|  |  |
| --- | --- |
|  | **GRADUATE SCHOOL OF ENGINEERING AND MANAGEMENT**  **DASC 522 Machine Learning Course Syllabus**  **Fall 2023** |

**Meeting Times** Asynchronous

**Location** Online

**Instructor** Torrey Wagner, PhD

**Office hours** Flexible - please contact me to set up a phone or video call

**Contact Information** [torrey.wagner.2@us.af.mil](mailto:torrey.wagner.2@us.af.mil) cell 808-937-0674

**Course Description:**

This course prepares students to apply machine learning methods to solve military problems using the Python environment. Supervised and unsupervised learning techniques including classification, regression and clustering will be covered, along with an overview of reinforcement learning techniques. After reviewing classical machine learning methods, the course will focus on the structure and application of neural networks. Students will modify & extend Python scripts and apply them to several types of data analytics problems. These techniques will give students the ability to answer questions from data, including inference, estimation, and prediction. This course will use the CRISP-DM process with a focus on understanding & preparing data, selecting a modeling approach, model building, performance assessment and both evaluating and interpreting model results.

**Credits** 4

**Prerequisites** DASC 511 Object Oriented Programming Using Python, or equivalent

**Student Learning Objectives:**

1. Students will thoroughly understand classical and neural network machine learning algorithms and how to analyze data using the CRISP-DM method.

2. Students will apply machine learning processes and methods in a systematic manner to support the Air Force & Space Force to become more data-driven organizations.

3. Students will gain proficiency using Python to implement data preparation, model building, performance assessment and the evaluation and interpretation of model results.

4. Students will effectively communicate technically complex ideas and concepts in both written and spoken formats.

**Required Books and Resource Materials:**

* Hands-On Machine Learning with Scikit-Learn, Keras & Tensorflow, by Aurélien Géron, 3rd edition. Published by O’Reilly, 2023. ISBN: 978-1098125974. “HOML”
  + All textbook python code is available at <https://github.com/ageron/handson-ml3>
  + DoD members may be able to gain free electronic access to the textbook at <https://www.dodmwrlibraries.org/> through O’Reilly. Have your CAC # handy (EDIPI) when creating an account.
* An Introduction to Statistical Learning with Applications in Python, by Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. Published by Springer, 2023. “ISLP”
  + Free pdf download available at <https://www.statlearning.com/>. Links to purchase are also listed at that website.
  + Python notebook files, slides & datasets available at <https://www.statlearning.com/resources-python>
* A Python environment that supports the Jupyter Notebook .ipynb format, such as Google [Colaboratory](https://colab.research.google.com/notebooks/intro.ipynb)

**Grading Scheme/Policy:** Homework will make up 70% of your grade, participation is 12% and the take-home final project will account for 18% of your grade.Participation will be completed by contributing to course discussion boards. Subject to the instructor’s discretion, final course grades may be curved in favor of the student.

|  |  |
| --- | --- |
| **Percentage of Points Earned** | **Letter Grade** |
| grade 93% or greater | A |
| 90% ≤ grade < 93% | A- |
| 86% ≤ grade < 90% | B+ |
| 83% ≤ grade < 86% | B |
| 80% ≤ grade < 83% | B- |
| 76% ≤ grade < 80% | C+ |
| 73% ≤ grade < 76% | C |
| 70% ≤ grade < 73% | C- |
| 67% ≤ grade < 70% | D+ |
| 63% ≤ grade < 66% | D |
| 60% ≤ grade < 63% | D- |
| 0% ≤ grade < 60% | F |

**Use and Attribution of AI Tools:**

* Students are permitted to use AI tools to assist in their discussion assignments and homework, including both code & narrative generation.  If AI tools are used in an assignment, submit a separate “AI assistance” file that contains up to 3 prompts (and the AI replies), and the tool name such as ChatGPT or BARD. In the assignment, include a note that mentions which tool or tools were used to assist. This will attribute their use, and also help the instructor to better understand the capabilities of this technology.
* AI tools can give biased or fallacious results. In order to effectively leverage AI tools, students must apply critical thinking and interpretation to AI-generated content in the process of producing original work that reflects their understanding, expertise, personal insights, analysis, and creativity.
* This section was written with assistance from ChatGPT.

