Dimensionality Reduction

Introduction to Artificial Intelligence - 2023 Summer

Aug 3, 2023
Thu 4 PM

Kwangwoon University MI:RU
Artificial Intelligence Study

Agenda

In this course, you will learn

Part 1 – Quick Review: Norm and Covariance

Part 2 – What is Curse of Dimensionality?

Part 3 – Projection and Dimensionality Reduction

Quick Review: Norm in Linear Algebra

What is Norm?

$$\vec{x} = [x_1, x_2, \ldots\ldots, x_n]$$

X: a vector

Norm: a function that returns length/size of any non-zero vector

Conditions for norm function

$$\text{if } f(x) > 0 \text{ then } x \neq 0$$
$$\text{if } f(x) = 0 \text{ then } x = 0$$

$$f(Kx) = Kf(x)$$

$$f(x+y) \leq f(x) + f(y)$$

If these conditions are satisfied,
then the functions is a norm

Quick Review: Norm in Linear Algebra

Commonly used notations

$$\|u\| \qquad |u| \qquad \|x\|_p$$

The p-norm of vector x

If p has not given, then it usually means 2-norm

Quick Review: Norm in Linear Algebra

Commonly used notations

$$||x||_p = (x_1^p + x_2^p + x_3^p + \ldots.. + x_n^p)^{1/p}$$

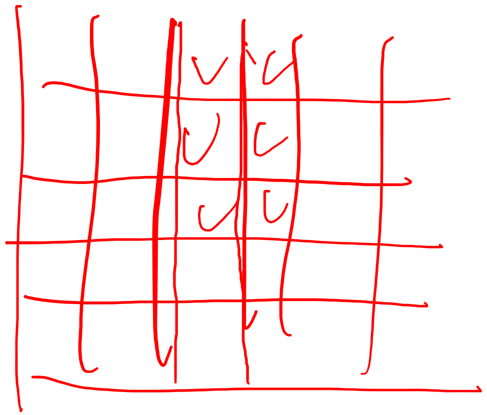
$$\boxed{||x||_p = (\sum_{i=1}^{n} x_i^p)^{1/p}}$$

Simplified form

Quick Review: Norm in Linear Algebra

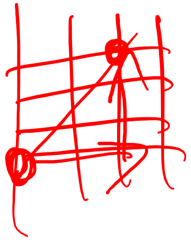
Manhattan Distance
(1-norm, L1-norm)

Grid-like path



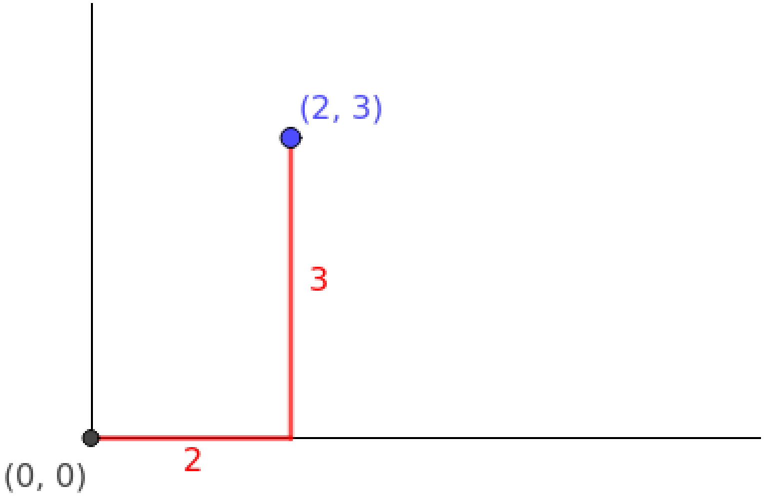
Quick Review: Norm in Linear Algebra

Manhattan Distance
(1-norm, L1-norm)



$$\vec{a} = [2, 3]$$

Let's assume we
have a vector "a"



1-norm could be represented in this form

$$||a||_1 = (2+3)^1$$

$$||a||_1 = 5$$

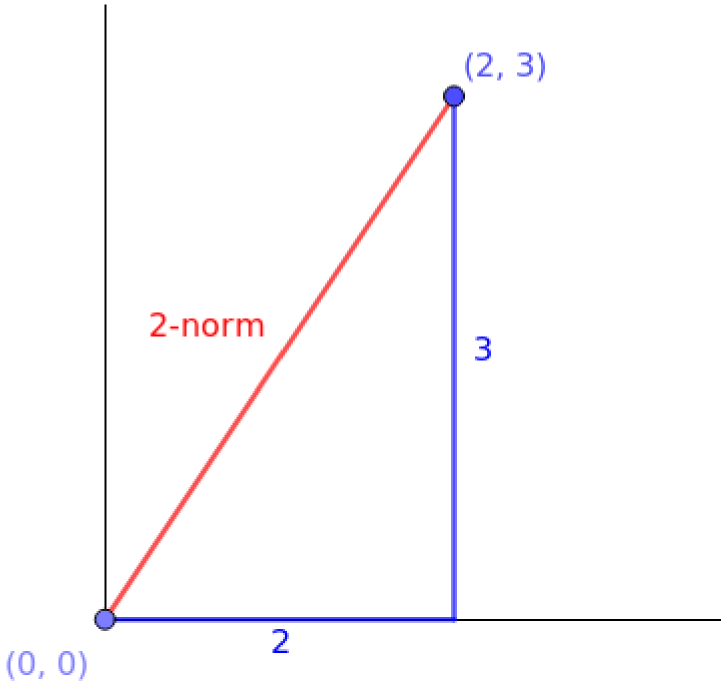
Quick Review: Norm in Linear Algebra

Euclidean Distance
(2-norm, L2-norm)

