Representative approach for big data dimension reduction with binary responses

Xuelong Wang, Jie Yang

July 16, 2019

Fundamental assumption

Let random variable $X \in \mathbb{R}^p$, $Y \in \mathbb{R}$ and $\eta \in \mathbb{R}^d$, where $d << p$

$$Y|X \sim Y|\eta^T X$$

Example

Sufficient dimension reduction

Dimension-reduction subspace

$$Y\mathbf{X}|\eta^T X \rightarrow Y\mathbf{X}|(\eta A)^T X \rightarrow Y\mathbf{X}|P_S^T X,$$

Where $P_{\mathcal{S}}$ is the projection matrix of subspace $\mathcal{S}$
$\mathcal{S}$ is called the dimension-reduction subspace
However, the $\mathcal{S}$ is not unique, i.e. if $\mathcal{S} \subset \mathcal{S}_1$, then $\mathcal{S}_1$ is also a dimension-reduction space.

Central Subspace

$$S_{Y|X} = \cap S_{SDR}$$

The target of sufficient dimension reduction is to estimate the structure of S_{Y|X}

Estimating the central subspace

Sliced type regression

One potential issuse is that we cannot recover all the central space.

Problem

Binary response

Limited the number of sliced For sir, it can only find one dimension the save stil possible find all the dimensions based on two sliced, but it still surfured from this

Probability Enhanced method

Representative approach

method

clustering steps also reducing the data size

parallel method