Slow Learning for Quadratic Classification of Poorly-Posed Observations

An Iterated Linear Dimension Reduction Approach

Dr. Gabriel Odom

January 16, 2017

Overiew of Topics

- Introduction of the problem
- Four HLDR Techniques
- Two Dimension Reduction Methods
- Simulation study
- Real Data Case
- Conclusion
- Future work
- References

Smile! :) You got this!

Introduction

Introduction

- Classification of observations from different elliptical distributions
- Real examples are often poorly posed ($n_i < p^2/2$)
- Introduce a competetor to PCA using the bias-variance tradeoff
- Dimension reduction via iterated singular value decomposition application

Four HLDR Techniques

- Ounpraseuth et al (SY)
- Loog and Duin (LD)
- Li's Sliced Inverse Regression (SIR)
- Cook and Weisberg's Sliced Average Variance Regression (SAVE)

Current vs Slow Learning *LDR* Algorithms

Sufficient Data Matrices

Consider p-dimensional multivariate normal data from K classes, with class means $\boldsymbol{\mu}_k$ and class covariances $\boldsymbol{\Sigma}_k$, for $k \in 1, \ldots, K$. For classes with unequal covariance matrices, we choose the data sufficiency matrix

$$ \tag{1}$$

PCA

The Principal Components Analysis (PCA) Algorithm for dimension reduction is as follows:

1. Whiten the $n \times p$ data matrix.
2. Construct $\mathbf{M}$.
3. Take the Singular Value Decomposition of $\mathbf{M} = \mathbf{UDV}^T$.
4. Choose a target dimension q such that the largest q singular values of $\mathbf{M}$ account for at least $(1-\alpha)\%$ of the sum of all singular values (the energy) of $\mathbf{M}$.
5. Select the first q singular vectors of $\mathbf{U}$ to be the $p \times q$ projection matrix.
6. Multiply the $n \times p$ data matrix by the $p \times q$ projection matrix to linearly reduce the data from p to q dimensions, while preserving $(1-\alpha)\%$ of the energy of the data matrix.

Slow Learning

The Slow Learning Linear Dimension Reduction ($SLLDR$) Algorithm for dimension reduction is as follows:

1. Whiten the $n \times p$ data matrix. $\text{for}(i \text{in} 1 : \text{steps})$
2. Construct $\mathbf{M}$.
3. Take the Singular Value Decomposition of $\mathbf{M} = \mathbf{UDV}^T$.
4. Choose a target dimension q_i such that the largest q_i singular values of $\mathbf{M}$ account for at least $(1-\alpha/\text{steps})\%$ of the sum of all singular values (the energy) of $\mathbf{M}$.
5. Select the first q_i singular vectors of $\mathbf{U}$ to be the $p \times q_i$ projection matrix.
6. Multiply the $n \times p$ data matrix by the $p \times q_i$ projection matrix to linearly reduce the data from p to q_i dimensions. $\text{end}(\text{for})$
7. Return the new $n \times q_{\text{steps}}$ data matrix

Summary and References

Summary

- Looked at two samples of high-dimensional data
- Considered four methods of testing equality of two high-dimensional mean vectors
- Used singular valued decomposition to reduce dimension of data sets
- Considered two methods of quantile estimation
- Compared methods via power simulation

References

. . .