**DEPARTMENT OF THE ARMY**
**UNITED STATES MILITARY ACADEMY**
**DEPARTMENT OF MATHEMATICAL SCIENCES**
**WEST POINT, NY 10996**

MADN-MTH 16 Dec 2025

MEMORANDUM FOR Students of MA477 Theory and Application of Data Science

SUBJECT: Instructional Memorandum

1. **Purpose**. This memorandum specifies materials, describes the goals and objectives, and announces policies and procedures for MA477 during AY 26-2.

2. **Course Overview**. This course builds on foundations in probability and statistics, linear algebra, and computer programming to give students experience with a wide range of concepts and models in data science. We present the mathematical theory for each model and demonstrate its application on interesting and relevant data sets. Topics may include classification and regression trees, regularization methods, splines and localized regression, and model validation. We use the computer programming language Python daily to explore data.

3. **Required Course Materials and Resources**. Please obtain a digital or printed copy of the 3rd edition of the Géron's Hands-On Machine Learning textbook prior to the first day of class.

a. James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. An Introduction to Statistical Learning. Vol. 112. New York: Springer, 2013. Available digitally at https://www.statlearning.com/.
b. Géron, Aurélien. Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow. O'Reilly Media, Inc., 2023.
c. Course webpage: MA477 Canvas Page.
d. Google Colab: We distribute and collect all lesson materials through Canvas and will primarily use Colab. Please ensure you have a personal Google account prior to the first day of class if you don't have one.

4. **Optional Additional Resources.** Python is one of the most widely used programming languages in the world. There are extensive resources available online. Here are a few other resources:

a. VanderPlas, Jake. Python data science handbook: Essential tools for working with data. O'Reilly Media, Inc., 2016.
b. ISL-python github repository: https://github.com/emredjan/ISL-python. Contains labs and exercises from Introduction to Statistical Learning in Python.
c. Ana Bell, Eric Grimson, and John Guttag. 6.0001 Introduction to Computer Science and Programming in Python. Fall 2016. Massachusetts Institute of Technology: MIT OpenCourseWare, https://ocw.mit.edu.
d. DataCamp – Online courses and training in Python and R. www.datacamp.com

5. **Course Design.** The course is divided into five blocks with the following themes and outcomes. Each block includes both conceptual lessons and hands-on coding exercises in Python.

**Block I: Regression**

1. Understand the steps involved in completing an end-to-end machine learning project.
2. Incorporate quantitative and qualitative predictors in regression models.
3. Explain key concepts in regression, including overfitting, train-test-validation data splitting, and the importance of scaling predictor variables.
4. Apply data visualization techniques to identify prominent predictor variables.
5. Understand and implement gradient descent and stochastic gradient descent algorithms.
6. Understand and implement regularization methods to prevent overfitting.

**Block II: Classification**

1. Understand and apply logistic regression to classification problems.
2. Understand and apply Linear Discriminant Analysis (LDA) to classification problems.
3. Understand and implement Support Vector Machines (SVM) for classification.
4. Apply the K-Nearest Neighbors (KNN) algorithm in classification problems.

**Block III: Advanced Tree-Based Techniques**

1. Use tree-based methods for regression and classification by segmenting the predictor space into simpler regions.
2. Understand and apply decision trees to statistical learning problems.
3. Implement and utilize advanced tree-based methods such as bagging, random forests, and boosting for prediction.

**Block IV: Neural Networks**

1. Understand and implement neural networks for machine learning tasks.
2. Understand and implement convolutional neural networks (CNNs) for various tasks.

**Block V: Unsupervised Learning and Introduction to NLP**

1. Apply unsupervised learning techniques to extract insights from data.
2. Understand the Linear Algebra concepts behind Principal Component Analysis (PCA).
3. Implement K-means clustering algorithms using Python.
4. Gain an introductory understanding of Natural Language Processing (NLP).

6. **Evaluation**. The following table shows the distribution of points in the course.

| ***Graded Event*** | ***Points*** | ***Total Points*** |
|---|---|---|
| *AI Certification Assignment* | *10* | *10* |
| *Lesson Preparation Checks* | *2* | *50* |
| *Quizzes (8) (Drop lowest)* | *50* | *350* |
| *Problem Sets (5)* | *70* | *350* |
| *Projects (3)* | *80* | *240* |
| ***Total*** | | 1000 |

a. Lesson Preparation Checks. For lessons with assigned reading, cadets will be *randomly selected* to provide a brief summary of the material. Each cadet should expect to present a 3-5 briefings during the semester. You may use any handwritten notes during your explanation. Each briefing will be evaluated on a 0–2 scale, and your scores will be averaged and then scaled to a maximum of 50 points.
b. Quizzes. After each group of lessons, you will complete an in-class quiz designed to assess both your conceptual understanding and your technical proficiency with the material.
c. Problem Sets. These assignments will generally be due every 3 weeks or so. Mastery of a skill requires repetition and formative feedback.
d. Projects. There will be three projects (two individual, one group). All projects must be submitted in LaTeX and will be evaluated on both the correctness of the analysis and the quality of the mathematical writing
e. Bonus Points. A Cadet can earn up to 10 bonus points throughout the course. Bonus points may be awarded at the instructor's discretion.
f. Final course grades will follow the Department of Mathematical Sciences grade cut-offs as shown in Table 1 below. Students can view their performance on all graded events during the course in Canvas. Late assignments receive a 10% deduction per 24 hours late.

