Lecture 1: Introduction

CSE4130: 기초머신러닝

Junsuk Choe (최준석)

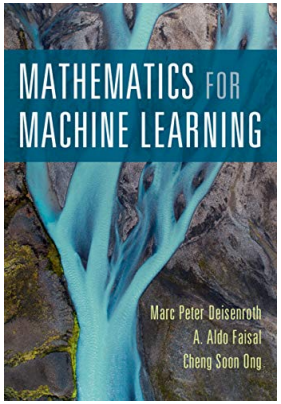
Course Intro

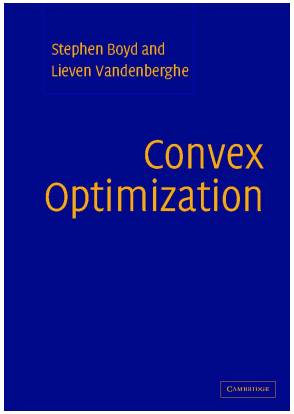
- Course Title: Basic Machine Learning (CSE4130)
- Time and Location: Tue/Thurs 09:00-10:15, In-person lectures

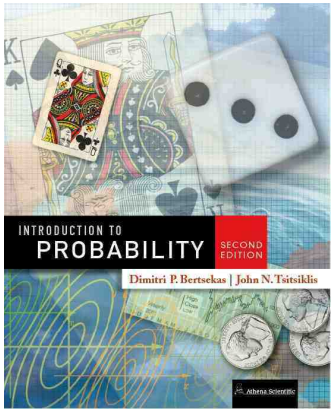
- Instructor:
  - Junsuk Choe 최준석 (AS913, jschoe@sogang.ac.kr)
- TA:
  - Yeji Park 박예지 (AS915, yjparkm@sogang.ac.kr)
  - Beomyun Kwon 권범윤 (AS915, beomyunkwon@gmail.com)

- Lecture slides and other information will appear at http://cyber.sogang.ac.kr

Textbook







- Mathematics for Machine Learning[1], Cambridge University Press, Marc Peter Deisenroth, A. Aldo Faisal, and Cheng Soon Ong
- Other books
  - Convex Optimization, Cambridge University Press, by Stephen Boyd and Lieven Vandenberghe
  - Introduction to Probability, 2nd edition, Athena Scientific, by Dimitri P. Bertsekas and John N. Tsitsiklis

[1] The entire textbook can be downloaded at `https://mml-book.github.io/`

Evaluation

- Mid-term Exam (40%)
- Final Exam (40%)
- Assignments (20%)

- Academic integrity
  - **What constitutes academic dishonesty?**
    - Cheating, fabrication, plagiarism, unauthorized collaboration (or facilitating any of these)
  - **What are the penalties and sanctions?**
    - receiving 0 point for any exam/assignment
    - Ignorance is no excuse

Course Policies

- Attendance and Work:
  - All students should attend class unless discussed with the instructor.
  - For the absence without notice, Sogang University regulation will be applied.
- Language:
  - The lecture will be given in Korean.

