Machine Learning (CS 5805) Fall 2023 Syllabus

Table of contents

Course description

The last decade has seen an explosive growth in database technology and the amount of data collected. This has created an unprecedented opportunity for machine learning, which is the process of efficient supervised or unsupervised discovery of interesting and useful information from collections of data. Some of the common tasks in machine learning are classification, clustering, the discovery of association rules/sequential patterns, and anomaly detection. Machine learning has seen several successful applications in diverse domains such as healthcare, economics, internet advertising, social sciences, and environmental studies. This course will give a rapid and vigorous introduction to the field of machine learning, as well as provide extensive hands-on experience via a course project. All course activities will be conducted online.

Essential Information

- Instructor: Martin Skarzynski Laptev PhD, MS, MPH
- Start Date: 2023-08-21
- End Date: 2023-12-06
- Start Time: 16:00 Eastern Time
- End Time: 17:15 Eastern Time

Class will be held

- on Mondays and Wednesdays
- according to the VT Academic Calendar
- via Zoom: https://virginiatech.zoom.us/j/86864816900?pwd=VmFjWGlhc1JCNnJVc2JzcE5SNGhadz09

Learning Aims

By the end of the course, students will:

- Be well-versed with common machine learning problems, concepts, and algorithms
- Be able to compare and contrast different machine learning algorithms and identify their

strengths and limitations in varying problem settings
- Gain practical understanding of machine learning algorithms through a course project

Materials

- Textbook (optional): Hands-on Machine Learning with Scikit-Learn, Keras and TensorFlow (3rd edition)
- Code: https://github.com/maptv/handson-ml3
- Anaconda: https://www.anaconda.com/download#downloads
- R: https://cran.rstudio.com/
- RStudio: https://posit.co/download/rstudio-desktop/
- Quarto: https://quarto.org/docs/get-started/

Anaconda includes VSCode and Jupyter (Lab and Notebook), and can be used to install RStudio, but I recommend that you install RStudio using the link above, instead of via Anaconda.

If you already use a different Python distribution that includes conda, such as miniconda, miniforge, or mambaforge, you do not have to use Anaconda. Personally, I use mambaforge. Nevertheless, Anaconda is a better choice for the vast majority of people.

Modules

- 01: The machine learning landscape
- 02: End to end machine learning project
- 03: Classification
- 04: Training linear models
- 05: Support vector machines
- 06: Decision trees
- 07: Ensemble learning and random forests
- 08: Dimensionality reduction
- 09: Unsupervised learning
- 10: Neural nets with Keras
- 11: Training deep neural networks
- 12: Custom models and training with Tensorflow
- 13: Loading and preprocessing data
- 14: Deep computer vision with CNNs
- 15: Processing sequences using RNNs and CNNs
- 16: NLP with RNNs and attention
- 17: Autoencoders GANs and diffusion models
- 18: Reinforcement learning
- 19: Training and deploying at scale

It is very unlikely that we will get through all of the modules.

Background Required

Some familiarity with Python will be very helpful, but there are no strict requirements. This course will start from zero and advance quickly.

Grading

There will not be any quizzes or exams. Instead, students must create a machine learning blog using Quarto publish it on GitHub Pages. This blog is the only graded deliverable for the course. The grading for the blog will be as follows:

- 1 point: a blog hosted on GitHub Pages
- 3 points: 3 blog posts that are each worth 1 point

To get a full point for the blog, its content and the code used to publish it must be available in a GitHub repository.

To get a full point for a blog post (full-point post), it must have - machine learning code and - at least one data visualization.

A blog post without any machine learning code or without any data visualizations will only get half a point (half-point post).

Example grades:

- A (4 points):
  - A blog with 3 full-point posts that each have
    - machine learning code and
    - at least 1 data visualization
- B (3 points):
  - A GitHub repository with the content for 3 full-point posts, but no blog
  - A blog with 2 full-point posts
  - A blog with
    - 1 full-point post and
    - 2 half-points posts
- C (2 points):
  - A GitHub repository with the content for 2 full-point posts, but no blog
  - A GitHub repository with the content for
    - 1 full-point post
    - 2 half-point posts
  - A blog with 1 full-point blog post
  - A blog with 2 half-point blog post
- D (1 points):
  - A GitHub repository with the content for 1 full-point post, but no blog
  - A GitHub repository with the content for 2 half-point posts, but no blog
  - A blog with no machine learning content
- F (0 points):
  - No deliverables

The course GTA(s) will provide feedback for the assignments, but final grading for the course project will be determined by the instructor.

Attendance and Class Participation

Attendance and class participation will not affect your grade. To really benefit from the course, you should attend and participate, but you can always review class recordings instead if your prefer.

Late Policy

Late submissions will not be penalized. As long as you submit the course project before the last class, I will give full credit. If you wait until the day of the last class to do all of the work, you will miss out on valuable feedback from the GTA(s) and probably not benefit much from the experience of rushing through the process.

