MSc in Bioinformatics for Health Sciences

AST. Advanced Statistics

Syllabus Information

Academic Course: 2019/20

Academic Center:

Study: 8045 – Bioinformatics for Health Sciences - MSc

Subject: 32547 - AST Advanced Statistics

Credits: 5.0

Course: 1st

Teaching languages: English

Teachers: Leandro Radusky

Teaching Period: 2nd term

Presentation

This course will focus on the statistical modeling of problems related to biomedical and biological data. Given the wide variety of methods available and the dizzying amount of libraries implementing this methods on different programming languages strong effort will be placed on 'learn to learn' how statistical modeling problems should be addressed. To accomplish this a strong accent will be devoted to the mathematical and algorithmic basis of the discussed methods.

Both frequentist and bayesian approaches will be discussed, while a stronger focus will be put on the last one. At the end of the course causal inference methods will be introduced.

Associated skills

General competences:

1. Developing abilities to develop solutions for real scientific problems that requires statistical modeling.

2. Developing abilities to reading, writing and listening scientific English related to the subject and communicating this scientific research by means of presentations.

Specific competences:

1. Developing R and Python programming skills

2. Developing algorithmic skills by implementing some of the presented methods.

3. Performing statistical modeling on real problems with emphasis in its formal aspects and a communicative visualization.

Learning outcomes

By taking this course, the students will be able to understand and apply statistical modeling into scientific problems.

Prerequisites

Students must have good knowledge in basic Statistics.
It is expected that the students possess some experience in python and/or R programming languages, since both languages will be used during the course.

Contents

The course will include lectures and hands-on exercises covering the following contents

Contents section 1: General concepts.
Estimators.
Random variables.
Discrete and continuous probability distributions.
Correlation and mutual information.
Deterministic and statistical models.
Parametric, nonparametric and semiparametric models.

Contents section 2: What kind of problem do I have?
What I want from my data? Minimization, classification, regression, dimensionality reduction.
What can I learn from my data? Supervised and unsupervised learning.
Feature selection.

Contents section 3: Bayesian inference
Bayes´ theorem.

From priors to posteriors.
Bayesian inference on models.
Hypothesis testing.
Maximum Likelihood estimation.
Markov Chain Monte Carlo.

Contents section 4: From bayesian networks to causal structures inference
Definition of bayesian networks, Hidden Markov Models.
Introduction to causal theory.
From bayesian networks to causal structures.
The PC algorithm.

Teaching methods

The course is mainly a practical course, and each topic is organized around a brief lecture to present the basic concepts behind each topic, followed by hands-on exercises using publicly available resources. At the end of some of the sessions, the student will be asked to deliver the answer of specific exercises. Practical projects emulating real problems will be developed by the students on each unit, justifying the implemented models, formal correctness and putting focus on the communication (visualization) aspects.

Evaluation

Each module will be evaluated through a practical project, representing the 60% of the final note. Also, at the end of the course a final exam will be taken, representing the remaining 40%.

Grading system:
A minimum performance of 50% on each item is required to pass the subject.

Bibliography and Information Resources

The following is the main bibliography to be utilized in each module:
Sections 1 & 2:
Introduction to Data Mining. Pang-Ning Tan, Michael Steinbach, and Vipin Kumar, Addison Wesley.

Section 3:
Davidson-Pilon, Cameron. Bayesian methods for hackers: probabilistic programming and Bayesian inference. Addison-Wesley Professional, 2015.

Section 4:
Shipley, Bill. Cause and correlation in biology: a user's guide to path analysis, structural equations and causal inference. Cambridge University Press, 2016.