

Data Science from Scratch

First Principles with Python

Stephen CUI 

March 20, 2023

Contents

1	Machine Learning	1
1.1	Modeling	1
1.2	What Is Machine Learning?	1

Chapter 1

Machine Learning

Data science is mostly turning business problems into data problems and collecting data and understanding data and cleaning data and formatting data, after which machine learning is almost an afterthought.

1.1 Modeling

What is a model? It's simply a specification of a mathematical (or probabilistic) relationship that exists between different variables.

1.2 What Is Machine Learning?

Everyone has her own exact definition, but we'll use machine learning to refer to creating and using models that are learned from data. In other contexts this might be called predictive modeling or data mining, but we will stick with machine learning.

We'll look at both supervised models (in which there is a set of data labeled with the correct answers to learn from) and unsupervised models (in which there are no such labels). There are various other types, like semisupervised (in which only some of the data are labeled), online (in which the model needs to continuously adjust to newly arriving data), and reinforcement (in which, after making a series of predictions, the model gets a signal indicating how well it did) that we won't cover here.