COMS4995W32 Applied Machine Learning (Fall 25)

Course Description

Applied Machine Learning is essential as it enables systems to learn from data and make reliable decisions at scale across industries. This course begins by walking-through the probabilistic foundations of machine learning - covering topics such as random variables, maximum likelihood estimation, Bayesian inference, and MAP priors. From there, the emphasis shifts to the end-to-end predictive modeling pipeline, including deployment considerations and production challenges, such as scalability, monitoring, and data drift. The curriculum then steps into the realms of neural networks, transformer-based architectures, and state-of-the-art LLM training. The final module explores agentic workflows, where models are chained with tools to autonomously tackle complicated tasks in the real world.

Learning Objectives

- **[Problem Formulation]** choosing the right question, business value and solutions
- **[Mathematics of Data Science]** statistical learning, optimization, algorithms
- **[Machine Learning Models]** supervised and unsupervised learning (classification, regression, clustering, neural networks, Transformers)
- **[Feature Analysis]** data visualization, feature engineering, feature selection
- **[Modeling Process]** training, validating, testing, evaluation metrics, LLM tuning
- **[MLops]** deployment, versioning, monitoring and maintenance
- **[Tools]** Python, scikit-learn, Google Colab, Visual Studio Code

Prerequisites

- Programming proficiency in Python3, including numpy, pandas, pytorch, and comfort with Google Colab or Visual Studio Code
- Solid grasp of data-structures and basic algorithms; students shall be able to implement and debug code, with the help of AI assistants
- Basic Linear Algebra knowledge (vectors, matrices, eigenvalues/eigenvectors, SVD, positive-definite matrices)
- Probability theory and mathematical Statistics (random variables, expectation, variance, Bayes Rule, MLE/MAP)
- Experience with exploratory data analysis and basic data-visualization techniques
- [Optional] Prior exposure to introductory Machine Learning or Data Science coursework

Course Logistics

- Classroom: Pupin Hall 301 (Morningside Campus)
- Schedule: Thursdays, 7:00 pm - 9:30 pm (Fall 2025)
- Credits: 3.0

Instructor

- Dr. Spencer W. Luo (Google)
- Email: swl2145@

Teaching Assistants

- Case Hallowell Schemmer (chs2164@)
- Mihika Riya Sanghvi (mrs2356@)
- Hamsitha Challagundla (hc3540@)
- Shashwat Kumar (sk5520@)
- Jiayi Niu (jn2941@)
- Tony Tian (jt3640@)

TA Office Hours: TBD, updated on Courseworks
Optional Recitation: TA-led sessions for coding and math refreshers.

Course Schedule

Class	Lectures	Assessment
1	Introduction to Applied Machine Learning • ML workflow • ML in production Tools: Colab, scikit-learn, PyTorch [Reading] PML ch1	
2	Data Preparations and Feature Engineering • Data cleaning, visualization, pipelines [Reading] DL ch 2	
3	Generative vs. Discriminative Approaches: Naive Bayes vs. Linear Regression • Applications in price prediction, CTR [Reading] ESL ch 4	Assignment 1 out
4	Model Evaluation & Bias-Variance • Metrics: Accuracy, AUC, RMSE, cross-validation • Case study: Sentiment analysis baseline [Reading] PML ch 7	
5	Tree-based Models & Ensembles • Decision Trees, Bagging, Boosting and Ensemble Methods [Reading] ESL ch 8	Assignment 2 out
6	Unsupervised Learning • Clustering (k-means, DBSCAN), PCA, t-SNE • Self-supervised learning [Reading] PML ch 14	Midterm Exam
7	Neural Networks Fundamentals • MLP, activation, backprop • PyTorch MNIST hands-on [Reading] DL ch 5	
8	Convolutional Neural Networks • Convolution, pooling, ResNet • CIFAR-10 hands-on [Reading] DL ch 9	

