

Machine Learning: A Probabilistic Perspective

immediate

December 28, 2018

1 Introduction

1.1 Types of machine learning

Predictive or Supervised Learning The goal is to learn a mapping from inputs to outputs given a labeled set of input-output pairs.

Descriptive or Unsupervised Learning The goal is to find interesting patterns in a given dataset.

1.2 Supervised Learning

Classification Learning a mapping from inputs x to outputs y where $y \in \{1, \dots, C\}$, with C being the number of classes. Classification can be formalized as **function approximation**. Assume $y = f(x)$ for some unknown function f . We estimate f given a labeled training set, and make predictions from the estimated function, which we denote $\hat{f}(x)$.

Binary Classification Where the number of classes is 2.

Multiclass Classification Where the number of classes is greater than 2.

Multi-label classification Where each instance may belong to multiple classes.

Generalization Making predictions on inputs not in the training set.

Probabilistic Predictions Given a model that outputs a probability for each class, we can make our "best guess" for the class of the instance by returning the most probable class label; the mode of the distribution. This is called the **MAP Prediction** or **Maximum A Posteriori**, meaning that the class has the highest likelihood given the instance.

$$\hat{y} = \hat{f}(x) = \operatorname{argmax}_c p(y = c|x, D) \quad (1)$$

where D is the training dataset.

Real World Applications of Classification

- Document Classification
- Spam filtering
- Image Classification
- Handwriting Recognition
- Face detection and Recognition

Regression Like classification except that the response variable you are trying to predict is continuous.

1.3 Unsupervised Learning

The goal is to discover *structure* in the data; this is sometimes called *knowledge discovery*. Can be formalized as **density estimation**, meaning building models that return a probability density for each input: $p(x_i|\theta)$. **Supervised and unsupervised learning can be distinguished as conditional and unconditional density estimation, respectively.**