MATH 7339-Machine Learning and Statistical Learning Theory 2 – Fall 2024

Instructor: He Wang
Class time and room: see Canvas
Office Hours: see Canvas
Email: he.wang@northeastern.edu

Recommended textbooks (with links):

1. The Elements of Statistical Learning: Data Mining, Inference, and Prediction by Trevor Hastie, Robert Tibshirani, Jerome Friedman.
2. Machine Learning: A Probabilistic Perspective, Kevin Murphy, MIT Press.
3. Probabilistic Machine Learning: An Introduction , Kevin Murphy, MIT Press.
4. Probabilistic Machine Learning: Advanced Topics , Kevin Murphy, MIT Press.
5. Pattern Recognition and Machine Learning, Christopher M. Bishop, Springer.
6. More textbook (for time series and NLP) chapters or paper references will be provided on Canvas. Complementary lecture notes will also be provided.

Prerequisite:
Knowledge of machine learning at the level of Math 7243, or CS 6140, or EECE 5644, or DS 5220. Knowledge about linear algebra, multivariable calculus, probability and statistics are required. Programming skill in Python or MATLAB or R is required.

Overview:
Continues Math 7243. The course further covers theory and methods for regression and classification, along with other advanced topics in machine learning, statistics and deep learning. Starts with reviewing basics of machine learning in a broader and deeper way. Further topics will be drawn from smoothing methods, clustering, latent variable models, mixture models and EM, Markov decision processes and reinforcement learning, neural networks and deep learning. We will also discuss current research topics from image segmentation, generative adversarial network, neural style transfer, natural language processing and topological data analysis. The course includes student presentations and a semester project.

Course Format:
About three-fourths of the classes will involve traditionally formatted lectures. For the other one-fourth of the classes, we will read and discuss two seminal papers relevant to the course topic. These classes will involve presentations by groups of students of the paper contents followed by breakout discussions about the material. Topics will be discussed in the context of research reported in the literature in recent years. Students will participate in class discussions and will present the results of a semester-long research project of their own choosing.

Student Learning Outcome:
At the end of the course, students will achieve the following goals and objectives:

1. Student will have a solid understanding of the theory and methods of advanced machine learning models and frameworks, such as kernel methods, latent variable models, neural network and backpropagation, Markov decision processes.
2. Students will understand and be able to explain more advanced regression and classification machine-learning and deep learning concepts, algorithms and methods, such as deep neural network, generative adversarial network and neural style transfer, reinforcement learning.
3. By working on computer labs and semester project, students will be able to apply the methods and algorithms of machine learning and deep learning to real world probems.
4. Along with the study of the advanced topics, students will be able to read, evaluate and review new research papers.

Course topics: The courses will further cover nonparametric and parametric learning; supervised, unsupervised, and semi-supervised learning; graphical models; ensemble methods; and reinforcement learning. More specifically, we will cover the following topics:

I. **Theory and methods (About 8 weeks)**

1. Bias-variance trade-off. Review linear methods and exponential families.
2. Model Complexity, Hyperparameters and tradeoffs (VC dimension, degree of freedom, hyperparameters optimization)
3. Bayesian Methods, Bayesian inference/ Bayesian Statistics/ methods
4. Basis expansions and regularization
5. Kernel smoothing methods
6. Latent variable models (PCA, FA, ICA, Gaussian Mixture Models,EM Algorithm)
7. Kernel methods

II. **Further topics.**

8. **Time Series (about 3 weeks)**
   1) Time Series and Forecasting
   2) Overview, intro and examples,
   3) AR ; MA ; ARIMA
   4) RNN for time series data
   5) FFT
   6) Facebook: Forecasting at scale
9. **Natural language processing (depending on time, case study) (about 1 week)**
   1) Vectorization (Word Embeddings)
   2) RNN,NN for sequences
   3) Transformer
   4) Bidirectional Encoder Representations from Transformers (BERT)

III. **Applications**

The machine learning is almost defined by its application. We will have frequent in class labs wherein we implement the algorithms we have discussed in the theory portion on real world data sets. We will encounter applications in image processing, natural language processing, finance, medical data, etc.

