

## 8주차\_차원 축소

### 0. 차원 축소

차원의 저주 : 샘플 특성 많으면 학습 어려움

차원 축소 : 특성 수를 줄여 학습 가능한 문제로 만드는 기법

예) MNIST 784 픽셀 대신 154개 대상만으로 충분히 학습가능, 데이터 시각화 (2,3차원으로 줄이면 가능)

### 1. 차원의 저주

특성 수 커질수록 과대 적합 위험 커짐

해결책 : 샘플 수 늘리기 -> 사실상 불가능 / 저차원으로 낮춤

### 2. 차원 축소 기법

대표 기법 : 사영, 다양체 학습

1) 사영 기법 : 낮은 차원 공간으로 사영

2) 다양체 학습

- 쉽게 룰레이크 펼치기

- 다양체 가설 : 대부분의 고차원 데이터셋이 더 낮은 차원의 다양체에 가깝다는 가설 (일반적으로 사실 아님)

### 3. 주성분 분석 (PCA)

: 훈련 데이터에 가까운 초평면에 데이터셋을 사영하는 방법

- 초평면 설정 기준 : 분산 보존 최대, 특잇값 분해 (SVD)

- 랜덤 PCA

- 점진적 PCA

### 4. 임의 사영

존슨-린덴슈트라우스 정리 : 고차원의 데이터를 적절한 크기의 저차원으로 임의로 사영

하더라도 데이터셋의 정보를 많이 잃어버리지 않음을 보장

## **5. 국소적 선형 임베딩 (LLE)**

: 대표적인 다양체 학습 기법

## **6. 기타 차원 축소 기법**

- 다차원 스케일링(MDS)
- ISomap
- t-SNE
- 선형 판별 분석 (LDA)
- 커널 PCA