#### Introduction to the Course

Statistical Machine Learning (ENGG\*6600\*08)

School of Engineering, University of Guelph, ON, Canada

Course Instructor: Benyamin Ghojogh Fall 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, over<u>fitting and cross validation</u> are explained.
- Regression models, including linear regression, ridge regression, and lasso regression, are introduced.
- Classic <u>classification</u> methods, such as <u>LDA</u>, <u>QDA</u>, <u>SVM</u>, <u>kernel SVM</u>, <u>Bayes</u>, <u>KNN</u>, <u>trees</u>, and <u>random forest</u> are covered.
- Bagging and boosting (AdaBoost) are then explained.
- Afterwards, mixture distributions and Gaussian mixture models are covered.
- In the meantime, we cover point estimation including <u>MLE</u> and <u>EM algorithms</u>.
- Then, spectral and probabilistic feature extraction, including PCA, FDA, MDS, Isomap, LLE, variational inference, factor analysis, probabilistic PCA, t-SNE, UMAP, and metric learning, are explained.
- If time allows, other topics can be covered including clustering algorithms, probabilistic graphical models (Markov models, factor graphs, HMM, and MCMC), causal inference, and outlier (anomaly) detection.

# The Tentative Schedule of Weeks

Week 1		Week 7	
Topics:	Preliminaries (probability, random variable, expectation, centering matrix, norm, derivative, Rayleigh-Ritz quotient, eigenvalue and singular value decompositions)	Topics:	Mixture distribution and Gaussian mixture model, Principal component analysis (PCA), dual PCA, kernel PCA, supervised PCA
Week 2		Week 8	
Topics:	Overfitting, cross validation, regularization, Linear discriminant analysis (LDA) and Quadratic discriminant analysis (QDA)	Topics:	Midterm exam
		Week 9	
Week 3 Topics:	Linear regression, ridge regression, Lasso regression	Topics:	Fisher discriminant analysis (FDA), kernel FDA, Multidimensional scaling (MDS), Sammon mapping, Isomap
Week 4		Week 10	
Topics:	Support vector machine (SVM) and kernel SVM	Topics:	Locally linear embedding (LLE), Variational inference, factor analysis, probabilistic PCA
Week 5			
Topics:	K-nearest neighbors (KNN), Bayes and naive Bayes classifiers, logistic regression, tree and random forest	Week 11	
		Topics:	Stochastic neighbor embedding (SNE) and t-SNE, Uniform Manifold Approximation and Projection (UMAP)
Week 6			
Topics:	Boosting and AdaBoost		

# **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 (these books are available at the University's book store):
  - Benyamin Ghojogh, Mark Crowley, Fakhri Karray, Ali Ghodsi, "<u>Elements of</u> 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 of the course:

- Instructor: Benyamin Ghojogh, email address: bghojogh@uoguelph.ca
- The course does not have any TA in this semester :(

Course info:

- Classes will be in-person.
- Discussion chats and questions will be in <u>Microsoft Teams group</u> of the course. The students will be added to the Teams group. Please mention (tag) my 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-6600-08-f23/
- 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: Assignments will be posted on CourseLink along with the due dates. They
  are not mandatory but are for you to practice. They are not graded. I will provide the
  solution keys to you after the due times.
- The midterm exam (50%): 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 <u>industrial projects</u> and <u>academic</u> 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.