$$\vec{a} = [2, 3]$$

$(2^2+3^2)^{\frac{1}{2}}$

$= \sqrt{2^2+3^2}$



2-norm could be represented in this form

$$||a||_2 = (2^2 + 3^2)^{1/2}$$

$$||a||_2 = (4 + 9)^{1/2}$$

$$||a||_2 = \sqrt{13}$$

Quick Review: Norm in Linear Algebra

Infinity-norm

example

$$\vec{a} = [2, 3] \qquad ||a||_{\infty} = 3$$

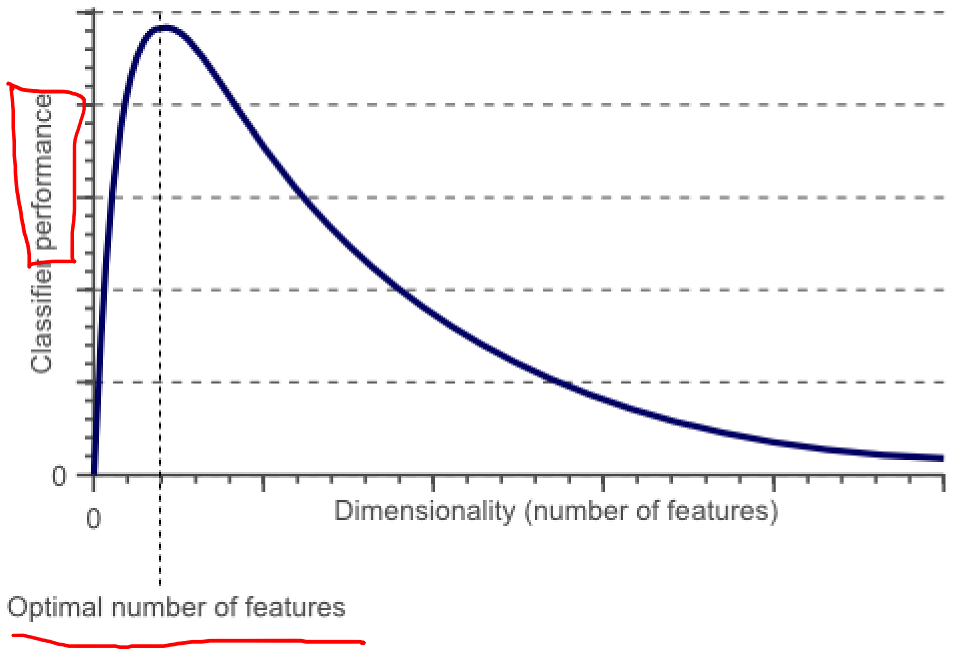
$$\vec{b} = [4, 3, -6] \qquad ||b||_{\infty} = 6$$

$\cancel{p \to d}$ $\lim_{p \to \infty} \Rightarrow |\text{vector}|$ max

Infinity-norm returns maximum absolute value in the given vector

What is Curse of Dimensionality?

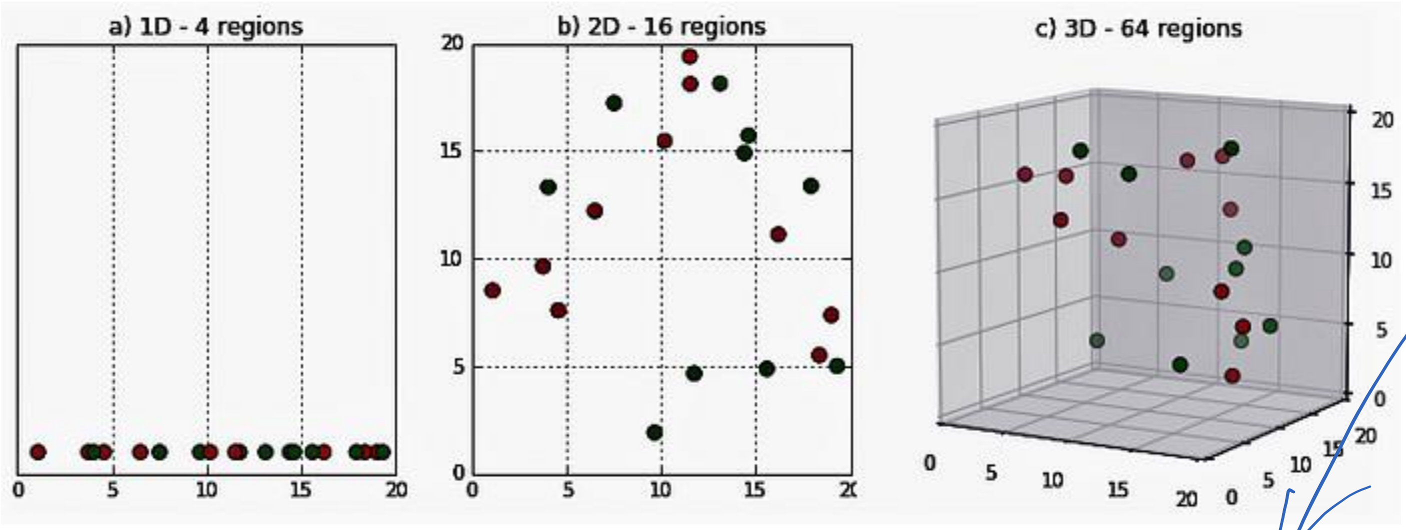
Wrong example

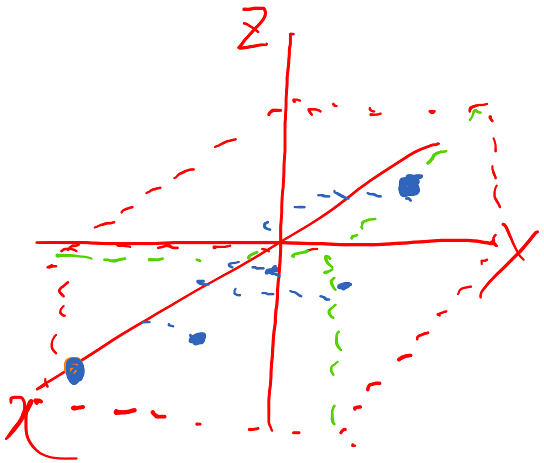


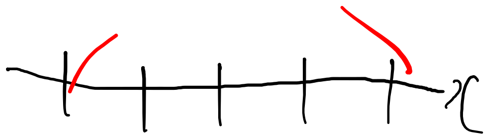
What is Curse of Dimensionality?

