THE UNIVERSITY OF TENNESSEE KNOXVILLE

The University of Tennessee
College of Engineering
Applied Data Science
Spring 2022

Course Information

Course Number: IE 465/565
Credit hours: 3
Time: 1:10 pm-2:25 pm – Tuesdays/Thursdays
Place: Online via Zoom: https://tennessee.zoom.us/j/91923762183
Prerequisites: An introductory course(s) in probability and statistics and Python (numpy, pandas, matplotlib, functions, classes)
Videos: Later available on YouTube/Canvas

Instructor

Yaoping Wang, Ph.D.
Research Assistant Professor
Institute for a Secure and Sustainable Environment
600 Henley St Ste 311
Knoxville, TN 37902

Office hours

Online via Zoom, time TBD

Course Description

An introduction to applied data science using machine learning tools. The class will mainly cover the analysis of data using supervised learning, the evaluation and improvement of model performance, and the interpretation of the results. Some unsupervised learning methods will be introduced toward the end. The theory of each method will be described and followed by the implementations in Python using the sklearn, xgboost, or pytorch packages. The course is project-based and emphasis will be put on students using machine learning to tackle real-world questions of interest to them. All course materials and other relevant information will be posted on Canvas.

Course Resources

Software

- Python programming language: Open-source software for computing and graphics. Download and install Python from here: https://www.python.org/
- Jupyter Software: https://jupyter.org/install. I recommend using Anaconda for installation: https://docs.anaconda.com/anaconda/install/. Alternatively, use Google Colab: https://colab.research.google.com/

Extension reading materials (not required)

- Sebastian Raschka & Vahid Mirjalili, Python Machine Learning – Machine learning and deep learning with Python, scikit-learn, and Tensorflow (2nd edition), 2017
- Applied Machine Learning Cornell Tech CS 5787 https://www.youtube.com/playlist?list=PL2UML_KCiC0UlY7iCQDSiGDMovaupqc83
- Christopher M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006
- Numerous other tutorials/videos/books that exist online …

Value Proposition

Ability to analyze large datasets to understand the hidden patterns provides an opportunity to improve/augment decision making processes in various engineering fields and business operations. This course will address the topics that are essential for developing these skills.

Student Learning Outcomes

- Ability to judge when it is appropriate to use machine learning
- Knowledge of some basic classification, regression, and clustering algorithms
- Ability to evaluate the performance of a machine learning model
- Ability to apply basic machine learning methods and to tweak them to improve performance
- Ability to identify hidden patterns in data and prescribe actions for unseen data based on machine learning results

Grading

Homework and quizzes – 20%

Project presentations – 60% total, with 30% from peer-review and 30% from instructor

Final project report – 20%

(The instructor may adjust these percentages somewhat based on the class performance)

Course policies

- All assignments and quizzes must be submitted through Canvas. Homework will be graded based on the criteria expected of professional engineering submissions including mathematical correctness, organization, readability, correct grammar, and spelling

- Grades will be posted on Canvas and updated periodically. It is each student's responsibility to verify that their grades have been entered correctly. Students have two weeks to discuss any concerns in their grades after a graded assignment or exam is returned
- Regular class attendance (via zoom) is expected. If a class is missed due to an unforeseen event, it is the absentee's responsibility to get all missing notes or materials
- No late assignments will be accepted
- Students may work together on projects and homework assignments; however, offering and accepting solutions from others is an act of plagiarism, which is a serious offense and all involved parties will be penalized according to the Academic Honesty Policy
- Students are expected to work independently on quizzes. All questions regarding quizzes need to be directed to the instructor
- UTK Default Schema will be used for grading:

Grades scored between				will equal
97	% and	100%		A+
94	% and	less than	97%	A
90	% and	less than	94%	A-
87	% and	less than	90%	B+
84	% and	less than	87%	B
80	% and	less than	84%	B-
77	% and	less than	80%	C+
74	% and	less than	77%	C
70	% and	less than	74%	C-
67	% and	less than	70%	D+
64	% and	less than	67%	D
60	% and	less than	64%	D-
0	% and	less than	60%	F

Academic Honesty Policy

Students are expected to follow the honor system as stated in the student handbook.
"An essential feature of The University of Tennessee, Knoxville is a commitment to maintaining an atmosphere of intellectual integrity and academic honesty. As a student of the university, I pledge that I will neither knowingly give nor receive any inappropriate assistance in academic work, thus affirming my own personal commitment to honor and integrity."

Student Disability Services.

Any student who feels they may need an accommodation based on the impact of a disability should contact Student Disability Services in Dunford Hall, at 865-974-6087, or by video relay at, 865-622-6566, to coordinate reasonable academic accommodations.

Tentative Course Outline

The weekly coverage and topics might change as they depend on the progress of the class

Week	Date	Topic
1	1/24-1/28	Linear regression, logistic regression, artificial neural network
2	1/31-1/4	Decision trees, bagging and boosting
3	2/7-2/11	Machine learning workflow, performance evaluation, and hyperparameter tuning via examples
4	2/14-2/18	
5	2/21-2/25	First round of presentations (problem definition, descriptive statistics, potential methods)
6	2/28-3/4	Recursive neural network
7	3/8-3/11	Convolutional neural network
8	Spring break	
9	3/21-3/25	Second round of presentations (lit review, methods, evaluation and benchmarking)
10	3/28-4/1	One-on-one project discussions
11	4/4-4/8	
12	4/12	Unsupervised learning methods
13	4/18-4/22	
14	4/25-4/29	Interpretable machine learning
15	5/2-5/6	Final presentations and reports
16	5/10	No class