Introduction to Dimension Reduction

H.G Min

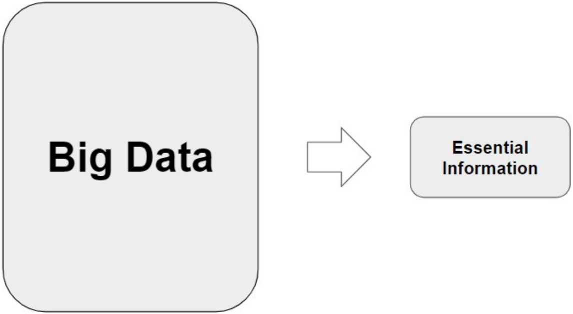
Yonsei University, Korea Quantum Computing

mhg9511@gmail.com

August 14, 2022

Overview

What is Dimension Reduction?



Dimension Reduction is a field to study shrinking big data into smaller data containing essential information of the original data.

Why we need the Dimension Reduction?

Curse of Dimensionality

The Curse of Dimensionality refers to various phenomena that arise when analyzing and organizing data in high-dimensional spaces

- Needs Huge Resources
- Big Sparsity Problem
- Degree of Freedom
- etc..

Various Meaning of the Dimension Reduction

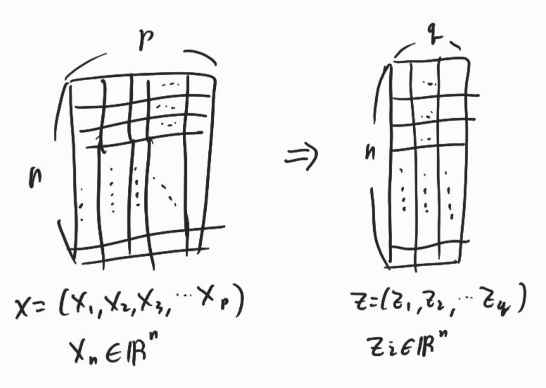
The expert of the Dimension Reduction usually consider it as **Statistics** itself.
ex) Expectation, Covariance, Modeling

Little bit narrowly, the dimension reduction is considered as shrinking $n \times p$ size of data matrix into either $m \times p$ or $n \times q$.

- Shrinking sample size n into m could be considered as **Clustering**.
- Shrinking feature size p into q is done by **Feature Selection** or **Feature Extraction**.

The most of the data scientists consider the **Dimension Reduction** as **Feature Extraction**.

Feature Extraction



The Feature Extraction means making new feature containing essential information of the Original Matrix. It could be formulated as $Z_i = f_i(X)$

Feature Extraction needs **Framework** to make the new feature and **criterion** to find parameters.

Principal Component Analysis (PCA)

1. Framework - Linear Transformation

PCA use linear transformation to make new feature.

$$Z_i = X\beta_i = X_1\beta_1 + X_2\beta_2 + X_3\beta_3 + ... + X_p\beta_p$$

2. Criterion - Variance

PCA make new feature to contain maximal variance.

$$\max_f Var(Z) \text{ subject to } \| f \| = 1$$

Principal Component Analysis (PCA)

$$\max_{\beta} Var(X\beta) \text{ subject to } \parallel \beta^t \beta \parallel = 1$$

$$\max_{\beta,\lambda} \beta^t \Sigma \beta - \lambda(\beta^t \beta - 1)$$

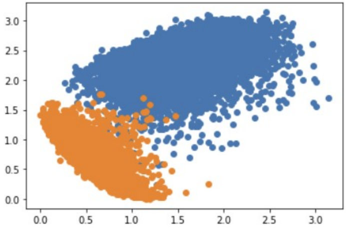
$$\frac{\delta}{\delta\beta}\left[\beta^t \Sigma \beta - \lambda(\beta^t \beta - 1)\right] = 2\Sigma\beta - 2\lambda\beta = 0$$

$$\frac{\delta}{\delta\lambda}\left[\beta^t \Sigma \beta - \lambda(\beta^t \beta - 1)\right] = -\beta^t \beta + 1 = 0$$

$$\Sigma\beta^* = \lambda\beta^*, \beta^{*t}\beta^* = 1$$

PCA Example


0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9


3.0
2.5
2.0
1.5
1.0
0.5
0.0
0.0
0.5
1.0
1.5
2.0
2.5
3.0

Properties of the PCA

- Ordered by Importance
  Eigenvalues are directly used as importance index.
- Orthogonality
  All of principal components are orthogonal.
- Linear Transformation
  New variables are made by linear transformation.
- Linear Correlation
  Only linear correlation is used at criterion.
- Sensitive at the Scale.
  All of the variables have to be scaled.

Various Algorithms of Dimension Reduction

1. Kernel Principal Component Analysis (Kernel PCA)
   Make new variables using Nonlinear Transformation.
2. Sparse Principal Component Analysis (Sparse PCA)
   Use smaller numbers of original variables.
3. Multiple Correspondence Analysis (MCA)
   Used at the Categorical Variables.
4. Independent Componenet Analysis (ICA)
   Statistical correlation is used at the criterion.

Sufficient Dimension Reduction (SDR)

Sufficient Dimension Reduction (SDR) is a field to study dimension reduction methods fitted for the modeling. Unlike the other dimension reduction methods, it uses dependent variable as reference so it is supervised learning.

The main target of the SDR Span(β) is called as **Central Space** or **SDR Space** and denoted by $S_{Y|X}$.

$$Y \perp X \mid \beta^t X$$

$$f(Y \mid X) = f(Y \mid \beta^t X)$$

Properties of SDR space

- Ordered by Importance
  Eigenvalues are directly used as importance index.
- Linear Transformation
  New variables are made by linear transformation.
- Use Dependent Variables
  It use the dependent variables
- Target to find general distribution of the Y.
  Target of the SDR is $f(Y \mid X)$ not just $E(Y \mid X)$.

Algorithms of SDR

- Sliced Inverse Regression

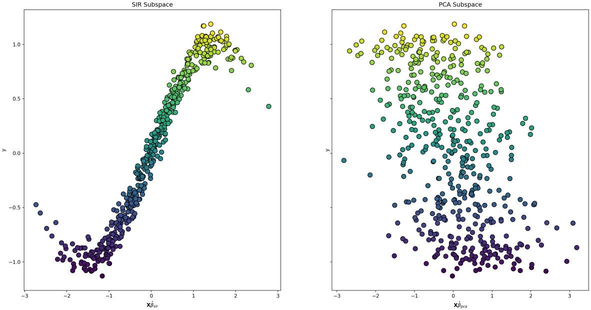
$$Span(Cov(X)^{-1}(E(X \mid Y) - E(X))) \subset S_{Y|X}$$

- Contour Regression

$$Span(2I_p - A(\epsilon)) \subset S_{Y|Z}$$

$$A(\epsilon) = E((Z - \tilde{Z})(Z - \tilde{Z})^t \,||\, Y - \tilde{Y} |< \epsilon)$$

SDR Example


SIR Subspace
PCA Subspace
y
1.0
0.5
0.0
−0.5
−1.0
−3
−2
−1
0
1
2
3
$\mathbf{X}\hat{\beta}_{sir}$
$\mathbf{X}\hat{\beta}_{pca}$

Reference

- https://hgmin1159.github.io/
- Bing Li (2018) Sufficient Dimension Reduction: Methods and Applications with R