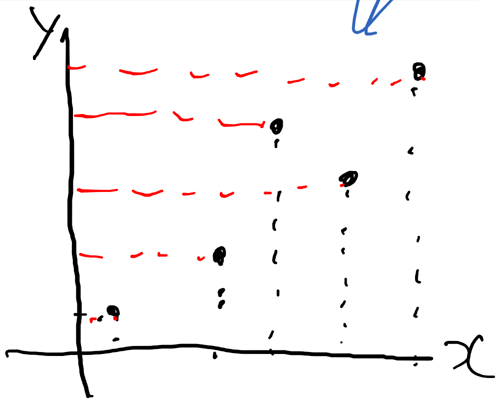
In perspective of density

data sample
$n=20$







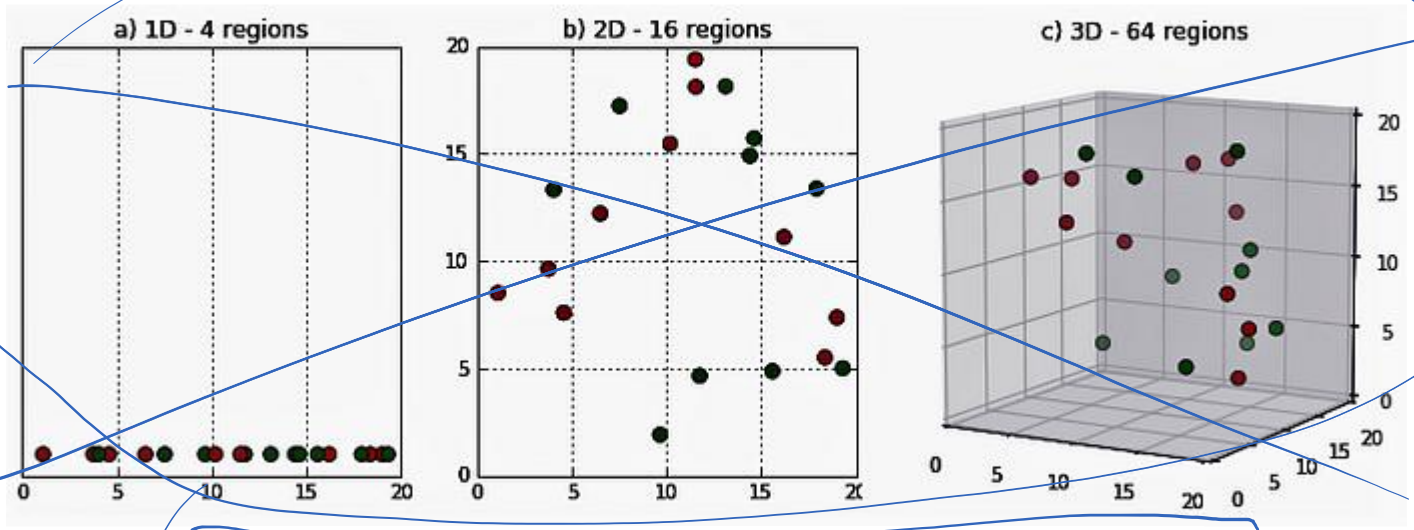


What is Curse of Dimensionality?

In perspective of density

n=5 10

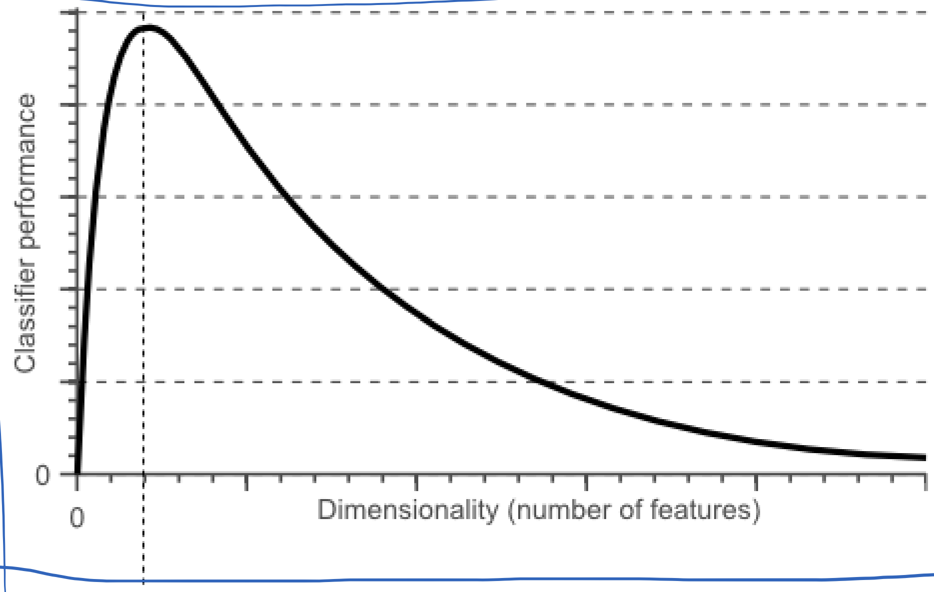
20



100

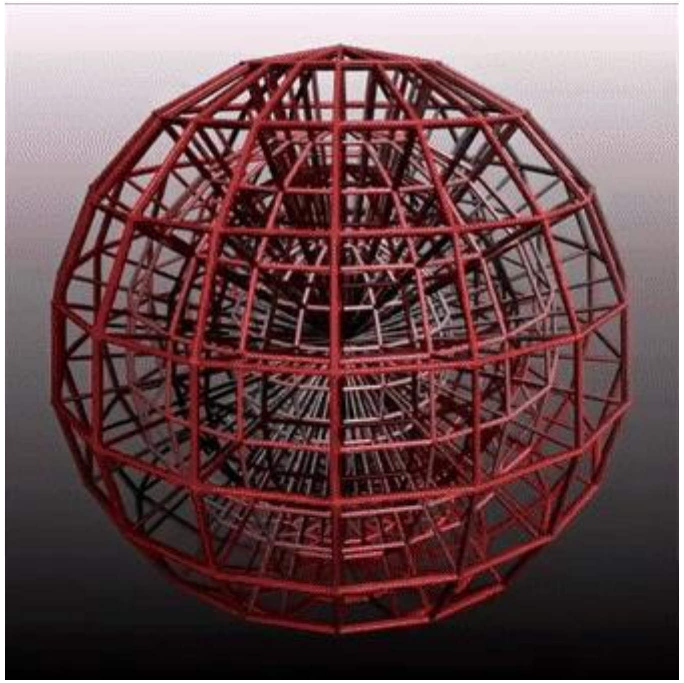
200

1000



What is Curse of Dimensionality?

