Chapter 1. Introduction

Yang Lin

September 4, 2018

1 Machine learning: what and why?

We define **machine learning** as a set of methods that can automatically detect patterns in data, and then use the uncovered patterns to predict future data, or to perform other kinds of decision making under uncertainty (such as planning how to collect more data!).

This books adopts the view that the best way to solve such problems is to use the tools of probability theory. The probabilistic approach to machine learning is closely related to the field of statistics, but differs slightly in terms of its emphasis and terminology. In the footnote, Murphy mentioned Rob Tibshirani's work, which is demonstrated as follow:

1.1 Types of machine learning

Machine learning is usually divided into two main types — **predictive** or **supervised learning**, and **descriptive** or **unsupervised learning**.

In the supervised learning approach, the goal is to learn a mapping from inputs $\mathbf{x}$ to outputs y, given a labeled set of input-output pairs $\mathcal{D} = (\mathbf{x}_i, y_i)_{i=1}^{N}$. Here $\mathcal{D}$ is called the **training set**, and N is the number of training examples.

Glossary

Machine learning	Statistics
network, graphs	model
weights	parameters
learning	fitting
generalization	test set performance
supervised learning	regression/classification
unsupervised learning	density estimation, clustering

Table 1: Rob Tibshirani, a statistician at Stanford University, has created an amusing comparison between machine learning and statistics, available as `http://www-stat.stanford.edu/ tibs/stat315a/glossary.pdf`.

In the simplest setting, each training input $\mathbf{x}_i$ is a D-dimensional vector of numbers, which are called **features**, **attributes** or **covariates**. Similarly the form of output can in principle be anything, but most methods assume that y_i is a categorical or nominal variable from some finite set ($y_i \in \{1, ..., C\}$), or that y_i is a real-valued scalar. When y_i is categorical, the problem is known as **classification** or **pattern recognition**, and when y_i is real-valued, the problem is known as **regression**.

In the unsupervised learning approach, we are only given inputs, $\mathcal{D} = (\mathbf{x}_i)_{i=1}^{N}$, and the goal is to find "interesting pattern" in the data. This is sometimes called knowledge discovery.

There is a third type of machine learning, known as reinforcement learning, which is somewhat less commonly used. This is useful for learning how to act or behave when given occasional reward or punishment signals.

2 Supervised learning

Supervised learning is the form of ML most widely used in practice.

2.1 Classification

In the case of classification, the goal is to learn a mapping from inputs $\mathbf{x}$ to outputs y, where $y \in \{1, ..., C\}$, with C being the number of classes. According to the value of C, we divided it into two categories:

- If $C = 2$, this is called **binary classification**, in which case we often assume $y \in \{0, 1\}$.
- If $C > 2$, this is called **multiclass classification**, in which case we often assume $y \in \{1, ..., C\}$.

We will denote the probabilities over possible labels, given the input vector $\mathbf{x}$ and training set $\mathcal{D}$ by $p(y|\mathbf{x}, \mathcal{D})$. In general, this represents a vector of length C, whose elements are probabilities of the classes.

Regression is just like classification except the response variable is continuous.

3 Unsupervised learning

In the **unsupervised learning** approach, we are just given output data, without any inputs. The goal is to discover "interesting structure" in the data; this is sometimes called **knowledge discovery**. We formalize our task as one of density estimation, that is, we want to build our models of the form of $p(\mathbf{x}_i|\boldsymbol{\theta})$ instead of $p(y_i|\mathbf{x}_i, \boldsymbol{\theta})$. That is, supervised learning is *conditional density estimation*, whereas unsupervised learning is *unconditional density estimation*.

Unsupervised learning is arguably more typical of human and animal learning. It is also more widely applicable than supervised learning, since it does not require a human expert to manually label data.

4 Some basic concepts in machine learning

5 Exercises