9	Transformers ● Attention is All You Need [Reading] DL ch 10	Assignment 3 out
10	Transformers in Practice ● BERT, GPT, Hugging Face pipelines ● Text classification, QA [Reading] DL ch 12	Final project out
11	Pre-training and Supervised Fine-tuning ● Understanding pre-training paradigms and the role of SFT [Reading] DL ch 18-19	
12	Reinforcement Learning ● Exploration and exploitation ● RL in LLMs [Reading] RL ch 1-4	
13	Agentic Workflow ● LangChain, tool integration ● Go beyond the LLM blackbox [Reading] Google Agentspace	

Recommended Textbooks

- The Elements of Statistical Learning (ESL) by Trevor Hastie et al.
- Deep Learning (DL) by Ian Goodfellow et al.
- Reinforcement Learning: An Introduction (RL) by Richard Sutton et al.
- Probabilistic Machine Learning: An Introduction (PML) by Kevin Murphy et al.

Assessments

Students will learn twice - once when they solve a problem themselves and again when they share it to peers. As a result, this course requires the following activities:

- Class participation
  - Enrich discussions with personal insight, critical questions, and constructive critiques
- **1 Mid-term test**
  - Closed-book, single-sheet cheat sheet allowed
  - Mix of multiple choice and short-answer derivations covering fundamental concepts.
  - Emphasis on understanding key concepts, not memorizing formulas
- **3 Homework assignment**
  - Each student will work on a concrete ML problem and come up with an end-to-end solution, covering feature engineering, advanced models and hyper-param sweeping
  - The deliverables: 1) well-documented code repo, 2) N-page experimental report (formulation -> modeling -> result -> lesson)
- **1 Final LLM project**
  - Build up a team around 5 team members, and work on a practical LLM training project
  - The deliverables for each: 1) presentation slides, 2) 10-min project video presented by all team members, 3) detailed experimental report and codes
  - The instructor and the team will share guidelines of accessing GPU resources

Assessment	Final Grade %
Class participation	5%
Mid-term test	30%
Homework assignment	35%
Final Project presentation	30%

Grade	Percentage Interval
A	90-100%
B	80-89%
C	70-79%
D	60-69%
R	59% or below

Course Policies

Attendance & Participation: Class participation will be graded by in-class engagement, including asking relevant questions based on a critical review of required readings, on lectures, and on comments made by your peers. The lack of attendance will count against your participation grade

Late-work policy: Late work will be accepted with the following conditions:
- Work submitted 24 hours or less past the original due date & time will be accepted with no penalty
- Work submitted more than 24 hours past the original due date & time will be accepted with a 10% penalty for up to 3 days (72 hours) past the original due date & time
- Work will **not** be accepted more than 3 days (72 hours) past the original due date & time

Laptops & Mobile Devices: Students are permitted to use laptops and mobile devices during class sessions, such as for note-taking purposes. Please ensure audio sounds are muted so as not to disturb the class

Re-grade policy: Students may request a re-grade within one week from the date the graded assignment is returned to students. If requesting a regrade, students should be prepared to make a compelling argument as to why the original grade needs to be reconsidered. *The student should discuss the grade with the TA first before requesting instructor re-grade.* The process for requesting an instructor re-grade is to email the instructor with the request within one week from the date the graded assignment is returned. The instructor and student(s) will then decide on a mutually agreeable time and place to discuss the assignment

Use of Generative AI: AI is everywhere now, and thus assignments/projects in this course will *permit or even encourage* the use of AI tools, such as ChatGPT and Gemini. In the assignments/projects, AI use must be appropriately acknowledged and cited. For instance, if you generated a whole paragraph through ChatGPT and edited it for accuracy, your submitted work would need to include a note such as "I generated this work through Chat GPT and edited the content for accuracy". **It is each student's responsibility to assess the validity and applicability of any AI output that is submitted.** Please email the instructor if you have questions regarding what is permissible and not for a particular assignment/project.