8주차 숙제

차원 축소:

차원의 저주: 특성이 많아질수록 샘플 간 거리가 커져 학습이 어려움

차원 축소: 정보 손실을 최소화하며 특성 수를 줄여 학습 성능 및 속도 개선

MNIST 데이터 784차원 → 154차원으로 줄여도 분류 가능

데이터 시각화 (2D/3D): 군집 구조 등 통찰 획득

차원축소 기법:

사영: 고차원 데이터를 선형적으로 저차원 평면에 투영

다양체 학습: 비선형 구조를 보존하며 펼치는 방식

주성분 분석:

고차원 데이터를 가장 잘 설명하는 초평면에 투영

분산을 최대한 보존하는 방향(축)을 선택 -> 주성분

SVD(특이값 분해)로 계산, 분산 누적합이 95%가 되는 시점이 적절한 차원이라고 함

랜덤 프로젝션: 고차원 데이터를 무작위로 낮은 차원에 사영해도 거리 정보가 대부분 유지됨(존슨-린덴슈트라우스 정리)

LLE(국소 선형 임베딩): 전체적으로는 비선형 구조지만, 국소적으로 선형성 가정.

각 점 주변의 선형관계를 유지하면서 저차원 표현 찾음