

# Curse of Dimensionality

차원의 저주

차원  $\uparrow$  data 조(조)하게 분포  $\rightarrow$  학습이 어려워진다

(빈공간이 늘어남)

(modeling의 난이도  $\uparrow$ )

데이터의 특징을 살리면서 낮은 차원에 표현한다

example

Mnist  $28 \times 28 = 784$  dimension

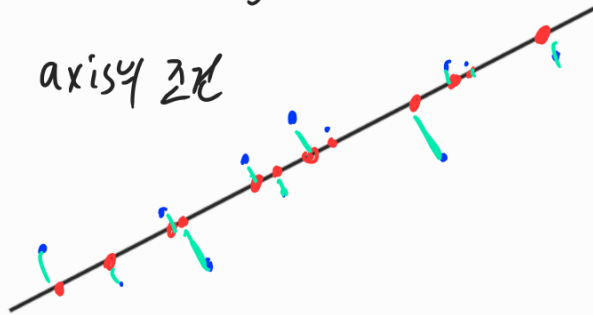
필요한 정보는 제(제)항(항)이 차원 축소 필요

PCA 축소방법

샘플들의 분포가 있을 때 샘플들 잘 설명하기 위해

새로운 axis(축)을 찾아서 데이터

새로운 axis의 조건



1. 빨간점 사이의 거리(거리)가 최대

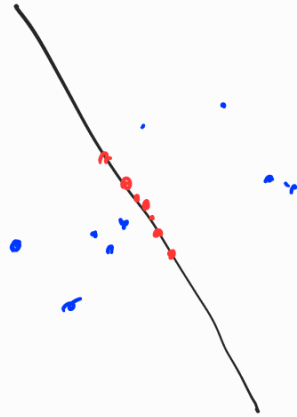
2. 파란점과 빨간점 사이의 거리(거리)가 최소

1. 정보의 손실

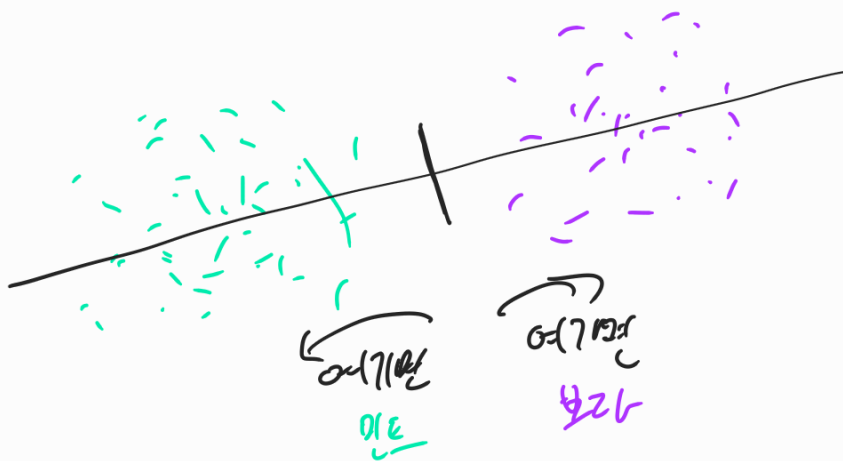
\* 빨간점과 파란점에서  
가장 가까운 점

(PCA, projection)

안좋은 axis

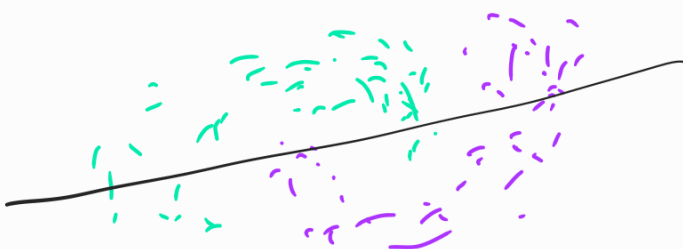


PCA를 통한 classification



Good

Limitation of Linear Dimensional Reduction

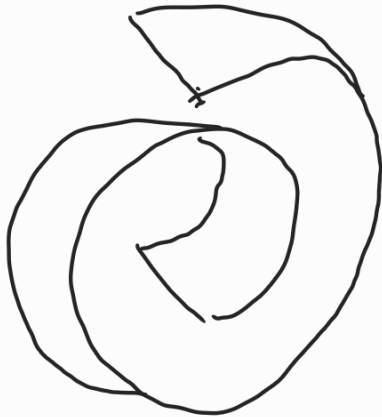


? 아니

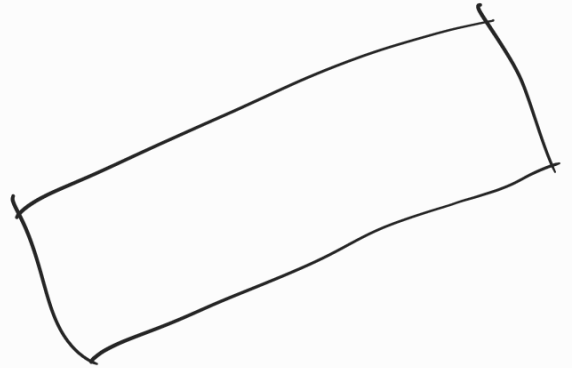
그래서 Deep Neural net  
비선형성까지 표현 가능

# Manifold Hypothesis

- 고차원 공간의 샘플들이 저차원 다양체(Manifold)의 형상을 나타내어 있다는 가설



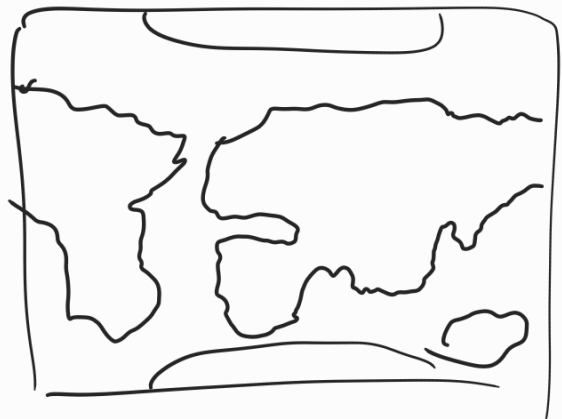
고차원 공간



저차원 공간

example

지구



2차원 좌표계로 인식

고차원 공간에서의 두 점 사이의 거리는 저차원 공간으로의 투영 후 거리와 다르다

저차원에서는 두 data 한 개 간에 관계가 존재한다.



비선형적 차원 축소 example