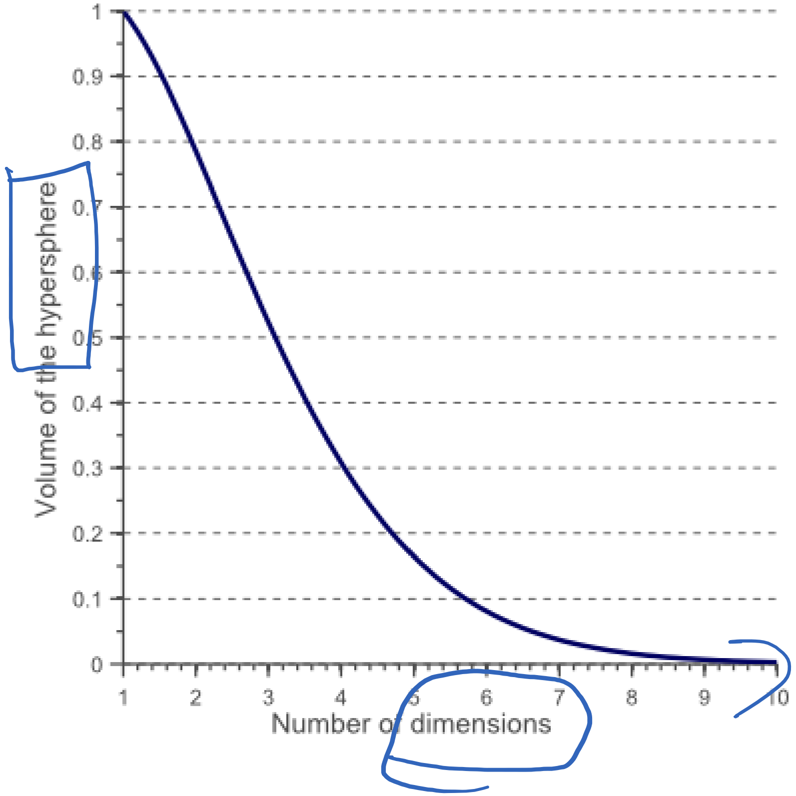
In perspective of density



A hypersphere

$$V(d) = \frac{\pi^{d/2}}{\Gamma(\frac{d}{2}+1)} 0.5^d.$$

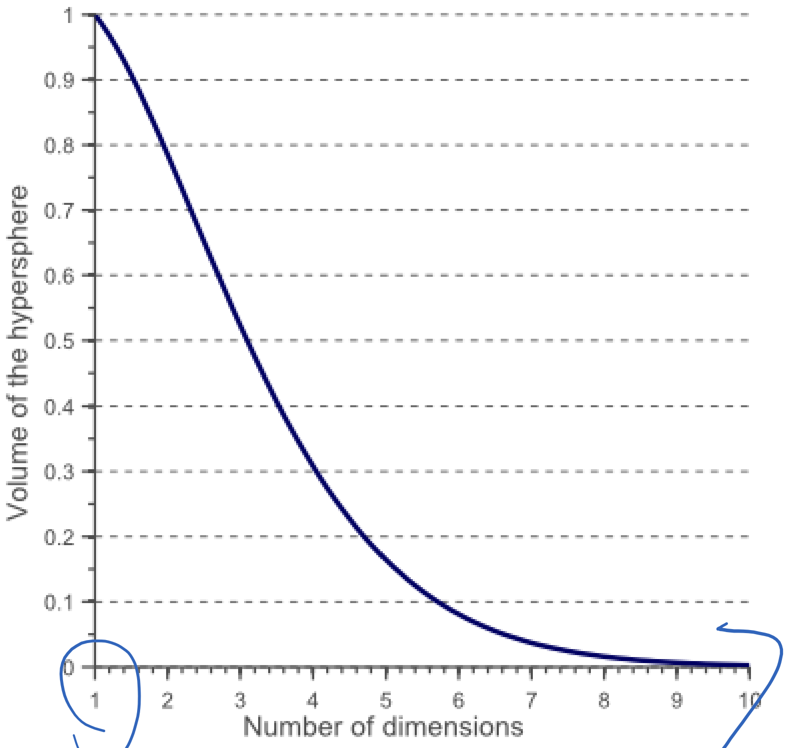
Volume of sphere with d dimension



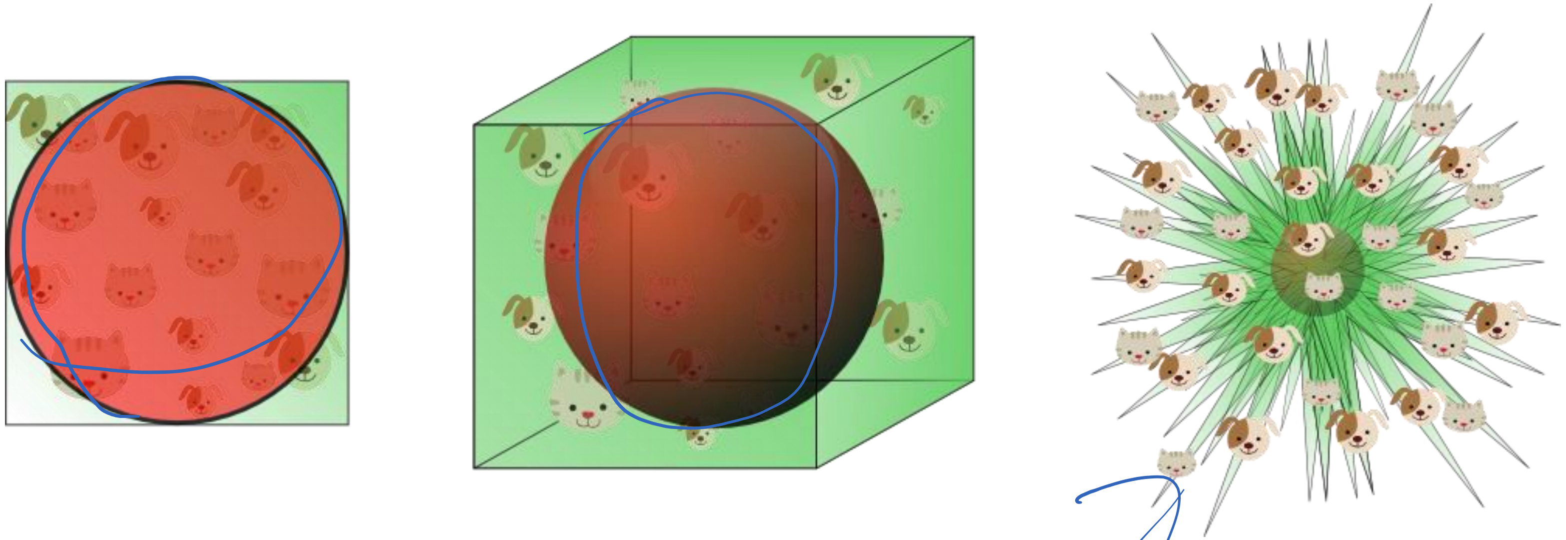
The volume of the hypersphere

What is Curse of Dimensionality?

In perspective of density



The volume of the hypersphere



What is Curse of Dimensionality?

Ways to handle this problem

Reduce the dimension

Or

Get more data

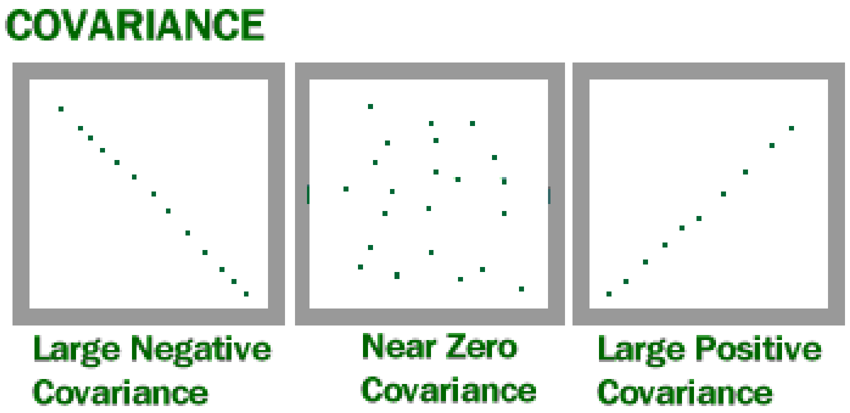
~~ANOVA~~

Projection and Dimensionality Reduction

Covariance

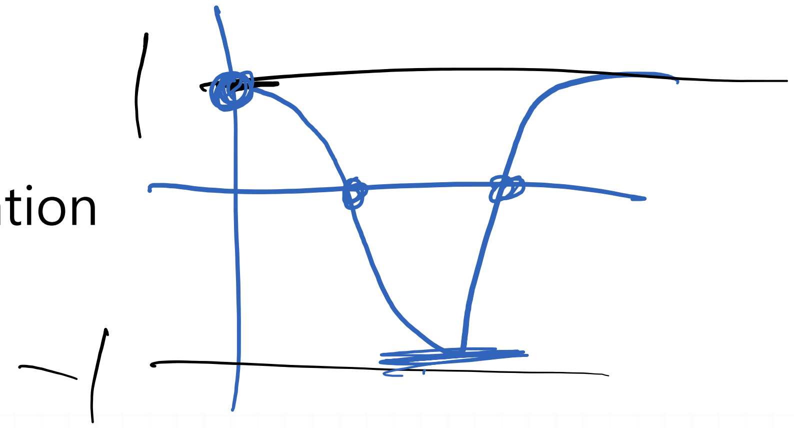
$$\text{Covariance}(x,y) = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})$$

