

# Week 9

## < 다양한 차원 축소 기법들 추가 >

### 1. PCA 변형

#### (1) Kernel PCA

kernel 사용  $\rightarrow$  non-linear 차원 축소

denoising, compression, structured prediction

#### (2) Incremental PCA

Input data size  $\sim$  depending.

PCA보다 더 memory efficient.

partial computation 가능

#### (3) Sparse PCA

데이터를 reconstruct 할 때 sparse components 사용.

amount of sparseness = tunable parameter

#### (4) Mini Batch Sparse PCA

(3)을 mini batch 사용 가능. faster /  $\downarrow$  정확도

### 2. ICA

(Independent Component Analysis)

- multivariate features dataset

$\rightarrow$  maximally independent 하거나

additive subcomponents로 나뉜다.

- 차원 축소가 사용X,

individual components를 separating.

### 3. Factor Analysis

simple linear generative model

w/ Gaussian latent variables

### 4. t-SNE

- t-distributed Stochastic Neighbor Embedding

- 고차원 데이터 타입을 사용하는 비선형 차원 축소 기법

- multi-dimensional data  $\rightarrow$  2 or more dimension (눈으로 보기 용이하게..)

- probability distribution

w/ random walk on neighborhood graph

$\rightarrow$  data 구조 파악

- 포인트들을 찾아 lower-dimensional space에

그 포인트를 나타내는 것 = 목표.

2-d plane 시각화 시 쓰인다!!

### 5. UMAP

- Uniform Manifold Approximation & Projection

- 비선형 차원 축소 기법

- t-SNE 보다 빠름.

$\rightarrow$  large dataset 비선형,

sparse matrix 데이터에 적합

- 새로운 데이터  $\rightarrow$  Embed Immediately 가능