Chapter 1

What is Machine Learning?

An ability to learn is one of the most prominent and intrinsic skills found in intelligent organisms. All of them have had to constantly adapt to changing environment since the day they were born to ensure their survival. It is hard to imagine a living organism that lacks this ability.

Consider a learning process of a newborn baby. First of all, the baby must learn to communicate, at least to her parent, to express what she wants. This may take a form of non-verbal expressions. To communicate more effectively, the baby must learns to speak. Learning to speak usually begins by learning to associate between simple words given by the parents and objects in the real world. Because the parent provided good examples, this form of learning is almost always the fastest learning process. However, the baby will sometimes have to rely on her own sense organs, for example, when learning to distinguish people or objects. Lastly, many tasks require trial-and-error learning. When the baby learn to walk, she will not succeed in a single trial, but will get better over time. As we can see, learning can take several forms. To read and understand this sentence is in part due to your ability to learn from past experience.

It is hard to believe that the baby has to learn everything from scratch. Seemingly, our brain has been engineered in a way that allows us to learn much faster than other animals. Hence, we may categorize the learning process into *long-term learning* and *short-term learning* depending on a time scale at which it happens. In long-term learning, a learning process happens at a large time scale and is beyonds our life expectency. The experience is encoded in our genes through the evolutionary process. On the other hand, short-term learning involves the process through which we experience things throughout our life.

Machine learning (ML) is a subfield of artificial intelligence (AI). It aims to equip computers with the ability to learn from past experience using math-

ematical, statistical, and computational techniques. Specifically, the computers have access to the past experience in the form of *empirical data*. Having access to an unprecedented amount of data provides challenging research frontiers and gives rise to novel applications. In this course, we will cover a wide range of successful ML techniques including mathematical foundations and examples of their uses in practice.

1.1 Real-world Applications

In this section, we look at some real-world applications of machine learning. Since machine learning can be generally defined as computational methods that make use of empirical data, it has been applied in a wide range of data intensive domains.

- *Computer vision*: Can we teach computers to see and understand images the way we do? Available data can be a still image, a collection of images, or videos. Important computer vision tasks are image recognition, object detection, and face detection. One of the most influential applications of computer vision is a pedestrian recognition used in self-driving cars.

- *Natural language processing*: If computers could understand our languages, it would have been easier for us to communicate with them and get things done much more effectively. In NLP, data may be collected from text messages, emails, e-books. Important problems in NLP are part-of-speech tagging, morphological analysis, parsing, and named-entity recognition. An important application is a language translation.

- *Speech recognition*: This involves teaching computers to understand what we speak and is important in the development of personal assistants like Siri. Important tasks in speech recognition include speech synthesis, speech recognition, and speaker verification.

- *Social network analysis*: Given a data in the form of a social network, can we figure out hidden relationships between any pairs of entities in the network? Online companies like Facebook and LinkedIn have been using machine learning techniques to help suggest *people you may know*.

- *Personalization*: Wearable devices and smart phones have become very common. These devices have generated tremendous amount of data about each individual. Based on personalized data, machine learning

techniques can be used to individually track activities and predict potential health risks. In medicine, personalized data has been used to design drugs that work specifically for each individual.

- *Recommendation systems*: Based on previous purchases by the customers, online commercial stores like Amazon and Alibaba have been using machine learning techniques to predict which products the customers may like. Hence, recommending such products will likely lead to more purchases in the future.

This list is by no means exhausive. In principle, machine learning can be applied in any domain in which a collection of data is available.

1.2 Methematical Definition

Scenarios given in Section 1.1 illustrate real-world applications of machine learning. In this section, we give a formal definition of machine learning. Roughly speaking, learning problems consists of these three main ingredients.

1. A **data set**, which we denote by D, consists of data points. The data points $\{z_1, z_2, \ldots, z_n\}$ where z_i are distributed according to some unknown distribution $P(Z)$. The data points are reminiscent of past experiences from which the learner will use to gain valuable insights about $P(Z)$. It is the most important ingredient of any machine learning problems. Data can come in different types such as texts, sequences, and images depending on applications.