for multivariate



Projection and Dimensionality Reduction

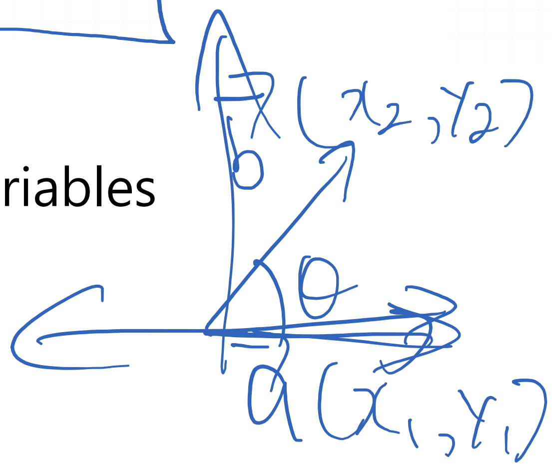
Covariance



Covariance != Correlation

$$\text{Covariance}(x,y) = \boxed{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}$$

$$\rho_{xy} = Correlation\ (x,y) = \frac{cov(x,y)}{\sqrt{var(x)}\sqrt{var(y)}}.$$



Joint variability of two random variables

Range: $[-infty, infty]$

Relationship between two variables

Range: $[-1, 1]$

$$\boxed{\frac{\vec{a}\cdot\vec{b}}{|\vec{a}||\vec{b}|}} = \cos\theta$$

Projection and Dimensionality Reduction

Principal Component Analysis (PCA)

4

rank 4

4

×2

rank 3

A

rank

Singular matrix

Principal Component Analysis

Extracting Principal Components

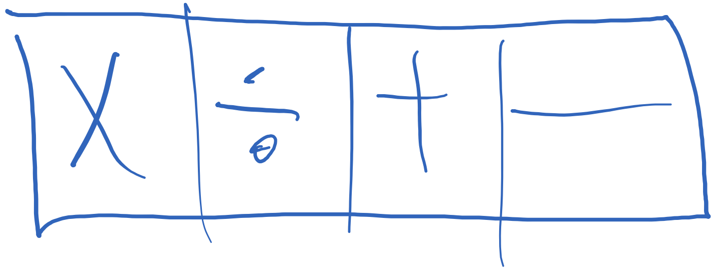
SVD Singular Value Decomposition

How do we extract?

Eigenvalue, Eigenvector!

Projection and Dimensionality Reduction

Principal Component Analysis (PCA)



row space
column space

Eigenvalue decomposition
= Linear transform of a matrix

$$Av = \lambda v$$

A is a square matrix
v is an eigenvector
lambda is an eigenvalue

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \lambda \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$$

n n

Number of Eigenvector can be 0 to n

A

$$\begin{pmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} = \begin{pmatrix} \lambda_1 v_{11} & \lambda_2 v_{12} & \lambda_3 v_{13} \\ \lambda_1 v_{21} & \lambda_2 v_{22} & \lambda_3 v_{23} \\ \lambda_1 v_{31} & \lambda_2 v_{32} & \lambda_3 v_{33} \end{pmatrix}$$

If the result of linear transform is multiplication non-zero itself: eigenvector

Multiplied constant: eigenvalue

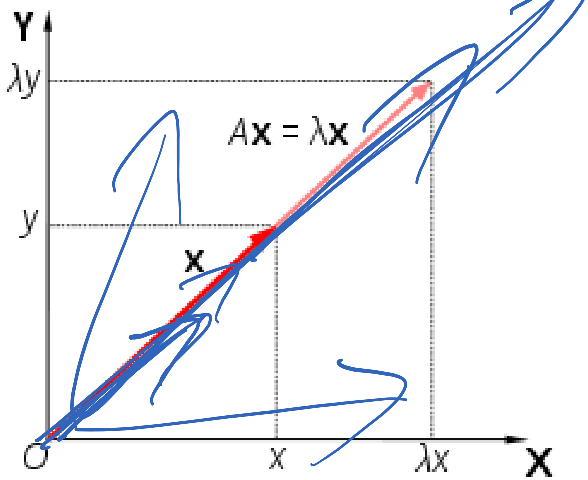
Eigenvalue and Eigenvector of A
: λ_i, v_i

Projection and Dimensionality Reduction

Principal Component Analysis (PCA)

Operation in the previous slide is only for square matrix

For non-square matrix?
➔ Singular Vector Decomposition (SVD)

