ISLR2 Ch12 Unsupervised Learning

Lydia Gibson

2022-07-02

12.2 Principal Components Analysis

Principal components allow us to summarize large datasets of correlated variables with a smalller number of representative variables that collectively explain most of the variability in the original dataset.

Principal component analysis (PCA) refers to the process by which principal components are computed, and the subsequent use of the components in understanding the data.

The first principal component of a set of features $Z_1, X_2, \ldots, X_p$ is the normalized linear combination of the features $Z_1 = \phi_{11}X + \phi_{21}X + \cdots + \phi_{p1}X_p$ that has the largest variance.

The first principal component loading vector solve the optimization problem:

$\underset{\phi_{11},\ldots,\phi p1}{maximize}\{\frac{1}{n}\sum_{i=1}^{n}(\sum_{j=1}^{p}\phi_{j1}x_{ij})^2\}$ subject to $\sum_{j=1}^{p}\phi_{j1}^2 = 1$

12.3 Missing Values and Matrix Completion

12.4 Clustering Methods

12.5 Lab: Unsupervised Learing

12.6 Exercises