# Computational Statistics

Sofia Triantafyllou

Spring 2023

E-mail: `sof.triantafillou@uoc.gr`
Web: **https://polyhedron.math.uoc.gr/2223/moodle/course/view.php?id=18**
Class Hours: Monday, Wednesday 09.15-11.00. A212
Office Hours: Thursday, 13.00-15.00

---

This is a tentative syllabus for the course Computational Statistics.

## Course Description

This course will cover fundamental topics of statistical inference using probabilistic graphical models. Probabilistic graphical modeling and inference is a powerful modern approach to representing the combined statistics of data and models, reasoning about the world in the face of uncertainty, and learning about it from data. This course will provide a solid introduction to the methodology and associated techniques. The second half of the course will be student-led, students will select a paper to present (see Section Project).

## Required Materials

The course notes are based on the following books:

1. Bishop, Chris, Pattern Recognition and Machine Learning
2. Murphy, Kevin, Machine Learning: A Probabilistic Perspective

## Prerequisites

Undergraduate students must have passed MEM 262 with grade 7 and above. For graduate students there are no formal prerequisites for this class, but I would advise that you have passed and introductory statistics course. I will only briefly cover basic statistical inference in the class, so you should be familiar.

## Course Syllabus

The course will cover the directed and undirected probabilistic graphical models (PGMs) and inference in PGMs.

| Date | Content | Readings |
|---|---|---|
| Feb 6 | Introduction/Probability Recap | Murphy: Ch. 2, Bishop: Ch. 2 |
| Feb 8 | Frequentist Inference | Murphy: 3.1-3.4 |
| Feb 13 | Bayesian Inference | Murphy: 3.3-3.5 |
| Feb 15 | Testing (In) Dependence | |
| Feb 20 | Directed PGMs | Murphy: 10.1-10.2, Bishop: 8.1- |
| Feb 22 | Undirected PGMs | Murphy: Ch. 19, Bishop: 8.3 |
| Feb 27 | NO CLASS | |
| Feb 29 | Exact Inference: Variable Elimination | Murphy: 20.3, Bishop: 8.4 |
| Mar 6 | Exact Inference: Sum-product algorithm (Project Proposal Due) | Murphy: 20.1-20.2 |
| Mar 8 | Exact Inference: Junction Tree | Murphy: 20.4 |
| Mar 13 | Sampling: Monte Carlo | Murphy: 23.1-23.4, Bishop: 11.1 |
| Mar 15 | Sampling: MCMC | Murphy: 24.1-24.3, Bishop: 11.2 |
| Mar 20 | Mixture Models | Murphy: 11.2-11.3, Bishop: 9.1- |
| Mar 22 | Expectation-Maximization | Murphy: 11.4, Bishop: 9.3 |
| Mar 27 | Structure Learning (HW 1 due) | Murphy: 26.5-26.6 |
| Mar 29 | Structure Learning | Murphy: 26.5-26.6 |
| Apr 3 | Causality Pt 1 | |
| Apr 5 | Causality Pt 2 | |
| Apr 10 | EASTER VACATION | |
| Apr 12 | EASTER VACATION | |
| Apr 17 | EASTER VACATION | |
| Apr 19 | EASTER VACATION | |
| Apr 24 | Variational Inference (HW 2 due) | Murphy: 21.1-21.2 |
| Apr 26 | Variational Inference | Murphy: 21.3 |
| May 1 | NO CLASS | |
| May 3 | Student Presetnations | |
| May 8 | Student Presentations | |
| May 10 | Student Presentations | |
| May 15 | Student Presentations | |
| May 17 | Student Presentations | |

## Grading

- Homeworks: 20%
- Project: 40%
- Final exam: 40

## Homeworks

There will be two hoemworks (due dates in the syllabus). They will have a programming component.

## Project

For your project, you will need to read, present and implement one of the papers in this book: M.I. Jordan (editor), Learning in Graphical Models The deliverables of the project are as follows:

- Project Proposal (due March 6th): One-page proposal including the paper you have picked, a paragraph describing the method in the paper, and the data set you will apply the paper to. (possible sources for data: Kaggle, UCI Machine Learning Repository. In some cases, it is convenient to apply the method to simulated data. In that case, your proposal needs to describe the simulation process.
- Implementation & Report: You will implement and apply it on a data set. You will submit your implementation and a report including your results.
- In-Class Presentation: You give a tutorial covering the paper in a 40-minute presentation (May 3- May 17).

## Learning Goals

The objective of this course is for students to develop a solid understanding of probabilistic graphical models, learn how to apply them to diverse problems . Students are expected to become familiar with the following concepts: Bayesian methodology, conditional independence, model selection, directed graphical models (Bayes nets), undirected graphical models (Markov random fields, factor graphs), exact inference on graphs using message passing, expressing model learning as inference, approximate inference for missing value problems using expectation maximization (EM), variational inference, sampling probability distributions using Markov chain Monte Carlo (MCMC). Specific Topics Include:

1. Creating both directed and undirected graphical models for data.
2. Identifying conditional independencies in graphical models.
3. Specifying distributions for parameters of model components that link the model to data.
4. Applying exact and approximate inference methods to compute marginal probabilities and maximally probable configurations given a model (sum-product and max-sum algorithms, respectively, Monte Carlo sampling methods).
5. Applying approximate inference to learn model parameters using expectation maximization (EM algorithm) and variational inference.