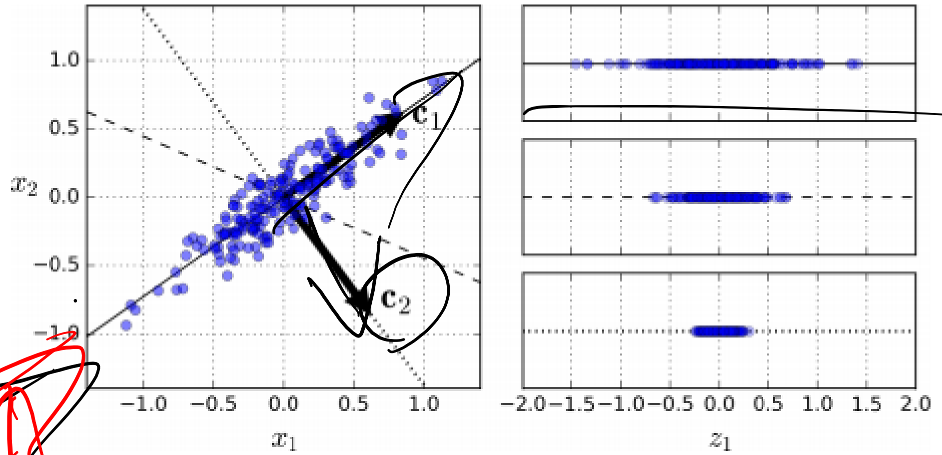


Matrix A: stretched by x

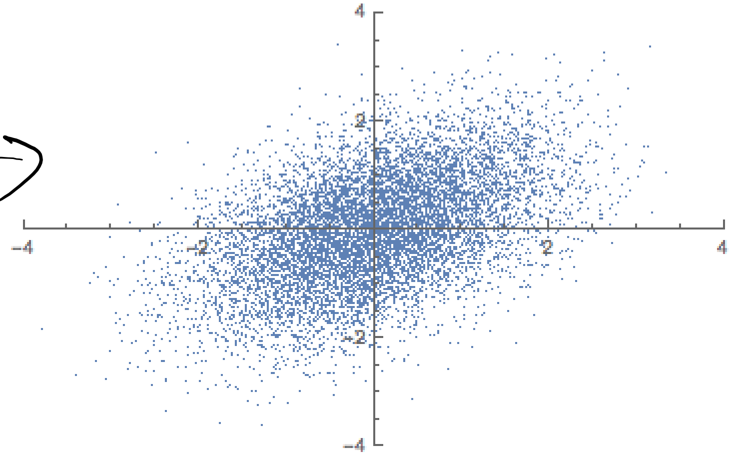
X didn't change A's direction
➔ X is an eigenvector of A!

Projection and Dimensionality Reduction

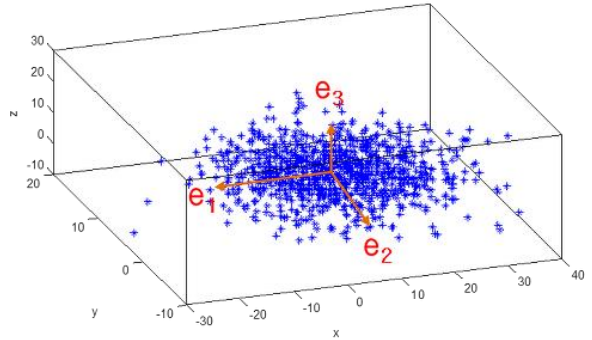
Principal Component Analysis (PCA)



Why should we find the lines that can keep the variance

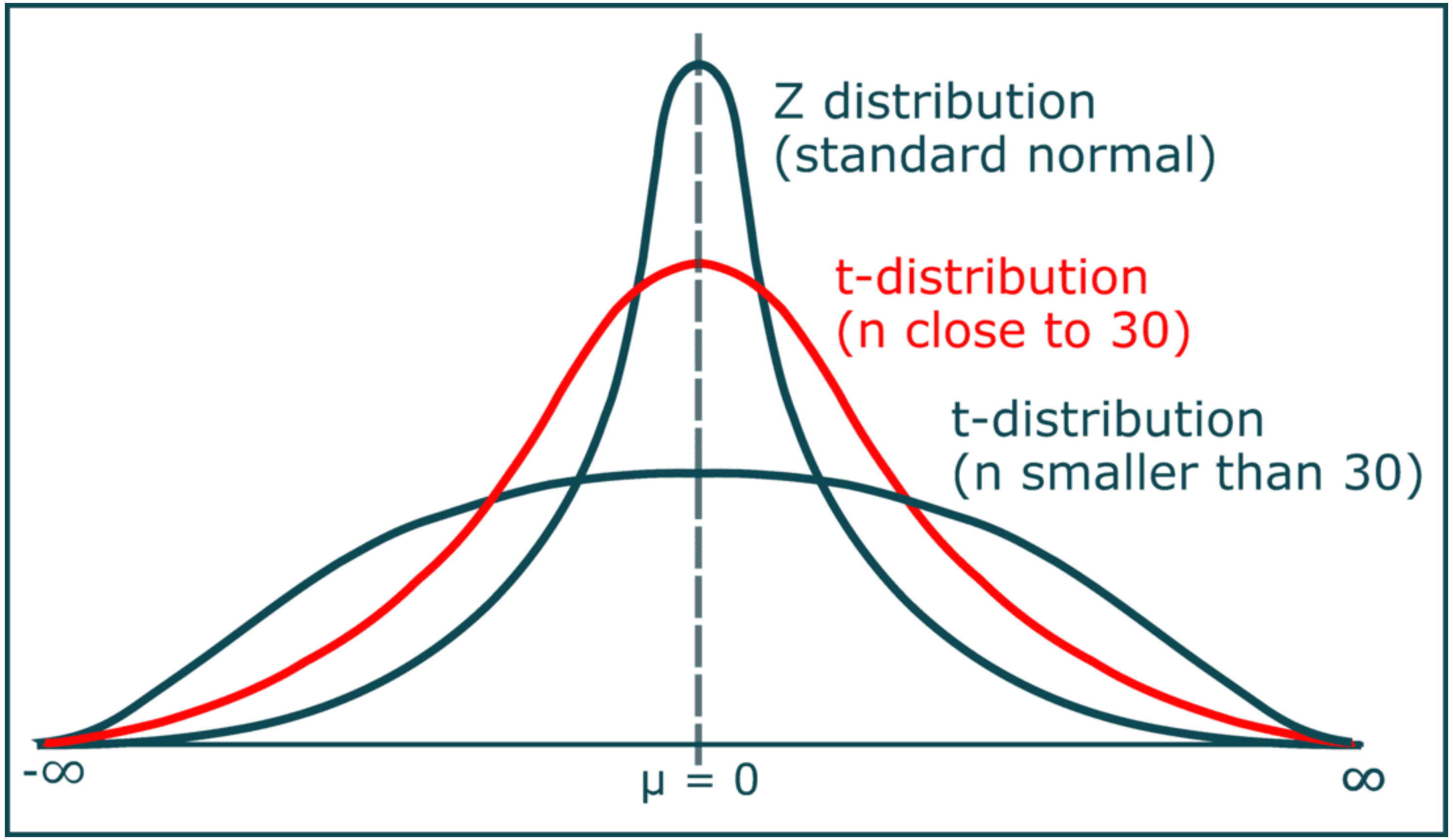


Variance == Information



Projection and Dimensionality Reduction

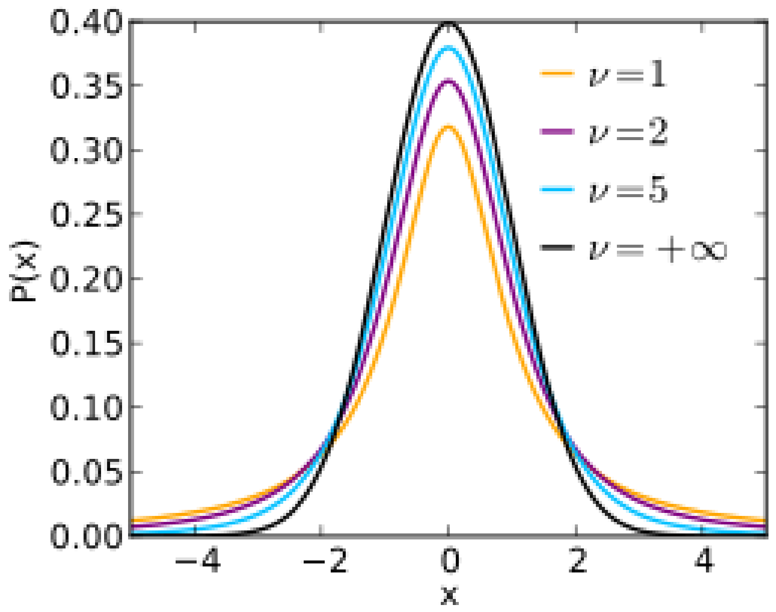
t distribution



Student's t-distribution
Inference about the population mean when the sample size is small
Or
Population standard deviation is unknown

