STATISTICAL METHODS FOR MACHINE LEARNING II

SYLLABUS: STA 414 / 2104 WINTER 2021

1. Instructors.

- Murat A. Erdogdu
  Email: sta414-2021-prof@cs.toronto.edu
  Office: Online
  Office Hours: Wednesday 10:00-12:00

2. Lectures. This course has two identical sections:

- 414 L0101 & 2104 L9101: Monday 14:00-17:00
- 414 L5101 & 2104 L6101: Tuesday 18:00-21:00

3. Teaching Assistants.

Email: sta414-2021-tas@cs.toronto.edu

- An, Ruijian
- Li, Mufan
- Yu, Lu

4. Course webpages. The course webpage contains all course information, additional readings, assignments, announcements, office hours, etc. You are expected to check the following sites regularly!

- erdogdu.github.io/sta414
- q.utoronto.ca
- piazza.com/utoronto.ca/winter2021/sta414

Instructions to join online lectures and office hours will be sent through quercus every week. Lectures will be recorded and posted on quercus for your convenience. **Please do not distribute!**

5. Course Evaluation.

- 4 assignments: 50%, due on 2/08, 2/22, 3/22, 4/05
- Midterm exam: 20% in class on 3/01 for Monday section, 3/02 for Tuesday section.
- Final exam: 30%

6. Course Outline. This course covers several commonly used machine learning algorithms and related methodological concepts. Topics may include:

1. Introduction to ML & Least Squares
2. Probabilistic Models
3. Regularization and Bayesian Methods
4. Linear Methods for Classification
5. Decision Trees, Ensembles: Bagging & Boosting, Support Vector Machines

6. Optimization in ML
7. Decision Theory
8. Neural Networks & Backpropagation
9. Unsupervised learning, Latent variable models, Autoencoders
10. Matrix Factorizations & Recommender Systems
11. Sampling Methods
12. Gaussian Processes
13. Reinforcement Learning
14. Algorithmic Fairness

7. Prerequisites. STA314H1/CSC411H1/CSC311H1/(STA314H5, STA315H5)/CSCC11H3/CSC411H5; STA302H1/STAC67H3/STA302H5; CSC108H1/CSC120H1/CSC148H1/CSCA08H3/CSCA48H3/CSCA20H3 /CSC108H5/CSC148H5; MAT235Y1/MAT237Y1/MAT257Y1/(MATB41H3, MATB42H3)/(MAT232H5, MAT236H5)/(MAT233H5, MAT236H5); MAT223H1/MAT240H1/MATA23H3/MAT223H5/MAT240H5 Exclusion: CSC412H1

8. Textbooks. There is no required course textbook. The following materials can be helpful.

- Christopher M. Bishop (2006). Pattern Recognition and Machine Learning
- Ian Goodfellow, Yoshua Bengio and Aaron Courville (2016), Deep Learning
- Kevin Murphy (2012). Machine Learning: A Probabilistic Perspective
- Trevor Hastie, Robert Tibshirani, Jerome Friedman (2009). The Elements of Statistical Learning
- Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani (2017). Introduction to Statistical Learning
- David MacKay (2003). Information Theory, Inference, and Learning Algorithms

9. Assignments. There will be 4 assignments in this course. The assignments will be released on the course webpage. Submission instructions will be provided with each assignment.

9.1. *Collaboration policy.* Collaboration on the assignments is not allowed. Each student is responsible for his/her own work. Discussion of assignments should be limited to clarification of the handout itself, and should not involve any sharing of pseudocode or code or simulation results. Violation of this policy is grounds for a semester grade of F, in accordance with university regulations.

10. Exams. There will be a midterm exam (date TBD, most likely on the week of Mar 1). Details will be announced in class and on the course webpage. You can use an optional A4 cheat sheet - double-sided. Final exam date is TBD. You can use two optional A4 cheat sheets - double-sided. You will need to scan and upload your solutions to a submission platform, instructions of which will be given to you.

11. Late policy. If you are traveling, you may email your submission to one of the course staff in advance of the deadline. Ten percent of the value will be deducted for each late day (up to 3 days, then submission is blocked). No credit will be given for assignments submitted after 3 days. Extensions will be granted only in special situations, and you will need a Student Medical Certificate or a written request approved by the course coordinator at least one week before the due date.

12. Grading concerns. Any requests to have graded work re-evaluated must be made within one week of the date the grade is released. Re-evaluation may result in a decrease in the grade.

13. Computing. In the assignments and project, you may need to write your own programs, debug them, and use them to conduct various experiments, plot curves, etc. You may use any programming language, but `Python` is preferable. For some of the assignments, we will provide you a starter code in `Python` only.

14. Missed Tests.

- If a test is missed for a valid reason, you must submit documentation to the course instructor.
- If a test is missed for a valid medical reason, you must submit the University of Toronto Verification of Student Illness or Injury form to your instructor within one week of the test.
- The form will only be accepted as valid if the form is filled out according to the instructions on the form.
- Important: The form must indicate that the degree of incapacitation on academic functioning is moderate, serious, or severe in order to be considered a valid medical reason for missing the term test. If the form indicates that the degree of incapacitation on academic functioning is negligible or mild then this will not be considered a valid medical reason.
- If the midterm test is missed for a valid reason then the final test will be worth 50% of your final grade. Other reasons for missing a test will require prior approval by your instructor. If prior approval is not received for non-medical reasons then you will receive a term test grade of zero.

15. Accommodation for Disability Policy. Please send your documented accessibility requests directly to the instructor, at least a week before the due date of each evaluation item. Extensions may be granted, and the duration will be determined based on the letter from the Accessibility Services at the University of Toronto.