Principal Components Analysis

Prof. Sharon McNicholas

STATS 780/CSE 780

Introduction

- When we looked at the mixture of factor analyzers model, and the PGMM family, we saw examples of simultaneous dimension reduction and clustering.

- In general, there are many reasons why we might want to reduce the dimensionality.

- In this "lecture", we will look at a famous data reduction approach; principal components analysis (PCA).

- PCA finds a number of uncorrelated variables (principal components) that explain "enough" of the variance in the observed data.

What is a Principal Component?

- The first principal component is the direction of most variation (in the data).
- The second principal component is the direction of most variation (in the data) conditional on it being orthogonal to the first principal component.
- The third principal component is the direction of most variation (in the data) conditional on it being orthogonal to the first two principal components.
- For $r > 1$: the rth principal component is the direction of most variation (in the data) conditional on it being orthogonal to the first $r - 1$ principal components.

Definition

- Let $\mathbf{X}$ be a p-dimensional random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.
- Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ be the (ordered) eigenvalues of $\boldsymbol{\Sigma}$, and let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ be the corresponding eigenvectors.
- The ith principal component of $\mathbf{X}$ is

$$W_i = \mathbf{v}_i'(\mathbf{X} - \boldsymbol{\mu}),$$

  for $i = 1, \ldots, p$.
- Which we can write as

$$\mathbf{W} = \mathbf{V}'(\mathbf{X} - \boldsymbol{\mu}),$$

  where $\mathbf{W} = (W_1, \ldots, W_p)$ and $\mathbf{V} = (\mathbf{v}_1, \ldots, \mathbf{v}_p)$.

Some Results

- $\mathbb{E}[\mathbf{W}] = \mathbf{0}$.
- $\mathbb{V}\text{ar}[\mathbf{W}] = \mathbf{\Lambda}$.
- $\mathbf{X} = \boldsymbol{\mu} + \mathbf{V}\mathbf{W} = \sum_{i=1}^{p} \mathbf{v}_i W_i$.
- $\sum_{i=1}^{p} \mathbb{V}\text{ar}[W_i] = \sigma_{11} + \cdots + \sigma_{pp}$.

STATS 780/CSE 780 Prof. Sharon McNicholas

Key Results

- Consider $W = \mathbf{v}'(\mathbf{X} - \boldsymbol{\mu})$ with $\mathbf{v}'\mathbf{v} = 1$.
  i. $\mathbb{V}\text{ar}[W]$ is maximized when $W = W_1$, the first principal component.
  ii. If W is uncorrelated with the first $k < p$ principal components $W_1, \ldots, W_k$, the $\mathbb{V}\text{ar}[W]$ is maximized when $W = W_{k+1}$, the $k+1$st principal component.
- The proportion of the total variation explained by the ith principal component is $\lambda_i / \text{tr}\{\mathbf{\Sigma}\}$.
- The proportion of the total variation explained by the first k principal components is

$$\frac{\sum_{i=1}^{k} \lambda_i}{\text{tr}\{\mathbf{\Sigma}\}}.$$

Comments

- Let's look at some PCA examples in R.
- Then, we will look at some examples of PCA followed by clustering.
- Bouveyron and Brunet-Saumard (2014)[a] and others have written about this topic.

[a]Bouveyron, C. and C. Brunet-Saumard (2014). 'Model-based clustering of high-dimensional data: A review'. *Computational Statistics and Data Analysis* **71**, 52–78.