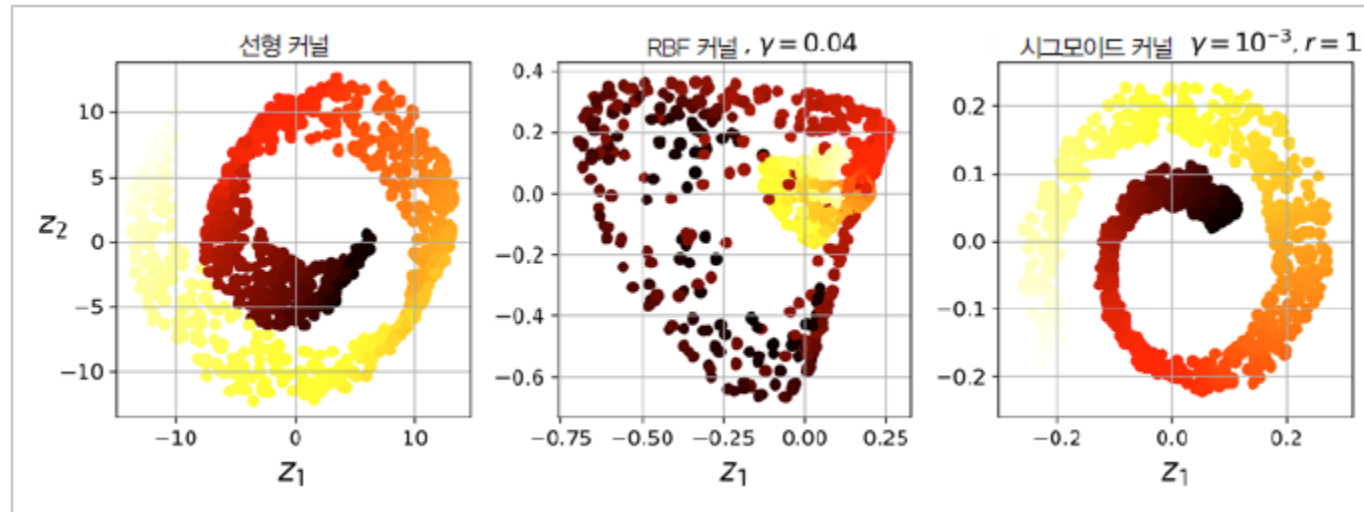
- 랜덤 PCA 실습
 - $O(m \times n^2) + O(n^3)$ vs $O(m \times d^2) + (d^3)$

