ECE 5984 SPECIAL STUDY: APPLICATIONS OF MACHINE LEARNING

Spring 2022 ♦ 3 credits

T Th 6:30 PM – 7:45 PM ♦ Durham 261, and Zoom (ID 857 0803 5647, passcode **hokies**!)

Instructor

Dr. Creed Jones Phone: 540-231-9850 Office: 462 Whittemore Hall
E-mail: crjones4@vt.edu (Please, put "5984" in the subject line of your email.)
Office hours: M W 1:30 PM – 3:30 PM, via Zoom, ID# 848 5781 1910
T Th 4:00 PM – 5:30 PM, via Zoom, ID# 813 4648 4450, from 462 Whittemore Hall
All course files, lecture videos, assignments and announcements are hosted in Canvas.

Description

Introduction to Machine Learning (ML) for predictive data analytics. Probability for ML including conditional probability, the product and chain rule, and the Theorem of Total Probability. Data preparation for ML algorithms, normalization, cleaning, and imputation of missing values. Information-based learning using decision trees. Similarity-based methods, data classification and clustering. Probability-based learning, conditional probability and Bayes' theorem, and applications. Linear and logistic regression and optimization-based learning. Performance evaluation of ML systems. Real-world applications of ML and case studies. Pre: Graduate Standing. (3H, 3C).

Formal learning objectives

Having successfully completed this course, the student will be able to

1. Apply standard Machine Learning (ML) approaches in real-world scenarios using software tools for predictive data analysis.
2. Prepare raw data sets for use by ML algorithms and software using appropriate techniques.
3. Formulate decision-tree solutions in information-based learning applications.
4. Perform data classification and clustering for ML applications using similarity metrics.
5. Compute probability-based solutions for inference and prediction using Bayes' theorem.
6. Apply optimization-based learning and regression techniques to engineering applications.
7. Evaluate ML approaches and systems using standard performance measures for specific case studies.

Prerequisites

Understanding of calculus and linear algebra. Ability to use a programming language.

Required text

Kelleher, Mac Namee and D'Arcy, *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies*, 2nd edition, MIT Press, 2020, ISBN 978-0-2620-4469-1.

Required software

We will be using two software environments to explore applications of machine learning: Tableau for data exploration and Python (with NumPy and Scikit-learn) for more advanced modeling. Both of these environments are freely available to you. More information will be coming on these packages.

This Course and Graduate Plans of Study

This course is currently being offered as a special study course (ECE 5984), but it is in the final stages of approval as a permanent course. It will be numbered ECE 5464. This course will not count as a special study course on your plan of study.

Note that this course is only approved to appear on plans of study for the MEng program; it cannot be used on the plan of study for MS and PhD students.

Assessment

Homework Assignments

There will be six homework assignments during the course; these are individual assignments! Students may discuss among themselves general approaches to solving homework problems. The final solutions are expected to be the original work of each individual student. Unless I specify otherwise in the assignment, you are welcome to use relevant snippets of code from online sources **as long as you clearly indicate which lines of code you have borrowed and give me the source!** Code from other students or acquaintances may not be used! If you have any questions, please ask me. Failure to clearly cite a source will be treated as an Honor Code violation.

All assignments should be submitted in electronic form via Canvas (check the assignment files for more detail). Homework assignments **only** will be accepted up to three days late and will be penalized 10% per day.

Exams

We will have a final exam on Saturday, May 7. You MUST be available at this time; make sure and clear your calendar now. More information will be provided prior to the exam.

Quizzes

We will have eight quizzes during the course. Each will be fully online, in Canvas, and will consist of around ten questions. You will have twenty minutes to complete the quiz, and it must be taken within the time window assigned (typically six hours). No late quizzes will be administered.

Projects

This course features a pair of implementation projects. These projects are more extensive and open-ended than a homework assignment. You will work in teams of three to four to implement your projects. You are welcome to choose your own team before February 1; if you have not chosen a team by that date, I will assign you to a team.

Technical Paper Review

You will be asked to select a suitable journal or conference paper in the field of machine learning and to write a review of it. More information will be provided on this assignment.

Grades are calculated per the following:

Graded Item	# of Items	Points per Item	Total Points	Percentage
Homework Assignments	6	25	150	30%
Projects	2	60	120	24%
Technical Paper Review	1	50	50	10%
Final Exam	1	100	100	20%
Quizzes	8	10	80	16%
			500	100%

I use the following mapping from percentage scores to letter grades:

Letter Grade		Range
A	100 %	to 94.0%
A-	< 94.0 %	to 90.0%
B+	< 90.0 %	to 87.0%
B	< 87.0 %	to 84.0%
B-	< 84.0 %	to 80.0%
C+	< 80.0 %	to 77.0%
C	< 77.0 %	to 74.0%
C-	< 74.0 %	to 70.0%
D+	< 70.0 %	to 67.0%
D	< 67.0 %	to 64.0%
D-	< 64.0 %	to 60.0%
F		below 60%

Course delivery

For those of you in Blacksburg, this course will be delivered in person - 6:30 to 7:45 PM on Tuesdays and Thursdays in room 261 of Durham Hall. Students in remote sections will attend via zoom. I invite and will reply to questions during lecture; if you are a remote student, use the chat feature of zoom to ask your question.

