Introduction to the Course

Statistical Machine Learning (ENGG*6600*02)

School of Engineering,
University of Guelph, ON, Canada

Course Instructor: Benyamin Ghojogh
Summer 2023

Introduction of the Instructor and Students

Let us know each other by introducing ourselves!

Feel free to let us know (if you would like):

- Your name
- Your major
- Whether you are studying MEng, MASc, or PhD
- What is your goal for taking this course? What are your expectations from this course?
- How much you know about statistical machine learning?
- How was your vacation between semesters? :)

Introduction of the Course

- This course focuses on statistical machine learning, which is almost most of the machine learning without a deep learning approach.
- We start with preliminaries and background, then the mapping model and tasks in machine learning.
- Then, regression models are introduced.
- Classic classification methods, such as LDA, QDA, SVM, kernel SVM, kernel techniques, Bayes, and KNN are covered.
- Then, spectral and probabilistic feature extraction are introduced, including PCA, FDA, MDS, Isomap, LLE, t-SNE, UMAP, metric learning, and random projection.
- Then, we cover point estimation including MLE and EM algorithms. Then, we cover mixture models.
- Probabilistic graphical models including Markov models, factor graphs, HMM, and MCMC are then explained.
- Generative models (including variational inference, moment matching, adversarial learning, and diffusion models) are the next covered topics.
- We also briefly introduce the concepts of causal inference.
- Ensemble methods such as bagging (e.g., random forest) and boosting are also covered. Clustering algorithms are reviewed.
- Energy-based learning such as Boltzmann machines and Physics-based models, such as Ising models, are covered.
- We will also cover outlier (anomaly) detection including LOF and isolation forest.
- If time allows, we briefly introduce PAC learning.

The Tentative Schedule of Weeks

Week 1

Topics: Preliminaries (probability and expectation, the learning model, overfitting, etc)

Week 2

Topics: Regression (linear, ridge, lasso, and nonlinear regression), Classification (LDA, QDA, SVM, kernel SVM)

Week 3

Topics: Classification (Bayes, KNN, logistic regression), Dimensionality reduction and feature extraction (spectral and probabilistic approaches)

Week 4

Topics: Continue of dimensionality reduction and feature extraction (spectral and probabilistic approaches), Point estimation (MLE, MAE, EM)

Week 5

Topics: Mixture models (Gaussian mixture model, etc), Probabilistic graphical models (HMM, Monte Carlo methods)

Week 6

Topics: Introduction to causal inference, Generative models (variational learning, adversarial learning, GAN, etc)

Week 7

Topics: Ensemble methods (Bagging, Boosting, AdaBoost)

Week 8

Topics: Midterm exam

Week 9

Topics: Energy-based learning (Boltzmann/Gibbs distribution, Ising model, Hopfield network, Boltzmann machines)

Week 10

Topics: Clustering (Centroid-based, Density-based, Distribution-based, Hierarchical, and Spectral Clustering)

Week 11

Topics: Outlier (anomaly) detection (LOF, One-class SVM, Isolation-based methods, geometrical methods, etc)

5.2 Seminar

Week 12

Topics: Group presentations of projects

Course Materials

- Lecture notes will be provided to you.
- YouTube channel of the course: [Link]
- Our tutorial papers: [Link]
- Additional resource for interested students: Prof. Ali Ghodsi's lectures at the University of Waterloo: [Link]
- Additional books:
  - Benyamin Ghojogh, Mark Crowley, Fakhri Karray, Ali Ghodsi, "Elements of Dimensionality Reduction and Manifold Learning", Springer, 2023, [Link]
  - Christopher M. Bishop, "Pattern Recognition and Machine Learning", Springer, 2006 [Link]
  - Trevor Hastie, Robert Tibshirani, Jerome Friedman, "The Elements of Statistical Learning: Data Mining, Inference, and Prediction", Springer, 2009, [Link]
- Research articles in the literature

Course's Websites

Introducing the instructor and TA of the course:

- Instructor: Benyamin Ghojogh, email address: bghojogh@uoguelph.ca
- TA: Hadiseh Moradisani, email address: hmoradis@uoguelph.ca

Course info:

- Classes will be in-person.
- Discussion chats and questions will be in Microsoft Teams group of the course. The students will be added to the Teams group. Please mention (tag) my name and TA's name when you post a question or message in Teams (so it notifies us).
- The course's website is:
  `https://bghojogh.github.io/pages/uoguelph/engg-6600-02-s23/`
- I will probably upload the videos of the classes to my YouTube channel [Click here].
  I will eliminate personal information of students (such as when they introduce themselves) in the videos.

Course Evaluation

- Assignments (20%): Assignments will be posted on CourseLink along with the due dates. They are performed individually. We will probably have several (two to three) assignments.
- The midterm exam (30%): Date will be around week 6. Details to be discussed in class.
- Course project (40%)
  - Date: Week 6 - 11
  - More details will be discussed in class. Report will be electronic submission due in CourseLink.
  - The number of people in each group will be announced in the class.
  - Pick a topic after 6 weeks.
  - Submit the title and proposal/objectives in CourseLink to be checked and approved.
- Group Presentation (10%):
  - Date: Week 11 - 12
  - During class time
- Bonus points: participation in class, participation in the discussions, asking questions, and answering questions.

Course's Goal

- Don't worry much about your marks!
- Focus on understanding the materials of the course.
- Our goal is to learn the important practical and theoretical algorithms in statistical machine learning, so you can use them in both your **industrial projects** and **academic research**.
- About theory and practice:
  - We will learn **some theory** to understand why these methods work.
  - We will also learn how to use the methods **in practice** for practical usage.

Ask Questions!

- Please ask questions whenever you do not understand something.
- Let the class be discussion-based. I do not want to be the sole speaker. We are gonna learn all together.