Note: Machine Learning a Probabilistic Perspective [1] by Murphy

Zheng Daqi

January 7, 2013

Abstract

This is a note taken when I read the book

1 Reading Plan and Progress

- Contents – Done
- Chapter 1 – today 3:00–5:00 pm
- Every Chapters' Summary – today

2 Notation Explanation

- Most of the contents are the direct quotation from the book, so without extra notation.
- Sections started with "My" are my notes.
- Items marks by "TODO" are my notes.
- Anything braced in square braces [my notes] are my notes.

3 Concepts (Definition)

Machine Learning 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!).

3.1 Types of Machine Learning 3

Supervised Learning (predictive)

Goal the goal is to learn a mapping from inputs $\mathbf{x}$ to outputs y. Our main goal is to make predictions on novel inputs, meaning ones that we have not seen before (this is called generalization), since predicting the response on the training set is easy (we can just look up the answer).

Given given a labeled set of input-output pairs $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^N$

Unsupervised Learning (descriptive, knowledge discovery)

Goal the goal is to find "interesting patterns" in the data.

Given only given inputs $\mathcal{D} = \{\mathbf{x}_i\}_{i=1}^N$

Reinforcement Learning (not included in this book) This is useful for learning how to act or behave when given occasional reward or punishment signals. (For example, consider how a baby learns to walk.)

3.2 Tasks in Supervised Learning 3.1

Classification Here the goal is to learn a mapping from inputs $\mathbf{x}$ to outputs y,where y $1, ..., C$, with C being the number of classes.

binary classification $C = 2$

multiclass classification $C > 2$

multi-label classification If the class labels are not mutually exclusive (e.g., somebody may be classified as tall and strong), we call itmulti-label classification, but this is best viewed as predicting multiple related binary class labels (a so-called multiple output model).

One way to formalize the problem is as function approximation.

3.3 TODO

Problem Formalization

Exploratory Data Analysis

4 Phenomenon

Long Tail In fact, data across a variety of domains exhibits a property known as thelong tail, which means that a few things (e.g., words) are very common, but most things are quite rare (see Section 2.4.6 for details).

5 Attitude

Problem involving uncertainty This books adopts the view that the best way to solve such problems is to use the tools of probability theory. Probability theory can be applied to any problem involving uncertainty.

Generalizing from small data This 4 means that the core statistical issues that we discuss in this book, concerning generalizing from relatively small samples sizes, are still very relevant even in the big data era.

Feature Extraction Such feature extraction is an important, but difficult, task. Most machine learning methods use features chosen by some human. Later we will discuss some methods that can learn good features from the data.

6 Experience

Exploratory Data Analysis 3.3 It is always a good idea to perform exploratory data analysis, such as plotting the data, before applying a machine learning method.

7 My Experience

7.1 What to Plot

Scatter Plot Visualize the pair-wised variables relationship.

8 My Idea TODO

Re-Plot the 20 newsgroup classification in Figure 1.2 sort the words in different order.

- total frequency
- other metric

References

[1] Kevin P. Murphy. *Machine Learning: A Probabilistic Perspective.* 2012. 1