清華大學
Tsinghua University

Reading Materials for Machine Learning Theory

Angelica Aviles-Rivero
aviles-rivero@tsinghua.edu.cn

Version: Oct 31st, 2024

What to Read? Key Readings for Machine Learning Theory

Machine learning is the study of algorithms that allow machines to learn patterns and make predictions based on data. Formally, it explores how a system can generalise from observed samples to unseen instances, minimising error whilst balancing constraints such as computational complexity and data availability. In essence, **learning** is a mathematical process that seeks to identify a hypothesis function from a class of possible functions that performs well on unseen data.

The goal of machine learning is to **approximate an unknown function** that maps inputs to outputs, given a finite set of observations. For example, the task of predicting whether an email is spam can be described mathematically: we seek a function $h \in \mathcal{H}$, from a hypothesis class $\mathcal{H}$, such that $h(x) \approx y$ for input-output pairs (x, y) drawn from an unknown probability distribution D. The challenge is to identify a function that minimises error on new, unseen samples—not just the observed data. Central to the theory of learning is the concept of **generalisation**: how well a hypothesis learned from a training set applies to unseen instances. This requires us to formalise notions of risk, such as the expected error (or generalisation error), and develop algorithms that can efficiently search the hypothesis space to minimise it.

However, learning comes with fundamental limitations. **Overfitting**, for instance, occurs when a model performs well on the training data but poorly on unseen data. This leads us to trade-offs between model complexity and accuracy, captured mathematically through frameworks like Empirical Risk Minimisation (ERM), Structural Risk Minimisation (SRM), and regularisation techniques.

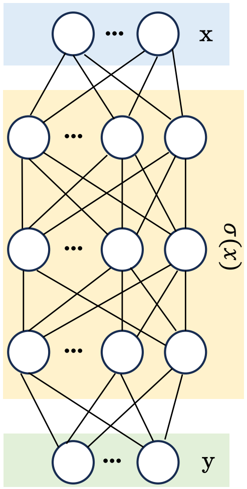
What to Expect from This Course

In this course, we will delve deeply into the mathematical underpinnings of machine learning. Our goal is not just to implement algorithms but to rigorously understand

why they work and when they are expected to fail. You will engage with mathematical concepts such as probability theory, optimisation, and linear algebra, as they are essential tools for analysing learning algorithms. Here is an outline of what you can expect:

- **Formal Foundations of Learning:** We will begin with the Probably Approximately Correct (PAC) learning framework to define what it means for a machine to "learn". You will also explore the VC-dimension, which quantifies the capacity of a hypothesis class to fit data, and understand how it relates to generalisation.
- **Algorithm Design and Analysis:** The course aims to revisit key algorithms such as linear regression, support vector machines (SVMs), and neural networks, framed as solutions to optimisation problems. We seek to study the bias-variance trade-off, a crucial concept that connects statistical learning theory with practical performance.
- **Computational Complexity and Feasibility:** Not all learning problems are solvable in a reasonable amount of time. We will explore the computational hardness of learning problems, discussing results such as the No-Free-Lunch theorem and analysing when certain tasks become intractable.

Activation Function — $\sigma(x)$



- **Optimisation Techniques:** A significant part of machine learning revolves around solving optimisation problems efficiently. We aim to investigate methods such as stochastic gradient descent (SGD), convex optimisation, and explore their convergence properties.
- **Robust Learning and Regularisation:** We seek to learn techniques to ensure that models generalise well, such as ridge regression, Tikhonov regularisation, and stability-based learning. These methods mitigate overfitting by balancing the complexity of the model with the available data.
- **Learning from Complex Data:** The course will extend beyond standard supervised learning, unsupervised learning (such as clustering), and dimensionality reduction techniques like Principal Component Analysis (PCA) and random projections.

Reading Materials

This course is supported by several core texts that provide the essential theoretical and practical knowledge required for this course on machine learning theory. In addition to the main readings, extra materials will be recommended throughout the course for those who are curious to explore advanced topics and gain deeper insights. Below is a brief description of the primary materials we will use, along with supplementary texts available for further study.

Main Materials

★Understanding Machine Learning: From Theory to Algorithms by (Shalev-Shwartz and Ben-David, 2014). This book offers a well-rounded introduction to the theory and algorithms that form the foundation of modern machine learning. It begins by addressing fundamental questions about learning, including how learning can be formally defined and under what conditions it succeeds. The text delves into key concepts such as PAC learning, VC-dimension, and empirical risk minimisation, which underpin the theoretical side of the field. Alongside theory, the book introduces practical algorithms such as linear models, neural networks, and support vector machines, and explores optimisation methods like gradient descent and regularisation. Special attention is given to challenges such as overfitting, model selection, and evaluation strategies, equipping readers with the tools needed to build effective models.
📖 The authors of that (Shalev-Shwartz and Ben-David, 2014) provide a copy for personal use, as indicated by the authors, at the following link.

★Convex Optimisation: Algorithms and Complexity by (Bubeck et al., 2015). This material offers a comprehensive introduction to convex optimisation with a particular focus on algorithms and their complexities. It begins by addressing fundamental aspects of convexity, such as the properties of convex functions and sets, and explains why convexity plays a central role in optimisation. The text also seeks to cover essential algorithms, including gradient descent, cutting plane methods, and stochastic optimisation, highlighting their convergence rates and computational feasibility. With a strong emphasis on both the theoretical underpinnings and practical implementation, this text is particularly valuable for optimisation and machine learning. The structured presentation makes it a great resource for the course, providing deeper insights into optimisation techniques critical for machine learning models.
📖 The pre-print version of this material is available online in this link.

★Practical Implementation Resources. Whilst this course focuses on machine learning theory, there will also be a practical component to reinforce your understanding. For this, homework exercises will primarily involve **Python-based implementations**. We encourage you to explore and become familiar with Python's scikit-learn library, as it will be the main tool used for practical assignments.

🌐 For reference and guidance, please use the Scikit-learn project's official documentation in here.

Supplementary Materials

★Fit without fear: remarkable mathematical phenomena of deep learning through the prism of interpolation by (Belkin, 2021). It explores foundational mathematical concepts related to deep learning, with particular focus on interpolation and over-parameterisation. The work is an attempt to bridge the gap between the theoretical underpinnings and practical success of deep learning models, which have outpaced traditional learning theory.
📖 The pre-print version of this material is available online in this link.

★High-dimensional probability: An introduction with applications in data science by Vershynin (2018). This supplementary material reading provides a rigorous exploration of probability theory specifically tailored for high-dimensional contexts, which are essential for understanding the theoretical foundations of machine learning. The book introduces essential concepts like concentration inequalities, random matrices, high-dimensional distributions, and random projections—all of which are central to the mathematical understanding of machine learning algorithms, particularly in high-dimensional and over-parameterised models. These topics align well with understanding how machine learning models generalise, optimise, and handle high-dimensional data, making it a valuable resource for students in this machine learning theory course.
📖 You can access a free draft of this material here, *provided it is used only for personal and classroom needs as indicated by the author.*

Bibliography

Mikhail Belkin. Fit without fear: remarkable mathematical phenomena of deep learning through the prism of interpolation. *Acta Numerica*, 30:203–248, 2021. 6

Sébastien Bubeck et al. Convex optimization: Algorithms and complexity. *Foundations and Trends® in Machine Learning*, 2015. 5

Shai Shalev-Shwartz and Shai Ben-David. *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014. 5

Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018. 6