Zoom Best Practices

We will be using Zoom for conducting all class activities and office hour discussions. The Zoom URLs are provided at the top of this document and can also be accessed from the Zoom tab on the left panel of the Canvas page of the class. Please familiarize yourself with Zoom and student tips for remote learning (see: https://tutorials.tlos.vt.edu/index/zoom.html and https://teaching.vt.edu/OurServices/StudentTips.html). You should keep your video turned on during the class to remain attentive and compensate for the lack of physical interactions in an online environment, unless restricted by low internet bandwidth. You may keep your audio muted unless you have a question to ask or need to respond to an on-going discussion to avoid interference and feedback. You can also type your question in Zoom's chat interface or provide nonverbal feedback and express opinions by clicking on icons in the Participants panel at any point during the lecture. We will be using Zoom Breakout rooms feature during office hours to facilitate one-on-one or group interactions of the instructor and TAs with students.

Communications and Feedback

We will be using Canvas Announcements as our preferred mode of communication to notify any changes to the class schedule and activities, so please ensure that your Canvas

Notification Preferences are set to notify you (typically via email) when an Announcement has been posted.

Academic Integrity

The tenets of the Virginia Tech's Honor Codes will be strictly enforced in this course, and all assignments shall be subject to the stipulations of the Undergraduate and Graduate Honor Codes. For more information on the Graduate Honor Code, please refer to the GHS Constitution at http://ghs.graduateschool.vt.edu. All paper reviews, project reports, and other submissions must represent your own individual effort. Students are encouraged to consult with one another about project design and evaluation issues, whether performed individually or in groups, as long as the individual submissions represent their individual efforts. Be particularly careful to avoid plagiarism, which essentially means using materials (ideas, code, designs, text, etc.) that you did not create without giving appropriate credit to the creator (using quotation marks, citations, comments in the code, link to URL, etc.). We will also adhere to Virginia Tech's Principles of Community for all in-class discussions and activities, to maintain a safe, welcoming, and respectful environment for every student in the class. For more information, see: https://www.inclusive.vt.edu/Initiatives/vtpoc0.html.

Accommodations for Students with Special Needs

Students with special needs will be provided additional resources and materials to aid in their learning. Mode of communication during the class will be adjusted in lieu of the respective needs of the student. Please discuss your requirements with the instructor so that we can work together to make a comfortable environment for everyone. Please see: https://www.ssd.vt.edu/ for more information. If you have an emergency medical information, please let me know privately as soon as possible.

Virginia Tech Community Wellness Commitment

Virginia Tech is committed to protecting the health and safety of all members of its community. **By participating in this class, all students agree to abide by the Virginia Tech Wellness principles** (https://ready.vt.edu/public-health-guidelines.html#wellness) **. Be respectful of the well-being of others, as well as individual choices about masking. Students who prefer to wear masks in class are always welcome to do so.**

Student Well-Being Support

Supporting the mental health and well-being of students in our class is of high priority to us and Virginia Tech. If you are feeling overwhelmed academically, having trouble functioning, or are worried about a friend, please reach out to any of the following offices:

Cook Counseling:

- 540-231-6557 to schedule an appointment and/or 24/7 crisis support
- ucc.vt.edu for more information

Dean of Students Office:

- 540-231-3787 for general advice
- 540-231-6411 for after-hours crisis
- dos.vt.edu for more information

Hokie Wellness:

- hokiewellness.vt.edu for more information about health and wellness workshops and consultations

Services for Students with Disabilities (SSD)

- 540-231-3788 or ssd.vt.edu for more information about accommodations and other disability-related supports
- Any student who has been confirmed by the University as having special needs for learning must notify me by email in the first week of the course--work through SSD so

they can send electronic notification of any accommodations.

Student Success Center:

- The Student Success Center helps students develop the skills needed to accomplish their academic goals and become self-directed learners. Their free services include individual and group tutoring, peer academic coaching, a Seminar Series on Academic Success, and more. Students can book appointments through Navigate. For instructions and more information, please visit www.studentsuccess.vt.edu.

For a full listing of campus resources check out well-being.vt.edu.

Please also feel free to speak with any of the instructors. We will make an effort to work with you; we care about you.

Academic Accommodations

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, chronic medical conditions, Deaf or hard of hearing, learning disability, mental health, or vision impairment, please contact the Services for Students with Disabilities (SSD) (540-231-3788, ssd@vt.edu, or visit ssd.vt.edu). If you have an SSD accommodation letter, please meet with your instructor privately during office hours as early in the semester as possible to discuss implementing your accommodations. You must give reasonable notice to implement your accommodations, which is generally 5 business days and 10 business days for final exams.

Technical support

Technical: For technical support assistance regarding any problems with Canvas, Zoom, or e-mail, please contact 4Help. For technical support issues related to Web-CAT or CodeWorkout, send an e-mail request to webcat@vt.edu providing specific details of the problem or issue. For questions related to programming assignments or homework, ask questions on the class general discussion area on Canvas, where the instructor or the TA(s) can provide answers.

Canvas privacy policy: http://www.canvaslms.com/policies/intl-privacy