Projects
Machine Learning Course
Fall 2025

Westlake University
Department of Artificial Intelligence, SOE
Tao Lin
https://github.com/LINs-lab/course_machine_learning

Course Overview — EST-3000

Machine learning methods are becoming increasingly central in many sciences and applications. In this course, fundamental principles and methods of machine learning will be introduced, analyzed and practically implemented.

Topics include: supervised learning (linear/non-linear models, generative/discriminative learning, parametric/non-parametric learning, support vector machines, neural networks); unsupervised learning (clustering, dimensionality reduction); learning theory (bias/variance trade-offs, practical advice).

The course will offer hands-on machine learning experiences through five graded homeworks, several coding exercises, and two graded course projects, followed by the final close-book exam.

Course Goals

By the end of the course, the student must be able to

- define the following basic Machine Learning (ML) models: regression, classification, clustering, dimensionality reduction, neural networks, and explain the main differences between them;
- explain and understand the fundamental theory behind these ML models;
- implement these ML models, and optimize the main trade-offs such as overfitting, and computational cost;
- implement ML models for real-world problems, and rigorously evaluate their performance using cross-validation;
- experience common pitfalls of ML in practice and how to overcome them.

Syllabus

We will cover the following ML methods and concepts:

- Basic regression and classification concepts and methods:
  Linear models, overfitting, linear regression, Ridge regression, logistic regression, SVMs, and k-NN
- Fundamental concepts:
  Cost-functions and optimization, cross-validation and bias-variance trade-off, curse of dimensionality, kernel methods.
- Neural networks:
  Basics, representation power, backpropagation, CNNs, transformer models, regularization, data augmentation, dropout.
- Unsupervised and self-supervised learning:
  k-means clustering, Gaussian mixture models, the EM algorithm, generative models, large language models, diffusion models, generative adversarial networks.

The syllabus provided on the website is more precise but subject to change.

Exercise sessions

Weekly every Thursday 08:00 - 09:35, in person in the following rooms.

All practical Labs and projects will be in *Python*. See lab 2 to get started. Don't worry if you have no experience in it yet, but in that case you should take enough time to thoroughly work on the first (and second) lab.

Prerequisites

We will revise some of basic ML concepts in the first and second weeks of the course. However, you are recommended to go through the list of pre-requisites here to make sure your knowledge is up to date.

Vector and Matrix Algebra. Vector and matrix multiplication, matrix inverse, rank, eigenvalue decomposition. Refer to first year courses, or the Interactive Linear Algebra on the website, or Gilbert Strang's book for example.

Vector and Matrix Calculus. Important: The definition of derivative with respect to vectors and matrices. For reference, see this blogpost explained.ai/matrix-calculus, or the Matrix Cookbook for example.

Scientific Computing Languages. Python Basics (see tutorial in lab 1).

Probability and Statistics. Conditional and joint distribution, independence, Bayes' rule, random variable and expectation, law of large numbers.

Gaussian Distribution. Univariate and multivariate, conditional, joint and marginals.

Writing Scientific Documents using Latex (not required but preferred). Many tutorials are available online, and we provide more resources when we come to Project 1.

Resources

Course Webpage

All materials will be made available on our public github repository, including annotated lecture notes, code and exercise solutions.

Lecture Notes

PDF notes for each lecture will be available on the website (and github) before the day of the lecture, and will often be annotated during the lecture. For revisions in case of errors, see also on github.

Recommended Textbooks

No book is mandatory for this course. Nevertheless, the following examples contain parts relevant to the course:

- G. Strang: Linear Algebra and Learning from Data
- S. Shalev-Shwartz and S. Ben-David: Understanding Machine Learning - From Theory to Algorithms
- G. James, D. Witten, T. Hastie and R. Tibshirani: An introduction to statistical learning
- T. Hastie, R. Tibshirani and J. Friedman: Elements of statistical learning
- C. Bishop: Pattern Recognition and Machine Learning
- K. Murphy: Machine Learning: A Probabilistic Perspective
- Michael Nielsen: Neural Networks and Deep Learning

Assessment and Practical Projects

- Assignments ($5 \times 4\%$)
- Project 1 (10%), due Oct. 23, 23:59
- Project 2 (30%), due Dec. 25, 23:59
- Final exam (40%)

Assignments (20%)

There are five graded homeworks, on the topics of (1) math foundation, (2) linear model, (3) kernel methods, (4) EM algorithm, and (5) neural network foundation.

Project 1 (10%)

The goal of this project is to help you prepare for Project 2.

In this first project, you will work in a group of 1 people.

You will implement the most important methods covered in the lectures and labs so far.

Additionally, we will provide you with an interesting real-world dataset, and organize our own competition.

A detailed project description will be posted on the public github repo very soon.

You will also submit your Python code, and a 2 page PDF report. *Deadline: Oct. 23*.

Project 2 (30%)

Project 2 is the final project and gives you more freedom and responsibilities.

Again, you will work in a group of 1 people.

You can freely choose between two options:

A) **Machine Learning for Science**: Pick a real-world challenge offered by any research group of the Westlake University.

   A list of potential project ideas will be made available to students later, and is subject to availability. It is also possible that you reach out directly to some research groups early in the semester, and ask for a potential project idea. Labs who are interest to host a student group can contact us here to offer their project idea, dataset or task. Projects ideas need to be approved by the group in question and by us, early November.

B) Pick one of two **pre-defined challenges** with real-word data problems.

   Submitting your predictions to a competition platform allows you to get immediate feedback on your performance.

In cases A) and B), you will submit your project as Python code, and a 4 page PDF report.

Deadline for all cases: Dec. 25.

Final exam (40%)

A very standard final exam.

It will contain questions on what you have learned during the lectures and exercise sessions.

We will give you a sample exam before for you to practice.

You are allowed to bring one cheat sheet (A4 size paper, both sides can be used).

No calculator, No collaborations. No cell phones. No laptops etc.

Contact Us

Please use the canvas for any questions and feedback on the course material or exercises, or email the respective assistants and teachers.

Teaching Assistants:

Yingming PU (Organizing TA)

Bangyan LIAO

The assistants will be helping you during the exercise sessions and projects.

Credits

Teaching material by Tao Lin and Kaicheng Yu.

Note that we heavily reused the course materials from EPFL CS-433 machine learning course.