Sketch of Chapter 1

DuLi

2019 年 3 月 13 日

目录

1 Definition of Machine Learning

1.1 Arthur Samuel's Definition

Machine Learning is a field of study that gives computers the ability to learn without being explicitly programmed.

这个定义看来是非常general了，基本上是从字面上解释了这个词组。稍微提一下Arthur Samuel这个人，直接把Wikipedia的简介搬过来。大体就是上古大神的意思了。

图 1: Arthur Samuel

Arthur Lee Samuel (December 5, 1901 – July 29, 1990)was an American pioneer in the field of computer gaming and artificial intelligence.He coined the term "machine learning" in 1959.The Samuel Checkers-playing Program was among the world's first successful self-learning programs, and as such a very early demonstration of the fundamental concept of artificial intelligence (AI).He was also a senior member in the TeX community who devoted much time giving personal attention to the needs of users and wrote an early TeX manual in 1983.

1.2 Tom Michell's definition

A computer program is said to learn from experience E with respect to some task T and some performance measure P,if its performance on T,as measured by P,improves with experience E.

乍一看这个写的比较拗口，但实际上也就是说通过对experience E的学习，使原有task T的performance P有了提高。对于Tom Michell的了解大概是因为那本机器学习的教材，薄薄一本，当时应该是看过同学的，没有看太多，也不好评价。这个定义就显得更像在描述一个事情而非定义。本书接下来用一个简单的例子来说明了一下，那就是spam filter，并将Tom Michell的定义套用做了讲解。

图 2: Tom Michell

2 Why use Machine Learning

依旧以写一个spam filter为例，当使用传统的编程技术时，需要一个长长的规则清单，并且这个规则是难以把握且复杂的。相反，当使用Machine Learning Techniques时，可以让机器自主的学习知识，学习垃圾邮件中的词频，这显然更加简洁高效。To summarize,Machine Learning is great for:

1. Problems require a lot of hand-tuning or long list of rules.
2. Complex problems is no good solution using traditional approach.
3. Fluctuating enviornment
4. Getting insights about complex problems and large amount of data

2.1 Types of Machine Learning Systems

- Whether or not they are trained with human supervision(Supervised Unsupervised Semisupervised Reinforcement Learning 监督非监督半监督强化学习)

- Whether or not they can learn incrementally(Online versus Batch learning 在线学习 VS 批量学习)

- Whether work by comparing new data to known data or instead detect patterns in the training data and build a predictive model(instance-based versus model-based learning 基于模型 vs 基于实例)

3 Supervised/Unsupervised Learning

3.1 Supervised Learning

The training data you feed to the algorithm includes the desire solutions,called *labels*.Here are some most important supervised learning algorithms(在后文中都有体现):

- k-Nearest Neighbors

- Linear Regression

- Logistic Regression

- Support Vector Machines

- Decision Trees and Random Forests

- Neural networks

3.2 Unsupervised Learning

The training data is unlabeled.Some important unsupervised learning algorithms:

- Clustering

- k-Means

- Hierarchical Cluster Analysis(HCA)

- Expection Maximization

• Visualization and dimensionality reduction

- Principle Component Analysis(PCA)

- Kernel PCA

- Locally-Linear Embedding(LLE)

- t-distributed Stochastic Neighbor Embedding(t-SNE)

• Associaton rule learning

- Apriori

- Eclat

3.3 Reinforcement Learning