2. A **model class**, which we denote by $\mathcal{M}$, consists of all possible solutions of one specific learning problem. It may be a finite, countably infinite, or even uncountable set of solutions. The choice of model class will depend largely on the nature of learning problem at hand and sometimes on the personal preference of domain experts. For example, one of the simplest model classes is the space of linear hyperplane in $\mathbb{R}^d$
$$\mathcal{M} = \left\{ f(z) = w^\top z \,|\, w \in \mathbb{R}^d \right\}.$$
An example of a more complicate model class is a *deep neural network*.

3. A **learner**, commonly called a *learning algorithm*, denoted by $F_{\boldsymbol{\theta}}$ is a strategy by which we pick the solution h from $\mathcal{M}$. It depends on a vector of hyper-parameters $\boldsymbol{\theta}$ which can be fine tuned. The learner

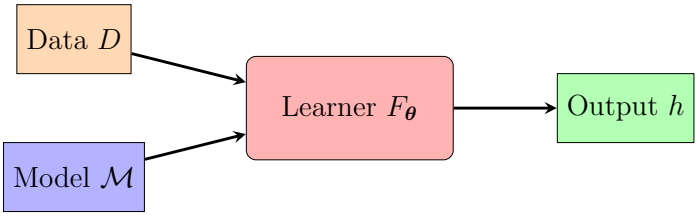

Figure 1.1: A simplified diagram of learning problem.

can be either deterministic or stochastic. Well-known learning algorithms include empirical risk minimization (ERM) and structural risk minimization (SRM), for example.

Hence, a learning problem is a triple $(D, \mathcal{M}, F_\theta)$ and can be defined by the following simple equation.

$$h = F_{\boldsymbol{\theta}}(D, \mathcal{M}). \tag{1.1}$$

An illustrative explanation of (1.1) is also given in Figure 1.1. In words, the learner $F_{\boldsymbol{\theta}}$ takes the data D and the model class $\mathcal{M}$ as inputs and systemetically finds the solution h from $\mathcal{M}$. An ultimate goal of machine learning is to design the learner $F_{\boldsymbol{\theta}}$ such that it learns the best possible output h with respect to some pre-specified criteria. Unfortunately, we will encounter several issues in practice that prevent us from accomplishing such goal. These issues give rise to challenges that form the core of research frontier in machine learning today. We discuss some of them here.

- *Small data*: In some applications, a data collection process can be expensive. Hence, the data that is actually available for learning are scarce.
- *Big data*: Too much data creates a problem too. It is less memory efficient and more computational expentive to deal with huge amount of data. That is, the more data we obtain, the longer the the learning algorithms take to learn from the data. Moreover, it is also challenging to preserve the privacy for a big personalized data.
- *Computational cost*: In certain situations, the complexity of the learning problem may call for a big model class which increases the computational cost and slows down the learning process.
- *The curse of dimensionality*: Some data sets can be explained using several number of modularities. It is generally very difficult to find

patterns and structures in data embedded in a high dimensional space. Moreover, in discrete domain, the sample size grows exponentially in the number of dimensions.

- *etc.*

In summary, machine learning techniques can be understood more broadly as data-driven methods fundamental concepts in mathematics, statistics, and computer science.

1.3 Types of Learning Problems

According to (1.1), we can roughly categorize learning problems into several categories depending on the types of data sets available to the learner, how they are presented to the learner, and how the learning algorithms are evaluated at the end.

1. *Supervised learning*
2. *Unsupervised learning*
3. *Semi-supervised learning*
4. *Reinforcement learning*
5. *Active learning*
6. *Online learning*
7. *Multi-arm bandit*

1.4 Exercises

1. What is machine learning?
2. What are three main ingredients of a learning problem? Explain.