Projection and Dimensionality Reduction

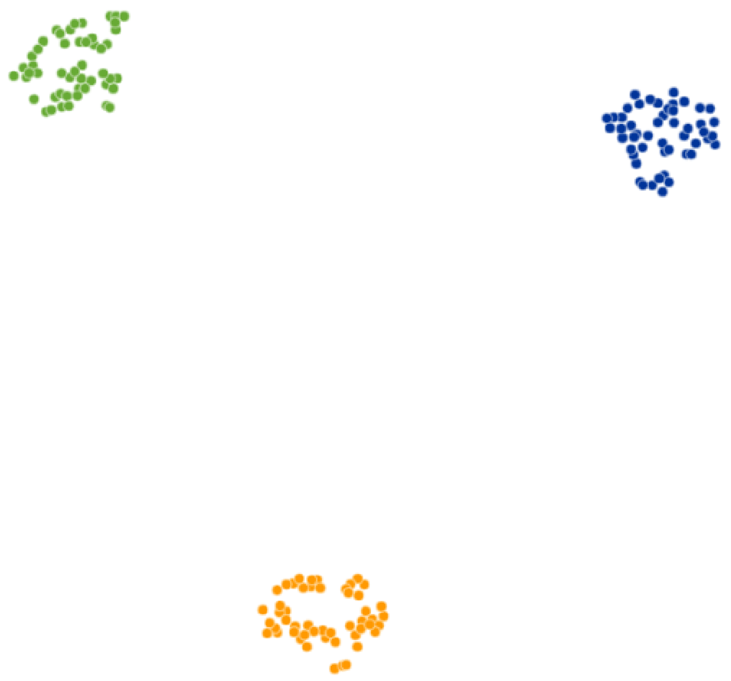
t distribution



∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291
One-sided	75%	80%	85%	90%	95%	97.5%	99%	99.5%	99.75%	99.9%	99.95%
Two-sided	50%	60%	70%	80%	90%	95%	98%	99%	99.5%	99.8%	99.9%