Table 1. Grade cut-offs

| SUBJECTIVE INTERPRETATION | LETTER GRADE | NUMERICAL GRADE | QUALITY POINT |
|---|---|---|---|
| Beyond expectations of course | A+ | 97-100 | 4.33 |
| Dominates the material | A | 93-97 | 4.00 |
| Mastery | A- | 90-93 | 3.67 |
| Excellent performance | B+ | 87-90 | 3.33 |
| Good understanding | B | 83-87 | 3.00 |
| Proficient; Aptitude for the subject | B- | 80-83 | 2.67 |
| Can build upon this foundation | C+ | 77-80 | 2.33 |
| Passing; Proficient now (short-range) | C | 73-77 | 2.00 |
| Short-range understanding | C- | 70-73 | 1.67 |
| Marginal performance with some elementary understanding | D | 65-70 | 1.0 |
| Failing; Definitely failed to demonstrate understanding | F | 0-65 | 0 |

7. **Philosophy**.

a. This course brings together your skills in statistics, linear algebra, and programming to develop models that learn from data, providing hands-on experience with real-world applications.
b. You will gain experience performing analyses similar to those used by data scientists and operations research analysts in industry, academia, and the public sector.
c. Each model is explored through its mathematical foundations and applied to engaging datasets, balancing theoretical understanding with practical implementation.
d. Active engagement is key. You will develop, code, test, and debug models yourself—building the skills and confidence needed to tackle challenges independently. Watching others may seem easy, but the real learning happens when you do it yourself.

8. **Academic Integrity**. At West Point, we are committed to the philosophy of academic integrity because it is solidly consistent with the ethical expectations of professional military officers and the fundamental tenets of the Cadet Honor Code. We rely on your documentation of submitted written work to assess the degree to which you understand the material of the course.

a. **DAAW**. Any work that you submit must maintain proper documentation in accordance with the USMA documentation guide. Projects will include a title page and follow the Dean's Policy on Documentation and Acknowledgement of Academic Work (DAAW).
b. **Generative AI.** Generative AI can both support and hinder learning. In this course, you may use AI tools (e.g., CoPilot, Gemini, ChatGPT) to prepare for class, help generate code, and assist with drafting or editing written responses, provided their use is documented according to the DAAW guidance. Keep in mind that AI-generated material may be inaccurate or incomplete. It is your responsibility to ensure all code meets project requirements and aligns with course principles. Specific guidance on using generative AI, along with expectations for its use, will be provided for each problem set and project.

   If you use generative AI:
   i. Describe how the material was generated, including prompts used.
   ii. Summarize the output (preferably with a link to the conversation).
   iii. Explain how you modified the output.

   AI is not an authoritative source. Always verify AI claims with academic literature. For instance, if AI states, "Logistic regression outperforms support vector machines when variables are collinear," you must confirm this with scholarly references.

9. **How to Succeed**

a. **Course Readings**. Read the assigned sections of the textbook for each lesson. The course texts are highly readable and widely used by professionals in the field. Plan to dedicate approximately 2 hours per lesson outside of class to ensure your success. Keep in mind that many lessons build upon each other, so staying on track is essential.
b. **Debugging Code**. Expect to encounter bugs, debugging is an essential part of the learning process and a core skill for data scientists. Don't be discouraged; it's an integral part of

learning. Before seeking assistance, make a strong effort to debug your code. To minimize errors:

i. Start with small, manageable pieces of code and scale up gradually.
ii. Test your program frequently.
iii. Modularize your code whenever possible.

c. **Understanding Concepts**. Ensure you fully understand the mathematical concepts behind your program before attempting to write code. A clear grasp of the underlying theory will make coding much easier and more effective.
d. **Additional Instruction**. Additional instruction (AI) may be scheduled with your, or any, instructor and is encouraged. The primary purpose of AI is to address specific questions that you may have for specific topics or problems. You should seek AI as soon as you need it.

10. **Chain of Command**. The Department of Mathematical Sciences has an open-door policy. Your initial attempt to address problems or concerns should go through the chain of command. However, if you are uncomfortable addressing an issue with your instructor, please approach any individual in the chain. Below is the chain of command for MA477.

a. MA477 Course Director and Operations Research Program Director: COL Andrew Lee, Thayer 246B, (845) 938-5988.
b. Head of the Department: COL Michael Scioletti, Thayer 238B, (845) 938-5285.

ANDREW C. LEE
COL, MI/47
Operations Research Program Director