**AFIT Policies:**

1. **Attendance:** Viewing and reading all assigned material is mandatory except for extenuating circumstances. Scheduled readings, videos, and exams are defined by the instructor and they are documented in the course schedule. (References: Student Handbook, Graduate School Catalog)
2. **Academic Integrity:** All students must adhere to the highest standards of academic integrity. Students are prohibited from engaging in plagiarism, cheating, misrepresentation, or any other act constituting a lack of academic integrity. Failure on the part of any individual to practice academic integrity is not condoned and will not be tolerated. Individuals who violate this policy are subject to adverse administrative action including disenrollment from school and disciplinary action. Individuals subject to the Uniform Code of Military Justice may be prosecuted under it. Violations by government civilian employees may result in administrative disciplinary action without regard to otherwise applicable criminal or civil sanctions for violations of related laws. (References: Student Handbook, ENOI 36 – 107, *Academic Integrity*)
3. **Academic Grievance:** AFIT and the Graduate School of Engineering and Management affirm the right of each student to resolve grievances with the Institution. Students are guaranteed the right of fair hearing and appeal in all matters of judgment of academic performance. Procedures are detailed in ENOI 36 – 138, *Student Academic Performance Appeals*.
4. **Student Accommodations:** The Graduate School of Engineering and Management makes reasonable accommodations for persons with disabilities within American Disability Act guidelines. Students should notify the Registrar, Ms. Kathy Burden of any requests for accommodation. Accommodations will be assessed only with official documentation of disability from a licensed psychologist or physician and requested accommodation.
5. **Testing Policy:** There are no tests in this course – only project-based coursework.
6. **Late Assignments and Make-Ups:** Late work will receive a 3% penalty per day unless there is prior coordination from the student. Exceptions may be made on a case by case basis.
7. **Homework:** Homework is expected to be an individual effort so the student has the opportunity to gain, improve, and sustain proficiency in course content. Discussion and collaboration with other students is encouraged, but each student’s homework submission must be entirely that student’s own work. Since this is a coding-intensive class, it is also permissible to use code from other sources (such as stackoverflow.com) with attribution, but it is not allowed to use code from prior DASC 522 students.

**Course Weekly Overview**

*Course assignments, due dates and other requirements may be subject to change.*

|  |  |
| --- | --- |
| Week | Topic |
| 1 | Introduction & Motivation |
| 2 | Classical Supervised Learning |
| 3 | Classical Unsupervised Learning |
| 4 | Selection & Regularization |
| 5 | Neural Network Foundations & Processes |
| 6 | Resampling Neural Network datasets |
| 7 | Supervised Learning with Neural Networks |
| 8 | Unsupervised Learning with Neural Networks |
| 9 | Neural Network Debugging |
| 10 | DoD Case Studies |

**Course Schedule**

**(subject to change)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Week | Topic | Subtopics | Certificate Link | Reading | Examples |
| 1 | Introduction & Motivation | - How machine learning (ML) fits into CRISP-DM  - When to use ML  - Tools used in the course  - DoD machine learning needs  - Joint Artificial Intelligence Center | - All courses | - DoD AI Strategy  - Executive Order on AI  - ISLP Ch 1/2  - HOML Ch 1\* | - DoD AI conference content (SecDef, VCJCS etc)  \* Please avoid examples that use the “pipeline” notation – they are not required in this course |
| 2 | Classical Supervised Learning | - Regression review  - Classification  - Overview of tree-based regression & classification | - Reinforce DASC 500 & DASC 512 regression  - DASC 512 week 6 regression | - Defense Innovation Board  - AI Principles  - ISLP Ch 3/4/8  - HOML Ch 3, 4\*, 6 | - DASC 500 week 5 & DASC 512 week 6 Python regression models  - Classification in Python  \* Ch 4 just LinearRegression & LogisticRegression |
| 3 | Classical Unsupervised Learning | - k-means clustering  - Hierarchical clustering  - Dimension Reduction | - Reinforce DASC 500 clustering material | - ISLP Ch 12  - HOML Ch 9 | - DASC 500 week 8 clustering in Python |
| 4 | Classical Selection & Regularization | - Selection methods  - L1, L2 & other regularization methods | - DASC 501 databases  - DASC 511 Python modeling | - 2019 RAND DoD Posture for AI (skim)  - ISLP Ch 6 | - Extract features from database, perform selection |
| 5 | Neural Network (NN) Foundations & Processes | - Neural network foundations  - Perceptron overview  - Perceptrons & perceptron learning  - Multilayer perceptrons | - DASC 512 regression | - ISLP Ch 10.1, 10.2, 10.6  - HOML Ch 10 | - Impact of hyperparameters on 2-d dataset training |
| 6 | NN Optimizers, Training & Resampling | - Backpropagation  - Optimizers  - 3 methods of resampling | - DASC 500  data preparation | - ISLP Ch 5, 10  - HOML Ch 11 | - Manual & automated dataset selection using housing data (val/test sets, feature cross) |
| 7 | NN Supervised Learning | - Regression  - Classification  - Hyperparameter tuning | - DASC 511 Python modeling | - Re-review HOML Ch 10  - Napierala article | - One-neuron network that performs linear regression  - Regression & binary classification on housing data |
| 8 | NN Unsupervised Learning, Autoencoders & Regularization | - Autoencoder unsupervised learning  - Image recognition using autoencoders  - 6 methods of NN regularization | - DASC 511 Python modeling | - Re-review HOML Ch 9  - HOML Ch 17 | - MNIST image processing  - F-MNIST image recognition  - Higgs boson prediction |
| 9 | NN Debugging | - Data attributes & errors  - Dataset splitting issues  - Hyperparameter symptoms  - Feature engineering errors  - Model structure | - DASC 500/512 statistics |  | - Python debugging in regression  - Python debugging in classification |
| 10 | DoD ML Case Studies | - Selected case studies |  | - ML to predict mental illness | - Predicting drug use in Air Force members  - Predicting global food crises |

Textbook abbreviations

* “HOML” Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems, Geron.
* “ISLP” Introduction to Statistical Learning with Applications in Python, James, Witten, Hastie, and Tibshirani.

The course syllabus is a general plan for the course; deviations announced to the class by the instructor may be necessary.