Computational analysis of biological structures and networks

Manuele Bicego

Master's degree in Medical Bioinformatics

Computer Science Dept.
University of Verona

Lecturer

	Manuele Bicego
Office	Ca' Vignal 2, Floor 1, Room 48b
Telephone	045 8027072
Email	manuele.bicego@univr.it
Office Hours	Currently: send an e-mail to get an appointment

The course

Timetable:

Tuesday	08.40 - 11.00 (Room G + Online)
Wednesday	11.10 - 13.40 (Lab Alfa + Online)

(note: lab will start in few weeks)

The course

- Lectures:
  - **At the Computer Science Dept.**: you should use the app "UNIVR Lezioni" to book your seat in the room (https://www.univr.it/it/app-univr-lezioni).
  - **On-line**: you should use the e-learning platform (moodle): the zoom link and related material will be posted there the days preceding the lecture.
- Lectures will be recorded and made available (in moodle) until the end of the semester

Please send me an email if you are interested in following the course but you are not allowed to access the moodle space (this is possible if your enrolment procedure is not completed yet)

Requirements

- Preferably: basic notions on Pattern Recognition (e.g. from the course “Riconoscimento e Recupero dell'informazione per Bioinformatica”, BS degree in Bioinformatics)
  - A brief recap will be given at the beginning of the course
- Basic notions of Algorithms, Probability, Statistics, Algebra
- Programming skills (for lab part)
  - Programming language used: Matlab

For students who don’t know Matlab: please send me an email asap!!

Overview

- Title: “Computational analysis of biological structures and networks”
- Many possible viewpoints: here we adopt the Pattern Recognition / Machine Learning perspective

- The course is about **Pattern Recognition / Machine Learning tools** and **techniques** to model biological ***complex*** objects
  - Objects with a ***complex structure*** (strings, 3D structures, sets, graphs, networks...)

Contents

The course is divided in two parts:

PART 1: **Theory**

PR/ML tools and techniques to model structured data

PART 2: **Laboratory**

Implementation of algorithms studied during the theory part (*matlab*)

Program (Theory)

- **Chapter 1.** Introduction
  - Basic Pattern Recognition concepts (recap from Bs. Course)
  - Introduction to structured data (data with complex structure)

- **Chapter 2.** Representation of structured data
  - The Bag of words representation
  - The dissimilarity-based representation
  - Advanced dimensionality reduction techniques

Program (Theory)

- **Chapter 3.** Models for structured data
  - Generative models (Bayes Networks)
  - Learning and inference

- **Chapter 4.** Advanced concepts
  - Kernels for structured data
  - Advanced learning paradigms for structured data (Multiple instance learning, semi supervised learning, transfer learning)

Material

- Slides, notes, suggested readings...
- Slides will be posted on the moodle platform
- All info and news can be found a the e-learning course homepage (news will be also posted in the public course webpage)

Reference books

- R. Duda, P. Hart, D. Stork *Pattern Classification.* Wiley, 2001 (2nd edition).
- S. Theodoridis, K. Koutroumbas: *Pattern Recognition*, Second edition, Academic press, 2003
- C.M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
- B. Frey: *Graphical Models for Machine Learning and Digital Communication*, MIT Press, 1998
- E. Pekalska, B. Duin, *The Dissimilarity Representation for Pattern Recognition*, World Scientific Press, 2005

Some specific readings will be suggested for every chapter

Assessment methods

Two parts:

- First part: **written exam** (during exam sessions)
  - Can be an oral depending on health situation
- Second part: **talk** within a thematic workshop (as in a conference)

Assessment methods

- First part: **written exam** (oral if the case)
  - few questions on course topics
    - Example: “Describe the main properties of Bayesian Networks”
  - one question on the lab part (typically understanding a small piece of code)
    - Example: “Does this matlab code compute the mean of the vector x? Why?”

```
% x is a vector of N entries containing numbers

m = 0;
for i = 1:N
    m = m+x(i);
end
m = m/(N-1);
```

Assessment methods

Second part: **Talk** within a thematic workshop

- The topic of the thematic workshop will be decided in advance (before middle of November)
- Each student has to choose a scientific paper to be presented in 10 minutes
- One thematic workshop will be held at the end of the course (registration needed by early December)
- Other sessions in June and September