Hands On Machine Learning Notes

Notes based on the Book "Hands-On Machine Learning with Scikit-Learn & TensorFlow" by Aurelien Geron

Krish Katariya

Last updated: **January 31, 2024**

Contents

1. Types of Machine Learning systems

There are many different Types of ML systems that it is often useful to classify them in broad categories based on :

- Whether or not they are trained with human supervision (supervised, unsupervised, reinforcement learning)
- Whether or not they can learn incrementally on the fly (online vs batch learning)
- Whether they work by simply comparing new data points to old data points, or instead detect patterns in training data

1.1. (Un)Supervised Learning

- Machine learning systems can be classified according to the amount and type of supervision they get during training. There are four major categories: **supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning**

Definition 1.1.1

In **supervised learning**, the training data you feed to the algorithm includes the desired solution, called **labels**

A typical supervised learning task is **classification**. The spam filter is an example of this, where models are filled with lots of emails with their class(label) of spam/normal emails.

Another task can be predict a target numeric value, such as the price of a car, known as *predictors*. This task is often called **regression**.

- One example of regression is actually **logistic regression**, which is used for classification. Basically, it outputs a certain probability than an object belongs to a class.

Example 1.1.1

Important Supervised Learning Algorithms:

- k-Nearest Neighbors
- Linear Regression
- Logistic Regression
- Support Vector Machines (SVM)
- Decision Trees and Random Forests
- Neural Networks (*can also be unsupervised*)

Definition 1.1.2

In **unsupervised learning** the training data is unlabeled. In this case, the system tries to learn without having a teacher.

Example 1.1.2

Important Unsupervised Learning Algorithms

- Clustering
  - K Means
  - DBSCAN
  - Hierarchical Cluster Analysis (HCA)
- Anomaly detection and novelty detection
  - One class SVM
  - Isolation Forests
- Visualization and dimensionality reduction
  - Principal Component Analysis (PCA)
  - Kernel PCA
  - Local Linear Embedding
  - t-distributed Stochastic Neighbor Embedding
- Association rule learning
  - Apriori
  - Eclat

For example, say you have a lot of data about your blog's visitors. You can run an unsupervised model to group them without ever giving info about what groups they are in.

1.2. Semisupervised Learning

Definition 1.2.1

Some algorithms can deal with partially labeled training data, usually a lot of unlabeled data and a little bit of labeled data. A good example is google photos when it detects some of your pictures faces to get the rest.

1.3. Reinforcement Learning

Definition 1.3.1

Reinforcement Learning has a learning system called the *agent* that observes the environment, selects and performs actions, and gets rewards/penalties based on its actions. It then find its best strategy, called its *policy*, to get the most reward.

1.4. Batch Learning

Definition 1.4.1

In **batch learning**, the system is incapable of learning incrementally: it must be trained using all the available data. This often takes a huge amount of time and lots of computational resources, so it is often done offline, then once fully done it can be used, This is called *offline learning*.

While simple, this can take many hours of training and is not often ideal, also requiring many computing hours.

Definition 1.4.2

In **online learning**, you train the system incrementally by feeding it data instances sequentially, in "mini batches". This is especially useful when you have a continuous flow of new information.

These systems are especially good because they don't take much computation power since you don't need to retrain many times.

Note 1.4.1

Stopped at top of page 17 of book