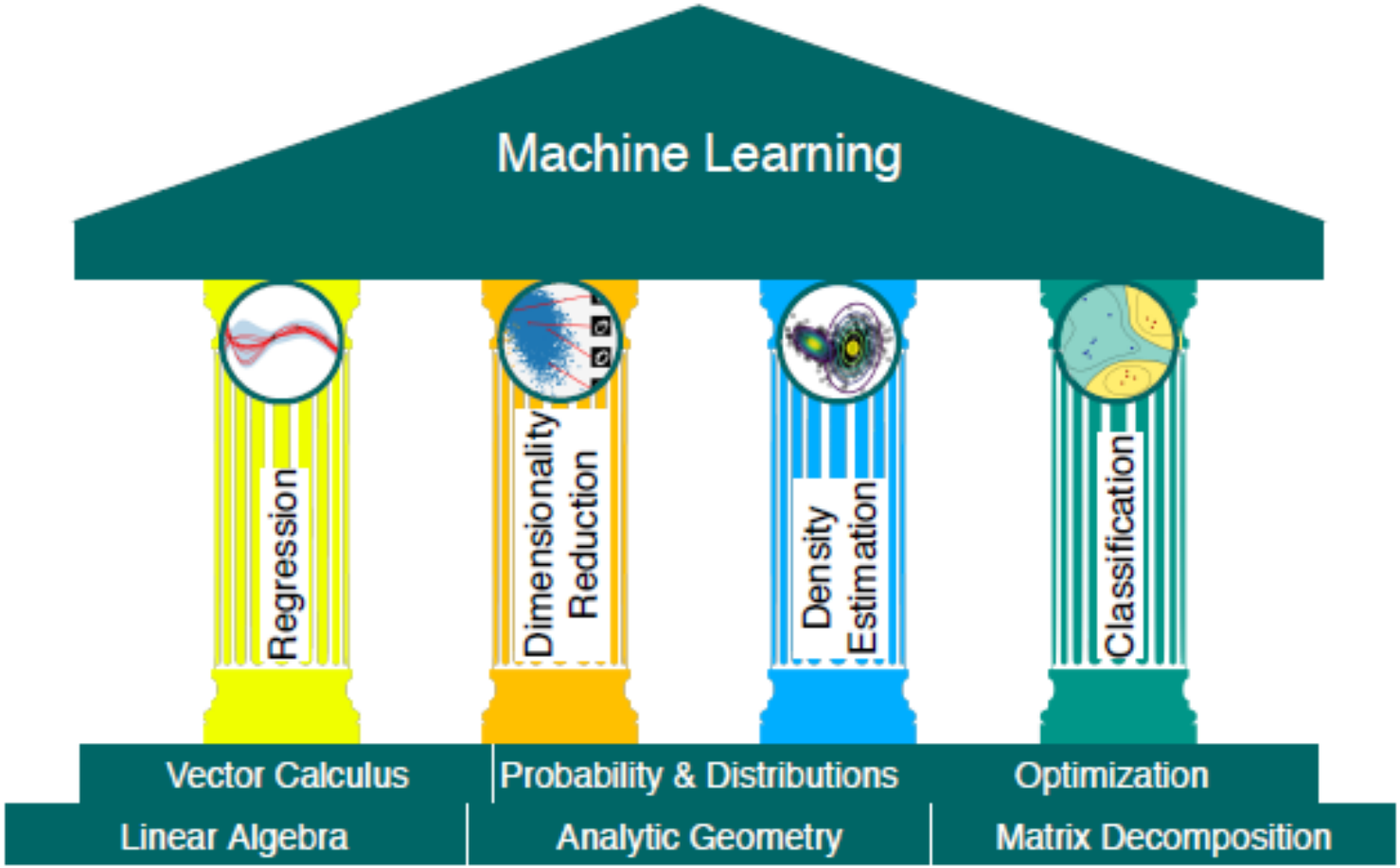
Course Policies

- Exams will be taken in-person.
- Questions will be based mainly on the assignments and textbook problems.
- **Honor code:**
  - Exam cheating is unacceptable in any academic environment.
  - If exam cheating is identified, it will get 0 point.

Organization

- Part I: Math
  1. Linear Algebra
  2. Analytic Geometry
  3. Matrix Decomposition
  4. Vector Calculus
  5. Probability and Distributions
  6. Optimization
- Part II: 4 Basic Machine Learning Problems
  1. When Models Meet Data
  2. Dimensionality Reduction with Principal Component Analysis
  3. Density Estimation with Gaussian Mixture Models
  4. Classification with Support Vector Machines

Organization



Suggestions on Course Schedules

Total 16 weeks

- Part I: Math
  1. Linear Algebra (2 weeks)
  2. Analytic Geometry (1 week)
  3. Matrix Decomposition (1 week)
  4. Vector Calculus (1 week)
  5. Probability and Distributions (2 weeks)
  6. Optimization (2 weeks)
- Part II: 4 Basic Machine Learning Problems
  1. When Models Meet Data (1 week)
  2. Dimensionality Reduction with Principal Component Analysis (1 week)
  3. Density Estimation with Gaussian Mixture Models (1 week)
  4. Classification with Support Vector Machines (1 week)
- Total 13 weeks + Midterm (1 week) + Final (1 week) + Extra (1 week)

Target Audience

- Undergraduate students
  - They may have partial backgrounds on the math (e.g., only vector calculus + linear algebra).
  - Depending on the students' background, the amount of time for math can be adjusted.
  - Some mathematical parts may need to be provided with some degree of rigorous proofs.