Projection and Dimensionality Reduction

t-SNE
t-distributed Stochastic Neighbor Embedding



Plotting high-dimensional data into 2D

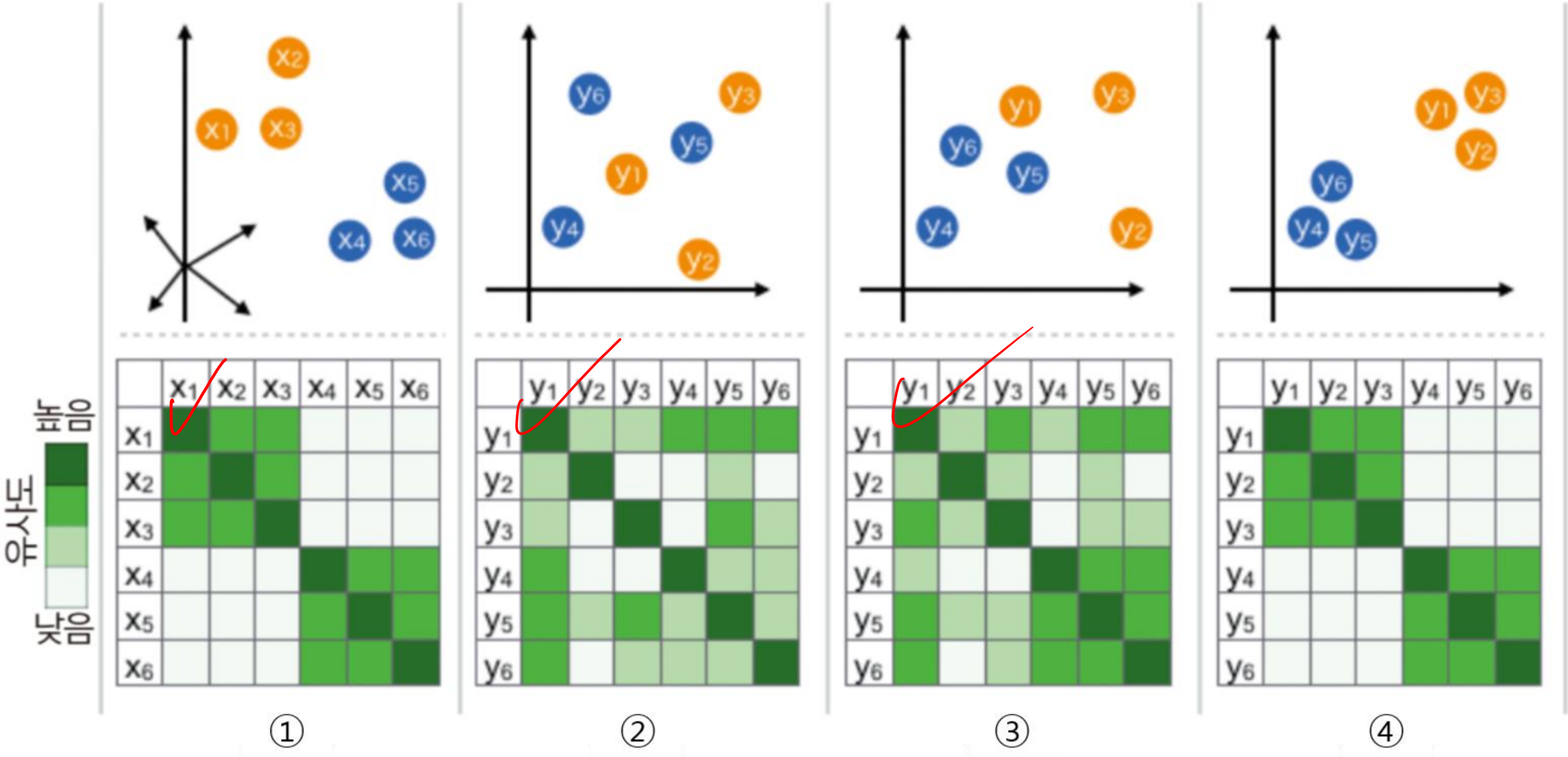
Projection and Dimensionality Reduction

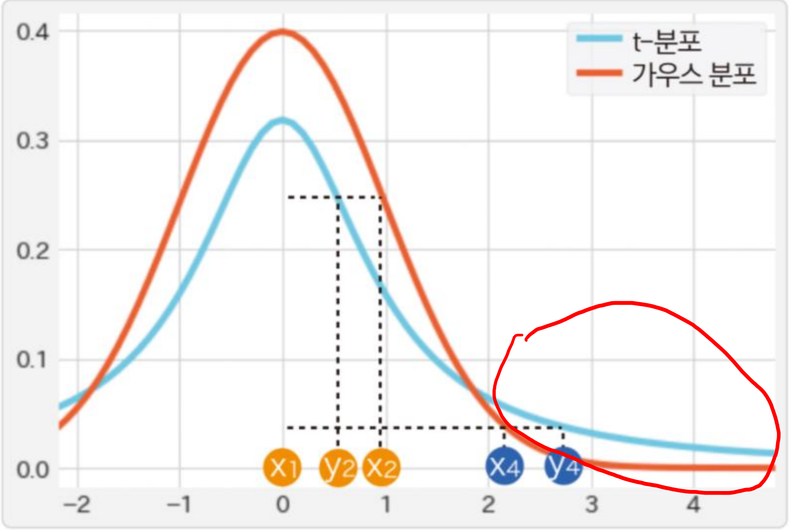
t-SNE vs PCA

t-SNE	PCA
- Nonlinear technique - Focuses on preserving the similarities between data points in a lower-dimensional space	- Linear technique - Focuses on preserving variance

Projection and Dimensionality Reduction

t-SNE





high dimensional data

Similarities in the heatmap are Gaussian distribution, similarity of x_i, x_j

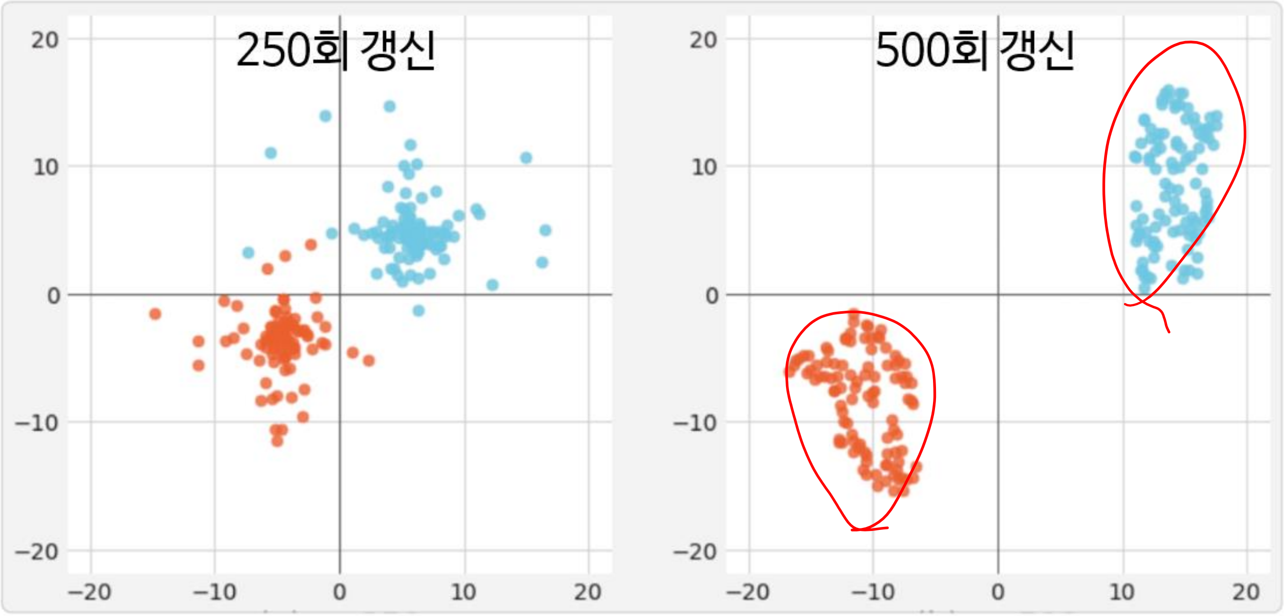
init Map y_i into lower dimension randomly

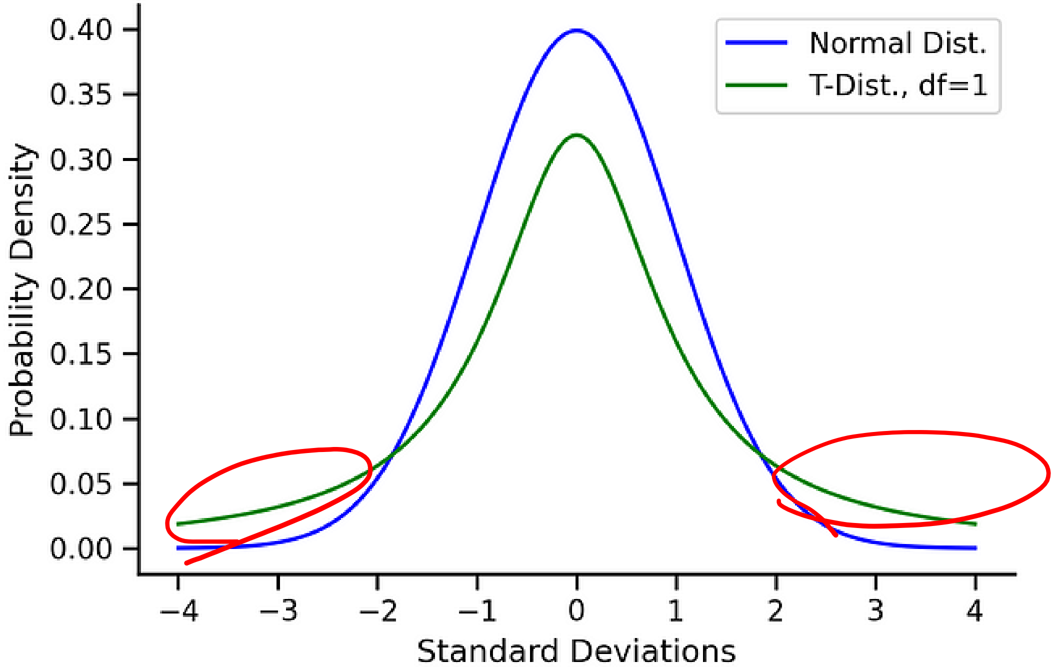
Calculate similarity of y_i, y_j for all i, j pairwise using t-distribution

plotted data

Projection and Dimensionality Reduction

t-SNE





Why they used t-distribution?

Heavily tailed ➔ Data in tail are more likely to be selected

References

A Basic Overview of Using t-SNE to Analyze Flow Cytometry Data – Marissa Fahlberg, PhD

t-SNE 개념과 사용법 - gaussian37

Eigen decomposition and Principal Component Analysis | Machine Learning | Clairvoyant Blog (clairvoyantsoft.com)