Grading: Students will be evaluated on the following basis.

1. **Homework.** (20%) There will be a few homework questions on the theory of machine learning
2. **Computer labs (**20%**)** (There will be a few computer labs focusing on the implementation of algorithms on real world data sets.)
3. **Attendance and Class discussion participation (**15%)
4. **Paper review/paper presentation (**15 %**)** For the paper presentation, (1) summarize the paper, (2) discuss the paper's strengths and weaknesses, and (3) discuss the paper's impact.
5. **Final project**. (30%) The final project be a computational analysis of a data set using sufficiently complicated or novel techniques from this course. It consists of a proposal, middle stage progress report, project report and presentation (with poster or slides). Project groups should contain 2-5 people.

Software: Python will be used throughout the course. Students should be prepared to use Python, (TensorFlow, Pytorch) for lab assignments and for the final project. Familiar with any one computer language is required, e.g., Python or Mathlab or R.

Late submission policy. Late submission within a week of any assignment without permission will receive at most 90% of the grade. Late submission within a week but after the posting of the solution will receive at most 70% of the grade. Other late submissions will depend on instructor's discretion. If you request an extension, you should contact the instructor at least one day before the due day.

Collaboration: You are welcome, even encouraged, to collaborate on the homework and lab assignments, though we urge you to first attempt working out all of the problems by yourself. However, you are expected to write answers **yourself** and understand everything that you hand in. **Copy** results from any sources will be considered as violating Academic Integrity. Collaboration is not allowed on the quizzes and exams.

Classroom disruptions policy: Each Northeastern student has the responsibility to respect the other students in classroom, including using electronic devices. (For example, messaging, game playing, watching videos, and internet surfing are not allowed during class time. However, you can use devices taking notes or quick search some class related information.)

Academic Integrity Policy: Cheating will not be tolerated. All incidents of cheating will be reported. From the Academic Integrity Policy: (see http://www.northeastern.edu/osccr/academic-integrity-policy/)

> "A commitment to the principles of academic integrity is essential to the mission of Northeastern University. The promotion of independent and original scholarship ensures that students derive the most from their educational experience and their pursuit of knowledge. Academic dishonesty violates the most fundamental values of an intellectual community and undermines the achievements of the entire University.

> As members of the academic community, students must become familiar with their rights and responsibilities. In each course, they are responsible for knowing the requirements and restrictions regarding research and writing, examinations of whatever kind, collaborative work, the use of study aids, the appropriateness of assistance, and other issues."

Title IX Policy: The University strictly prohibits sex or gender discrimination in all university programs and activities. Information on how to report an incident of such discrimination (which includes sexual harassment and sexual assault) is located at http://www.northeastern.edu/titleix .

Inclusion and Diversity: I value all students regardless of their background, country of origin, race, religion, gender, sexual orientation, ethnicity, or disability status, and am committed to providing a climate of excellence and inclusiveness within all aspects of the course. If there are aspects of your culture or identity that you would like to share with me as they relate to your success in this class, I would be happy to meet to discuss. Also, if you have any concerns in this area or are facing any special issues or challenges, I encourage you to discuss the matter with me as you feel comfortable, with assurance of full confidentiality (only exception being mandatory reporting of NU Academic Integrity Policy violations and Title IX sex and gender discrimination).

Students with disabilities: Students who have disabilities who wish to receive academic services and accommodations should follow the standard Disabilities Resource Center (DRC) procedures (see http://www.northeastern.edu/drc/getting-started-with-the-drc/).
College of Science Policies: The current College of Science Academic Course Policies is available at https://cos.northeastern.edu/wp-content/uploads/2012/10/COS-teaching-policies-April-2017.pdf .

TRACE: Every student is expected to complete the online TRACE survey at the end of the semester.