All lectures will be recorded and available for viewing later, but I strongly encourage you to participate in the lectures at the scheduled time! You are accountable for all material, regardless of any technical issues that may occur in recording and posting videos of the lectures.

General policies

Students may discuss among themselves general approaches to solving homework problems. The final solutions are expected to be the original work of each individual student. Course projects may be done independently or with a partner. All assignments should be submitted in electronic form via Canvas (check the assignment files for more detail). Only homework assignments will be accepted up to three days late and will be penalized 10% per day. No other work may be submitted late.

Office hours

To support the needs of students in all locations, I will be having office hours during which I will be available by Zoom, so that you can drop in and ask questions. These office hours will occur on Mondays and Wednesdays from 1:30 PM to 3:30 PM, and Thursdays from 4 PM until 5:30 PM. On Tuesdays and Thursdays I will be having my office hours from my office in 462 Whittemore, so you are welcome to drop by in person if you prefer.

M W 1:30 PM – 3:30 PM, via Zoom, ID# 848 5781 1910

T Th 4:00 PM – 5:30 PM, via Zoom, ID# 813 4648 4450, from 462 Whittemore Hall

I also use the Piazza forum tool in Canvas, accessible from the course site in Canvas. You are also welcome to contact me via email with questions; be sure to use your VT email account, and put "5984" somewhere in the subject line!

So, your options for asking questions are:

- During lecture (remote students use the Chat feature of Zoom)
- Piazza in Canvas (public question, anonymous is fine)
- Send me an email (put 5984 in the subject line!)
- Drop in during office hours sessions, either by Zoom or in person

Honor code

Compliance with the Graduate Honor Code requires that all graduate students exercise honesty and ethical behavior in all their academic pursuits at the university. Students enrolled in this course are responsible for abiding by the Honor Code. A student who has doubts about how the Honor Code applies to any assignment is responsible for obtaining specific guidance from the course instructor before submitting the assignment for evaluation. Ignorance of the rules does not exclude any member of the University community from the requirements and expectations of the Honor Code. Academic integrity expectations are the same for online classes as they are for in-person classes.

For additional information about the Graduate Honor System, please visit the following page: https://graduateschool.vt.edu/academics/expectations/graduate-honor-system.html .

Disabilities

Virginia Tech welcomes students with disabilities into the University's educational programs. The University promotes efforts to provide equal access and a culture of inclusion without altering the essential elements of coursework. If you anticipate or experience academic barriers that may be due to disability, including but not limited to ADHD, chronic or temporary medical conditions, deaf or hard of hearing, learning disability, mental health, or vision impairment, please contact the Services for Students with Disabilities (SSD) office (540-231-3788, ssd@vt.edu, or visit www.ssd.vt.edu). If you have an SSD accommodation letter, please meet with me privately during office hours as early in the semester as possible to deliver your letter and discuss your accommodations. You must give me reasonable notice to implement your accommodations, which is generally 5 business days and 10 business days for final exams.

Day	Module	Lec	Advance Reading	ECE5984 SP22 Daily Schedule Topics	Due
18-Jan	I - Foundations	1	1.1 - 1.7	Course introduction	
20-Jan		2	App. D	Review of linear algebra	
25-Jan		3	App. B & A	Review of statistics	
27-Jan	II - Data Prep	4	3.1 - 3.5	Data exploration	quiz 1
1-Feb		5		More on data exploration and presentation	
3-Feb		6		Python and sklearn	
8-Feb		7	3.4	Missing values	hw 1
10-Feb		8	3.6	Data preparation	quiz 2
15-Feb	III - Information and Similarity	9	4.1 - 4.3	Introduction to decision trees	hw 2
17-Feb		10	4.4 - 4.5	More on decision trees	
22-Feb		11	5.1 - 5.23	Similarity measures	
24-Feb		12	5.4	Classification	quiz 3
1-Mar	IV - Probability	13	6.1 - 6.2	Probability-based learning; Bayes' theorem	hw 3
3-Mar		14	6.3 - 6.4	Bayesian prediction	quiz 4
8-Mar	*No Class - Spring Break*				
10-Mar					
15-Mar	V - Gradient-based methods	15	7.1 - 7.2	Gradient-based methods	
17-Mar		16	7.3 - 7.4	Multivariate linear regression	
22-Mar		17	5.4.6	Variable selection	prj 1
24-Mar		18	7.4	Logistic regression	quiz 5
29-Mar	VI - Performance	19	9.1 - 9.3	Performance evaluation; misclassification	
31-Mar		20	9.4	ROC curves; other performance measures	
5-Apr		21	4.4.5	Model selection / ensemble models	hw 4
7-Apr	VII - Neural networks	22	8.1 - 8.3	Neural networks	quiz 6
12-Apr		23	8.4	More on neural networks	
14-Apr		24		Deep learning	
19-Apr		25	7.4.7	Support vector machines	hw 5
21-Apr	VIII - Other methods	26		Other modeling techniques	quiz 7
26-Apr		27	10.1 - 10.5	Unsupervised learning – a preview	
28-Apr		28	11.1 - 11.5	Reinforcement learning – a preview	hw 6
3-May		29		Course review	quiz 8 / prj 2
5-May	*No Class - Reading Day*				
7-May	**FINAL EXAM (7:00 to 9:00 PM)**				