



UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Statistica Multivariata

2425-1-F8204B002-F8204B003M

Learning objectives

The aim of the course is to introduce some multivariate techniques used in exploratory data analysis and in prediction. The student will be able to apply the aforementioned techniques to real data collected, for instance, in social and economic fields and that can be characterized by high-dimensionality.

Contents

- Multivariate normal distribution in statistics
- Linear methods for classification
- Models for categorical variables
- Models for ordinal variables
- Models for paired or clustered variables

Detailed program

Multivariate normal distribution in statistics

- Definition
- Mahalanobis distance
- Properties
- Maximum likelihood estimation
- Distribution of maximum likelihood estimators
- Wishart distribution
- Tools for testing the assumption of multivariate normality

- Multivariate Box-Cox transformation

Linear methods for classification

- Classification and decision theory
- Bayes classifier
- Naive approach: linear regression for classification
- Discriminant functions
- Linear discriminant analysis
- Similarities between linear discriminant analysis and linear regression
- Quadratic discriminant analysis
- Fisher's approach to linear discriminant analysis
- Naive Bayes method as a special case of discriminant analysis

Models for categorical variables

- Logistic regression
- Maximum likelihood estimation via Newton-Raphson
- Multinomial logistic regression
- Classification using logistic regression
- Parallel between linear discriminant analysis and logistic regression
- Alternative models for binary variables: probit model, complementary log-log model

Models for ordinal variables

- Cumulative logistic model
- Alternative models for ordinal variables: cumulative probit model, cumulative complementary log-log model
- Cox proportional hazards model
- Adjacent category logistic model

Models for paired or clustered variables

- Marginal and conditional approaches
- Conditional logistic model
- Conditional likelihood
- Random effects models
- Rasch model
- Item-response model
- Logistic-normal model
- Conditional and marginal properties of the logistic-normal model
- Potential generalizations
- Random effects models with nonparametric distribution

Prerequisites

Exploratory data analysis, statistical models, probability theory, statistical inference, programming

Teaching methods

Lectures (theory and exercises) and lab sessions.

Specifically, the schedule includes:

7 lectures of 3 hours each (instructional mode in the first part and interactive in the second part)

7 lectures of 2 hours each (instructional mode)

6 lab sessions of 2 hours each (interactive mode)

It is planned that 80% of the sessions will be conducted in-person and 20% remotely.

Assessment methods

The exam is a written test consisting of theoretical questions, theoretical exercises, and numerical exercises to be completed using R.

There are no mid-term exams.

Textbooks and Reading Materials

Hastie, T., Tibshirani, R., Friedman, J.H. and Friedman, J.H., 2009. The elements of statistical learning: data mining, inference, and prediction (Vol. 2, pp. 1-758). New York: springer.

Agresti, A., 2012. Categorical data analysis (Vol. 792). John Wiley & Sons.

Semester

Secondo semesters, first cycle

Teaching language

Italian

Sustainable Development Goals

QUALITY EDUCATION
