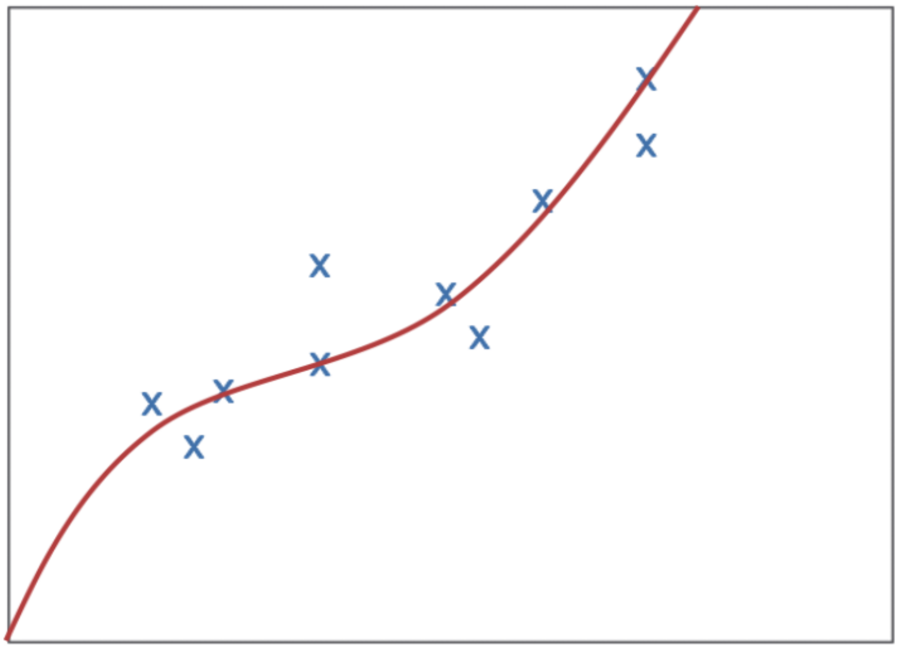
Read: Introduction to Statistical Learning 2.1–2.2 Basics of Learning



Our first reading is from Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani, *An Introduction to Statistical Learning*, Springer 2013. You can access a PDF of the book **[at the authors' website](https://static1.squarespace.com/static/5ff2adbe3fe4fe33db902812/t/6009dd9fa7bc363aa822d2c7/1611259312432/ISLR+Seventh+Printing.pdf)** **(https://static1.squarespace.com/static/5ff2adbe3fe4fe33db902812/t/6009dd9fa7bc363aa822d2c7/1611259312432/ISLR+Seventh+Printing.pdf)**. We will be reading the first chapter, along with the first two sections of the second (pages 1–42).

This text provides a good overview of many standard machine learning techniques. While they use code samples in R, and we will use Python, the text is still very useful, and we will just skip over the implementation sections.

Main topics

The book begins by introducing the basic idea of learning from data, before moving on to specific ML models:

- The first chapter introduces a variety of ways in which data may be represented, covering such ideas as the difference between *discrete* and *continuous* variables, both for input and output. This will be very important throughout this class, especially for output variables, since the type of ML approach we use—regression or classification—is often determined by the type of output we have.

- The first chapter also discusses the differences between cases in which we have training data—the information we use to build our ML models—consisting of *both* known input and output values, and cases in which we can observe input values only. The first part of this course will focus on the first case, investigating what is known as *supervised* learning; in the later parts of the course we will move to the second case, when we investigate *unsupervised* learning.

- Chapter 2 introduces the basic formulation of a learning problem in terms of estimating a function, a key concept in this course. It also discusses *parametric* and *non-parametric* methods. Much of what we do in this course concerns the first of these, in which the job of an ML model is to find a set of numerical *parameters* (also called *weights*), which are then applied to the numbers that make up input values, in order to compute some output function. In this context, the role of an ML algorithm is to adjust those parameters/weights to give us a good result. (In the later parts of the course, we will look at non-parametric models, which work without such numerical parameters.)

- Chapter 2 also introduces the key notion of *model fit*. When we do ML, we are always guided by some measure of success or failure. We are trying, always, to come up with models of data that get things right. As it turns out, there are many ways of measuring model fit, and many ways of determining whether what we are doing is working. This is something we will come back to many times in this course.

Where to focus (or not)

- As you read, some main things to think about:
  - For each type of learning, like regression/classification or supervised/unsupervised, note the features of problem and data that distinguish the approaches we take.
  - As you read about linear models, make sure you understand what exactly defines such a model.
  - There are a lot of ways of measuring how well a model fits data. As you read about them, note what each measures exactly and think about when you might use one of them, versus another.

- **One note:** this course is not meant to be exceedingly formal. We will dive into the mathematics of various algorithms and models enough so that we can develop intuitions for how those models work, but we will not be engaged in the business of formal proof. As you go through the readings, if the mathematical content is too hard to follow in detail, don't fret too much. Try to understand the *results* that the math is used to describe, or to prove, and make sure you understand the overall point of what you're reading, even if you don't get every little bit (and of course if you *do*, then that's great!). In this text, a lot of matrix algebra is used, but in the course we won't be dealing with that very much. Thus, you might want to skim (or skip) the section on matrices (pages 9–12) and come back to it later if you find it necessary.

- **Another note**: we have not assigned the final sections of the chapter. These concern implementations of some of the ideas covered, using R, and some exercises. As we have said, we are not using R in this course, but if you know it, you may find those sections interesting. Likewise, the exercises are there if you want to pursue them, but they are not expected to be completed. In general, if you wish to read deeper and further into any of the books we will use, you are more than welcome to do so. Reading is fun!