Modern Applied Statistics Chap 11: Exploratory Multivariate Analysis

Yongdai Kim

October 25, 2022

Seoul National University

Outline

Introduction

Introduction

Multivariate analysis is concerned with datasets that have more than one response variable for each observational or experimental unit.

Summary

- X : $n \times p$ data matrices
  - row : observation
  - columns : variables
  - x : the variables of a case by the row vector

The main division in multivariate methods is between those methods that assume a given structure and those that seek to discover structure from the evidence of the data matrix alone. Methods for known structure are considered in Chapter 12.

Introduction

In pattern-recognition terminology the distinction is between supervised and unsupervised methods.

- **Iris data**
  Iris data has 150 cases, which are stated to be 50 of each of the three species *Iris setosa, virginica and versicolor*. Each case has four measurements on the length and width of its petals and sepals.

A *priori* this seems a supervised problem, and the obvious questions are to use measurements on a future case to classify it, and perhaps to ask how the variables vary among the species. However, the classification of species is uncertain, and similar data have been used to identify species by grouping the cases.

Introduction

Krzanowski (1988) and Mardia, Kent and Bibby (1979) are two general references on multivariate analysis. For pattern recognition we follow Ripley (1996), which also has a computationally-informed account of multivariate analysis. Most of the emphasis in the

literature and in this chapter is on continuous measurements, but we do look briefty at multi-way discrete data in Section 11.4. Colour can be used very effective1y to differentiate groups in the

plots of this chapter, on screen if not on paper. The code given here uses both colours and symbols, but you may prefer to use only one of these to differentiate groups.

Visualization Methods

Visualization Methods

Cluster Analysis

Cluster Analysis

Factor Analysis

Factor Analysis

Discrete Multivariate Analysis

Discrete Multivariate Analysis