Machine Learning for 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	Thursday: 10.30-11-30 (in alternative send an e-mail to get an appointment)

The course

Timetable:

THEORY:	Thursday	16.30 - 18.30 (T.06)
LAB:	Wednesday	11.30 - 13.30 (Alfa)

(note: lab will start on 12th of October)

The course

- **Modality**: **in presence** at the CS department

- **Other facilities**: Lectures will be **recorded**, students not able to attend a given lecture in presence can send me an email asking for the access to the specific recording

The course

Please send me an email asap if you are interested in following the course but you are not allowed to access the moodle space (this is possible if your enrollment 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 at UNIVR)
  - A brief recap will be given at the beginning of the course
- Basic notions of Algorithms, Probability, Statistics, Algebra

Requirements

- Programming skills (for lab part)
  - Programming language used: Matlab

- Note for students who are not familiar with Matlab: the first lab lecture (12th of October) will be devoted to a brief overview of the Matlab language
  - (students who are familiar with Matlab can skip such lecture)

Overview

- Title: “Machine Learning for biological structures and networks”

- The course is about **Machine Learning / Pattern Recognition 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
  - 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 at the e-learning course homepage

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 (during exam sessions):

- First part (15 points): **written exam**: questions/exercises about course topics
- Second part (15 points): **oral exam**: talk within a thematic workshop (as in a conference)

Assessment methods

DETAILS

- The total grade is the sum of the grades of the two parts
- To pass the exam it is mandatory to get al least 9 points in both parts
- The two parts can be taken in different exam sessions
- Each grade part is valid until the end of the academic year (Feb 2024)

Assessment methods

- First part: **written exam**
  - few questions/excercises 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 in advance a scientific paper relevant to the thematic workshop theme
- During the thematic workshop, the student has 10 minutes to present the paper

- More details on the thematic workshop (deadlines, procedure, …) will be given at beginning of November