3. 다양한 PCA

- 점진적 PCA 실습
 - 메모리를 절약하기 위해 부분적으로 학습

3. 다양한 PCA

- 커널 PCA 실습
 - PCA 결과를 커널을 이용하여 적용

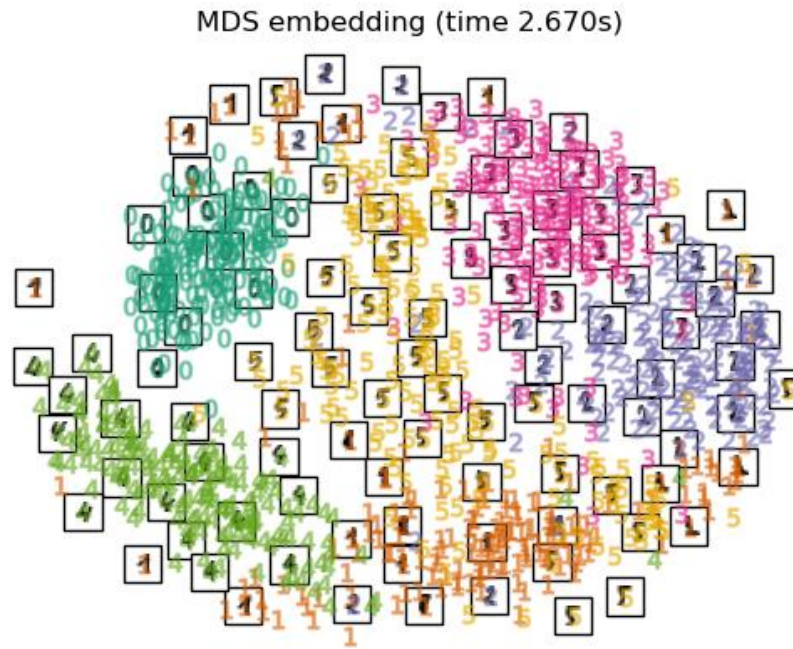


4. 비선형 차원 축소

- 지역 선형 임베딩(locally linear embedding) 실습
 - 이웃한 것에 선형적으로 연관되어 있는지 측정하여 표현

5. 다양한 차원 축소

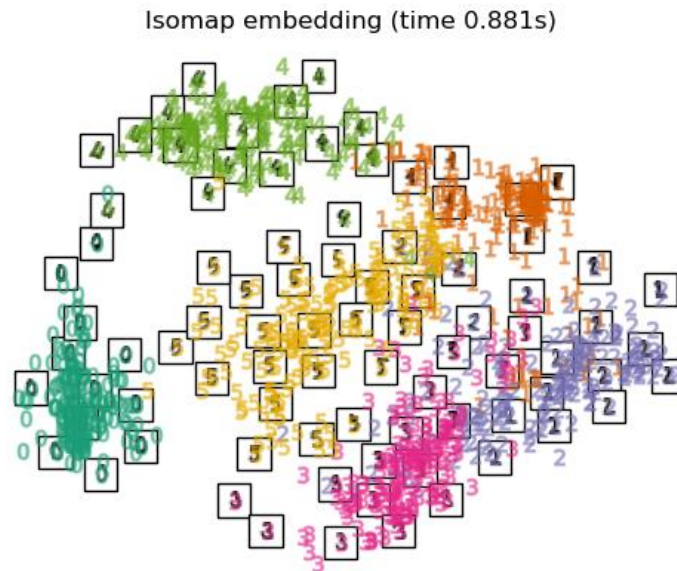
- 다차원 스케일링(Multidimensional scaling) 실습
 - 샘플 간의 거리를 보존하면서 차원 축소



5. 다양한 차원 축소

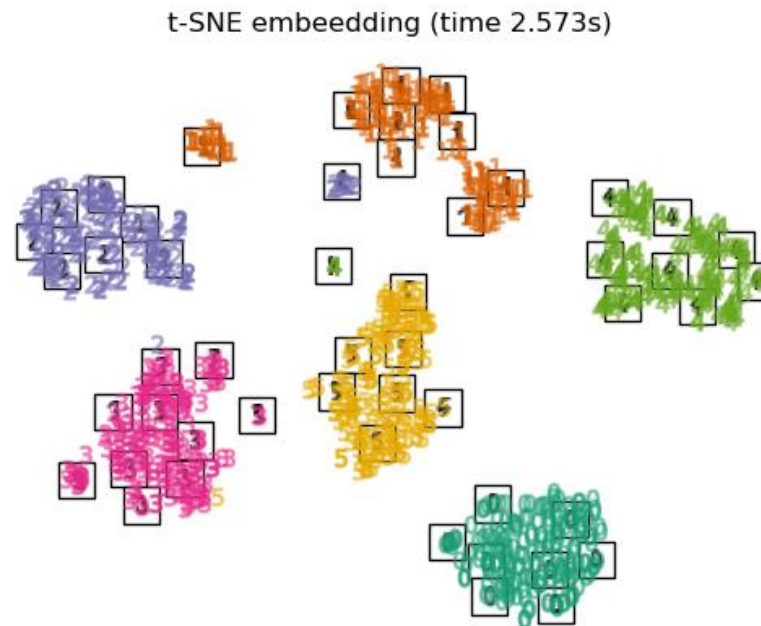
- Isomap 실습

- 샘플을 가까운 이웃과 연결하는 식으로 그래프를 생성



5. 다양한 차원 축소

- t-distributed stochastic neighbor embedding(t-SNE)실습
 - 비슷한 샘플을 가까이, 비슷하지 않은 샘플은 멀리 하도록 축소
 - 시각화를 위해 사용함(transform 미지원)



5. 다양한 차원 축소

- linear discriminant analysis(LDA) 실습
 - PCA와 다르게 얼마나 클래스 간 구분하는 축으로 학습하여 축소
 - 학습 시 target 데이터 필요
 - 이진 분류보단 다중 분류에 용이(분류 항목 -1 혹은 피쳐 항목 수 이하로만 축소됨)

