CSCE 50603: Machine Learning (3 credit hours)

Catalog Description: Introduction to supervised learning (KNN, regression, SVM, neural networks), unsupervised learning (k-means, EM), graphical models (causal graph), learning theorems (PAC, VC dimension), software and packages, and advanced topics (deep learning, reinforcement learning, and causal modeling).

Prerequisites: CSCE graduate standing.

Textbook: No textbook is required. The following books are good references.

- The Elements of Statistical Learning, by Trevor Hastie, et. al. (2009). Available online: https://web.stanford.edu/~hastie/ElemStatLearn/
- Machine Learning: A Probabilistic Perspective, by Kevin Murphy (2012)
- Understanding Machine Learning: From Theory to Algorithms, by Shai Shalev-Shwartz and Shai Ben-David (2014). Available online: https://www.cse.huji.ac.il/~shais/UnderstandingMachineLearning/
- Dive into Deep Learning, by Aston Zhang and Zachary C. Lipton and Mu Li and Alexander J. Smola (2020). Available online: https://d2l.ai/

Goals: This course will teach the theoretical foundation of various machine learning models, their programming implementation, and how to apply them to solve real-world problems.

Topics Covered:

- Introduction and preliminaries (linear algebra, calculus, probability theory) (1 week)
- Supervised learning (KNN, Bayes classifier, decision tree, linear regression, perceptron and logistic regression, support vector machine and kernel trick, neural networks) (5 weeks)
- Learning theorems (PAC, VC dimension) (1 week)
- Unsupervised learning (including clustering, EM algorithm) (2 weeks)
- Advanced topics in machine learning, including deep learning, reinforcement learning, and causal modeling (2-3 weeks)
- Software and packages for machine learning and deep learning (scikit learn, PyTorch) (1 week)
- Group project representation

Grading: Assignment 30%, mid-term 20%, group project 30%, final 20%.

Assignment Description:

There will be 3 assignments which are closely related to the materials and topics covered in the class. Students are not allowed to use ML packages. The topics of assignments are shown below.

- Implementation of linear regression with various regularization.
- Implementation of various gradient descent algorithms for support vector machine.
- Implementation of feed forward neural networks and backpropagation.

Group Project Description:

The purpose of the group project is to deepen the exploration of machine learning with real-world data. Students will need to write code, run it on the data, make some figures, write a few pages describing the task, the algorithm(s) used, and the results obtained. Students are free to use any online code or third-party sources as long as it is publicly available. Each group will give a presentation at the end of the semester. For the topic, there are three basic options:

- Kaggle: Titanic: Machine Learning from Disaster: Use the Kaggle competition dataset to predict which passengers were likely to survive in the Titanic sank, based on features such as class of travel, gender, age, etc. https://www.kaggle.com/c/titanic
- Kaggle: Digit Recognizer: Use the MNIST dataset to correctly identify digits from a dataset of tens of thousands of handwritten images. https://www.kaggle.com/c/digit-recognizer
- Custom: Find the topic and dataset based on students' interest. Students need to send the task to the instructor to make sure that the task can be executed reasonably before the deadline.

Academic Dishonesty Policy

As a core part of its mission, the University of Arkansas provides students with the opportunity to further their educational goals through programs of study and research in an environment that promotes freedom of inquiry and academic responsibility. Accomplishing this mission is only possible when intellectual honesty and individual integrity prevail. Each University of Arkansas student is required to be familiar with and abide by the University's 'Academic Integrity Policy' at honesty.uark.edu. Students with questions about how these policies apply to a particular course or assignment should immediately contact their instructor.

Class/laboratory schedule: Meets either 3 times a week for 50 minutes or 2 times a week for 75 minutes for 15 weeks.

Prepared by: Lu Zhang **Date**: August 13, 2025