Basic Notations

- Scalars: $a, b, c, \alpha, \beta, \gamma$
- Vectors: $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$
- Matrices: $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}$
- Sets: $\mathcal{A}, \mathcal{B}, \mathcal{C}$
- (Ordered) tuple: $B = (\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3)$
- Matrix of column vectors: $\boldsymbol{B} = [\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3]$ or $\boldsymbol{B} = \begin{pmatrix} \boldsymbol{b}_1 & \boldsymbol{b}_2 & \boldsymbol{b}_3 \end{pmatrix}$
- Set of vectors: $\mathcal{B} = \{\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3\}$
- $\mathbb{R}$, $\mathbb{C}$, $\mathbb{Z}$, $\mathbb{N}$, $\mathbb{R}^n$, etc
- Probability: We use both $p(\cdot)$, $\mathbb{P}[\cdot]$.

Introduction and Motivation

- Machine learning is about designing algorithms that automatically extract valuable information from data.
- There are three concepts that are at the core of machine learning:
  - **data**
    - The goal of machine learning is to design general purpose methodologies to extract valuable patterns from data, ideally without much domain-specific expertise.
  - **a model**
    - We design models that are typically related to the process that generates data, similar to model the dataset we are given.
    - The goal is to find good models that generalize well to yet unseen data, which we may care about in the future.
  - **learning**
    - Learning can be understood as a way to automatically find patterns and structure in data by optimizing the parameters of the model.

Introduction and Motivation

- Machine learning has seen many success stories, and software is readily available to design and train rich and flexible machine learning systems.
- However, the mathematical foundations of machine learning are important in order to understand fundamental principles upon which more complicated machine learning systems are built.
- Understanding these principles can facilitate creating new machine learning solutions, understanding and debugging existing approaches, and learning about the inherent assumptions and limitations of the methodologies we are working with.

Introduction and Motivation

- A challenge we face regularly in machine learning is that concepts and words are slippery.
- For example, the word “algorithm” is used in at least two different senses in the context of machine learning:
  - We use the phrase “machine learning algorithm” to mean a system that makes predictions based on input data. We refer to these algorithms as *predictor*s.
  - We also use the exact same phrase “machine learning algorithm” to mean a system that adapts some internal parameters of the predictor so that it performs well on future unseen input data. Here we refer to this adaptation as *training* a system.
- This lecture will not resolve the issue of ambiguity. However, I attempt to make the context sufficiently clear to reduce the level of ambiguity.

Introduction and Motivation: Data

- In this lecture, we assume that data has already been appropriately converted into a numerical representation suitable for reading into a computer program. **Therefore, we think of data as vectors.**

- Three different ways to think about vectors
  - a vector as an array of numbers (a computer science view)
  - a vector as an arrow with a direction and magnitude (a physics view)
  - a vector as an object that obeys addition and scaling (a mathematical view)

Introduction and Motivation: Model

- A **model** is a process for generating data, similar to the dataset at hand.
- Therefore, good models can also be thought of as simplified versions of the real (unknown) data-generating process, capturing aspects that are relevant for modeling the data and extracting hidden patterns from it.
- A good model can then be used to predict what would happen in the real world without performing real-world experiments.

Introduction and Motivation: Learning

- **Training** the model means to use the data available to optimize some parameters of the model with respect to a utility function that evaluates how well the model predicts the training data (e.g., a maximum of some desired performance measure).
- However, in practice, we are interested in the model to perform well on unseen data.
- Performing well on data that we have already seen (training data) may only mean that we found a good way to memorize the data. However, this may not generalize well to unseen data, and, in practical applications, we often need to expose our machine learning system to situations that it has not encountered before.

Introduction and Motivation: Summary

- We represent data as vectors.
- We choose an appropriate model, either using the probabilistic or optimization view.
- We learn from available data by using numerical optimization methods with the aim that the model performs well on data not used for training.

Questions?

References

[1] This lecture slide is mainly based upon `https://yung-web.github.io/home/courses/mathml.html` (made by Prof. Yung Yi, KAIST EE)

[2] Marc Peter Deisenroth, A. Aldo Faisal, and Cheng Soon Ong. Mathematics for machine learning. Cambridge University Press, 2020.