Mathematics for Machine Learning
— Introduction

Joseph Chuang-Chieh Lin

Department of Computer Science & Information Engineering,
Tamkang University

Fall 2023

Credits for the resource

- The slides are based on the textbook:
  - *Marc Peter Deisenroth, A. Aldo Faisal, and Cheng Soon Ong: Mathematics for Machine Learning. Cambridge University Press. 2020.*
  - *Howard Anton, Chris Rorres, Anton Kaul: Elementary Linear Algebra. Wiley. 2019.*
- We could partially refer to the monograph:
  Francesco Orabona: A Modern Introduction to Online Learning. https://arxiv.org/abs/1912.13213

Grading Policy

- Attendance (10%)
- Assignments & Quizzes (30%)
- Midterm Exam (30%)
  - 7 Nov. 2023.
- Final Exam (30%)
  - 2 Jan. 2024.

Outline

Three Core Concepts of Machine Learning

- Data
- Model
- Learning

Remark on the Data

- Machine learning is inherently data driven.

Remark on the Data

- Machine learning is inherently data driven.
  - Garbage in, garbage out.

Mitchell 1997

A model is said to learn from data if its performance on a given task improves after the data is taken into account.

Remark on the Data

- Machine learning is inherently data driven.
  - Garbage in, garbage out.

Mitchell 1997

A model is said to learn from data if its performance on a given task improves after the data is taken into account.

- **Goal:** Find good models that generalize well to yet unseen data

Four pillars of ML

The four pillars of ML:

- Regression
- Dimensionality Reduction
- Density Estimation
- Classification

Fundamentals:

- Calculus
- Linear Algebra
- Vector Algebra
- Analytic Geometry
- Matrix Decomposition
- Probability & Distributions
- Optimization

- Why are the mathematical foundations of machine learning important?

- Why are the mathematical foundations of machine learning important?
  - To understand fundamental principles upon which more complicated machine learning systems are built.
  - To facilitate creating new machine learning solutions, understanding and debugging existing approaches.
  - To learn about the inherent assumptions and limitations of the methodologies we are working with.

What's a machine learning *algorithm*?

- Predictor: A system that makes predictions based on input data.
- Training: a system that adapts some internal parameters of the predictor so that it performs well on future unseen input data.

Some Reasonable Assumptions

- Numerical representation of the data:

Some Reasonable Assumptions

- Numerical representation of the data: vectors.
  - An array of numbers (CS view)

Some Reasonable Assumptions

- Numerical representation of the data: vectors.
    - An array of numbers (CS view)
    - An arrow with a direction and magnitude (physics view)

Some Reasonable Assumptions

- Numerical representation of the data: vectors.
  - An array of numbers (CS view)
  - An arrow with a direction and magnitude (physics view)
  - An object that obeys addition and scaling (mathematical view; OOP view).

An Intuition of Learning/Training a Model

- Assume that we are given a dataset and a suitable model.
- Training a model: use the data to optimize parameters of the model w.r.t. some loss/utility function.
- The training process can be viewed as either climbing a hill to reach its peak moving downwards to the valley.

An Intuition of Learning/Training a Model

- Assume that we are given a dataset and a suitable model.
- Training a model: use the data to optimize parameters of the model w.r.t. some loss/utility function.
- The training process can be viewed as either climbing a hill to reach its peak moving downwards to the valley.
- However, at the same time, we are interested in the model which performs well on unseen data.

  Otherwise, it could be just that we find a way to memorize the data.

Part I.

Mathematics as the Foundation

Why linear algebra/vector algebra?

- The study of vectors and matrices.

Why linear algebra/vector algebra?

- The study of vectors and matrices.
- Represent numerical data as vectors.

Why linear algebra/vector algebra?

- The study of vectors and matrices.
- Represent numerical data as vectors.
- Represent a table of data as a matrix.

Why linear algebra/vector algebra?

- The study of vectors and matrices.
- Represent numerical data as vectors.
- Represent a table of data as a matrix.
- Formalize the *similarity* between vectors:
    - Analytic geometry (distance, norm, inner product, projection, ...)

Why linear algebra/vector algebra?

- The study of vectors and matrices.
- Represent numerical data as vectors.
- Represent a table of data as a matrix.
- Formalize the *similarity* between vectors:
  - Analytic geometry (distance, norm, inner product, projection, ...)
- Intuitive interpretation of the data and better efficiency for learning: *matrix decomposition*.

Part II:

Introductory Machine Learning

Topics

- Data, model & parameter estimation.
- Continuous Optimization.
- Linear regression.
  - Map the input $\mathbf{x} \in \mathbb{R}^d$ to corresponding observed function values $y \in \mathbb{R}$.
- Density estimation.
  - Find a probability distribution that describes the data.
- Principal Component Analysis
  - Matrix decomposition.
- Classification.

Terminologies

- i.e. $\Longrightarrow$ that is,
- e.g. $\Longrightarrow$ such as
- $\because \Longrightarrow$ because
- $\therefore \Longrightarrow$ therefore
- et al. $\Longrightarrow$ and others
- $\forall \Longrightarrow$ for any
- $\exists \Longrightarrow$ there exists
- a.k.a. $\Longrightarrow$ also known as
- w.r.t. $\Longrightarrow$ with respect to

Warm-up Exercise

Exercise

- Consider $\mathbf{x} = [x_1 \;\; x_2 \;\; x_3]^\top \in \mathbb{R}^3$ and $\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$.
- Compute $\mathbf{x}^\top \boldsymbol{A}\mathbf{x}$.
- Compute $\mathrm{tr}(\boldsymbol{A}\mathbf{x}\mathbf{x}^\top)$.

Discussions