CSE 575: Statistical Machine Learning (Spring 2021)
Instructor: Nupur Thakur

Dimensionality Reduction

Arizona State University

Table of contents

Dimensionality reduction

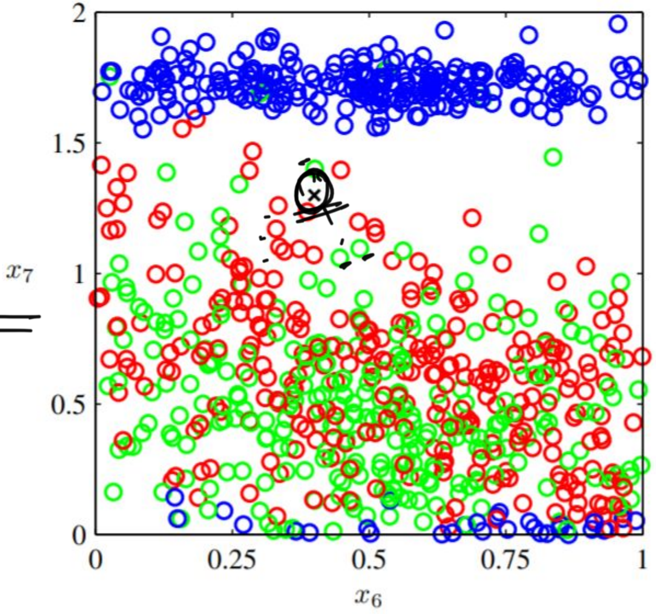
$784 \Rightarrow 28 \times 28$

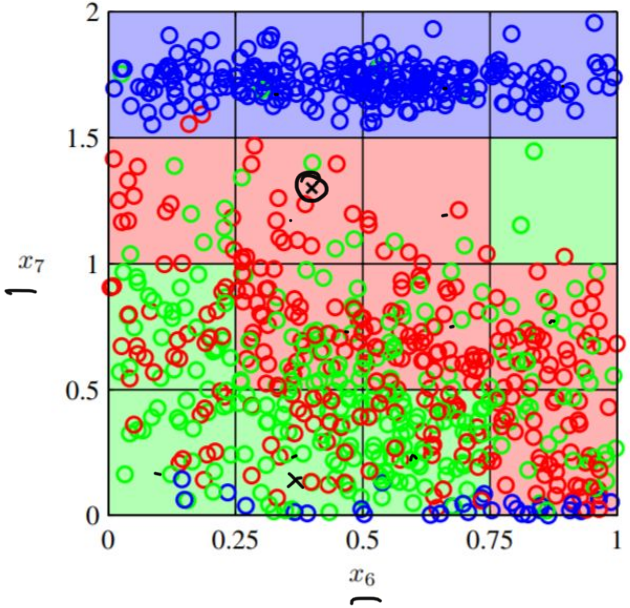
2 3

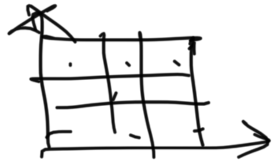
- Given N data points in a high-dimensional space (in the order of tens of thousands of dimensions), project them into some low-dimensional space.
- Why project them into some low-dimensional space?

– Curse of dimensionality.

Curse of dimensionality





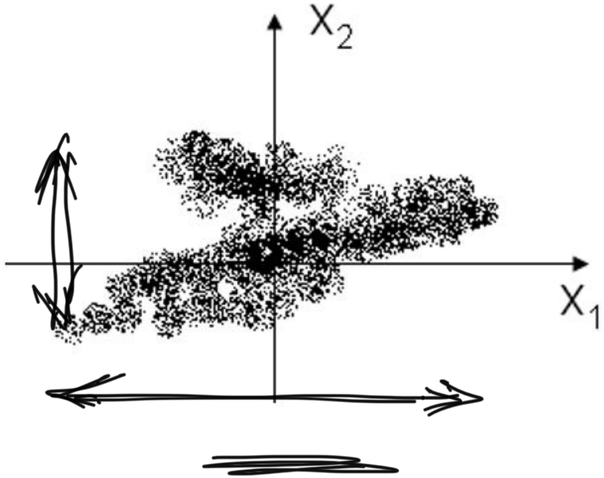


From ‘Pattern Recognition and Machine Learning’, by Bishop book. (fig no. 33, Section 1.4)

Principal Component Analysis (PCA)

- Basic idea -

2D to 1D



Different ways -

(1) Random

(2) Discard the less descriptive

(3) Project onto a space such that the max- variance is obtained.

PCA Problem Formulation

d dimensions

Problem

Given n samples $D=\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ in d-dimensional space, find a direction $\mathbf{e}_1$, such that the projection of D onto $\mathbf{e}_1$ gives the largest variance (compared with any other direction).

$e_1 \Rightarrow$ unit vector , $\|e_1\| = 1$

Projections – $y_i = x_i \cdot e_1$ Inner product.

Mean of projections – $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\sum_{i=1}^{n} x_i \cdot e_1 = \bar{x} \cdot e_1$

PCA Problem Formulation

Variance of projections –

$$\sigma^2(e_1) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 = \frac{1}{n}\sum_{i=1}^{n}\left[(x_i - \bar{x})e_1\right]^2$$

$$= \sum_{j=1}^{d}\sum_{k=1}^{d}\underbrace{\left[\frac{1}{n}\sum_{i=1}^{n}(x_{i,j} - \bar{x}_{i,j})(x_{i,k} - \bar{x}_{i,k})\right]}_{\text{Covariance}} e_j e_k$$

$$= \sum_{j=1}^{d}\sum_{k=1}^{d} e_j e_k C_{jk}$$ where $C \rightarrow$ covariance matrix.

$$\sigma^2(e_1) = e^t C e$$ $\Rightarrow$ Matrix form.

PCA Problem Formulation

$$e = \underset{e}{\operatorname{argmax}}\ \sigma^2(e) \quad \text{subject to } \|e\| = 1$$

Lagrange's multiplier

$$F(e) = \underset{e}{\operatorname{argmax}}\ \sigma^2(e) - \lambda(e^t e - 1)$$

$$\sigma^2(e) \rightarrow e^t C e$$

$$\frac{\partial F(e)}{\partial e} = 2Ce - 2\lambda e = 0 \Rightarrow \boxed{Ce = \lambda e}$$

$e \rightarrow$ eigenvector of covariance matrix C

$\lambda \rightarrow$ eigenvalue corresponding to eigenvector e.

PCA - Principal Components

- What are principal components?

— e_i vectors.

— First principal component – eigenvector with the largest eigenvalue.

- How many principal components to keep?

Total variance – sum of variance in all the projections.

$$= \sum_{j=1}^{d} \lambda_j$$

$d' \Rightarrow$ new no. of dimensions where $d' \ll d$

$$\boxed{\frac{\sum_{j=1}^{d'} \lambda_j}{\sum_{j=1}^{d} \lambda_j}} \sim$$

$D = [n \times d]$ $n \times d'$

PCA Algorithm

$d \to$ total no. of dimensions of data.

1. Compute the dxd sample covariance matrix C

2. Find the eigenvalues and corresponding eigenvectors of C

3. Project the original data onto the space spanned by the eigenvectors
   - The projection may be done onto a d'-dimensional subspace spanned by the first d' eigenvectors (ordered by the eigenvalue in descending order)
     - d' is determined by the desired accuracy

Questions?