1

Introduction

We rely on experience to shape our view of the unknown, with the notable exception of religion. But for most practical purposes we lean on experience to guide us through an uncertain world. We process experiences both naturally and statistically; however, the way we naturally process experiences often diverges from the methods that classical statistics prescribes. Our purpose in writing this book is to reorient common statistical thinking to accord with our natural instincts.

Let us first consider how we naturally process experience. We record experiences as narratives, and we store these narratives in our memory or in written form. Then when we are called upon to decide under uncertainty, we recall past experiences that resemble present circumstances, and we predict that what will happen now will be like what happened following similar past experiences. Moreover, we instinctively focus more on past experiences that were exceptional rather than ordinary because they reside more prominently in our memory.

Now, consider how classical statistics advises us to process experience. It tells us to record experiences not as narratives, but as data. It suggests that we form decisions from as many observations as we can assemble or from a subset of recent observations, rather than focus on

observations that are like current circumstances. And it advises us to view unusual observations with skepticism. To summarize:

Natural Process

- Records experiences as narratives.
- Focuses on experiences that are like current circumstances.
- Focuses on experiences that are unusual.

Classical Statistics

- Record experiences as data.
- Include observations irrespective of their similarity to current circumstances.
- Treat unusual observations with skepticism.

The advantage of the natural process is that it is intuitive and sensible. The advantage of classical statistics is that by recording experiences as data we can analyze experiences more rigorously and efficiently than would be allowed by narratives. Our purpose is to reconcile classical statistics with our natural process in a way that secures the advantages of both approaches.

We accomplish this reconciliation by shifting the focus of prediction away from the selection of variables to the selection of observations. As part of this shift in focus from variables to observations, we discard the term *variable*. Instead, we use the word *attribute* to refer to an independent variable (something we use to predict) and the word *outcome* to refer to a dependent variable (something we want to predict). Our purpose is to induce you to think foremost of experiences, which we refer to as observations, and less so of the attributes and outcomes we use to measure those experiences. This shift in focus from variables to observations does not mean we undervalue the importance of choosing the right variables. We accept its importance. We contend, however, that the choice of variables has commanded disproportionately more attention than the choice of observations. We hope to show that by choosing observations as carefully as we choose variables, we can use data to greater effect.

Relevance

The underlying premise of this book is that some observations are relevant, and some are not—a distinction that we argue receives far

less attention than it deserves. Moreover, of those that are relevant, some observations are more relevant than others. By separating relevant observations from those that are not, and by measuring the comparative relevance of observations, we can use data more effectively to guide our decisions. As suggested by our discussion thus far, relevance has two components: similarity and unusualness. We formally refer to the latter as informativeness. This component of relevance is less intuitive than similarity but is perhaps more foundational to our notion of relevance; therefore, we tackle it first.

Informativeness

Informativeness is related to information theory, the creation of Claude Shannon, arguably the greatest genius of the twentieth century.[1] As we discuss in Chapter 2, information theory posits that information is inversely related to probability. In other words, observations that are unusual contain more information than those that are common. We could stop here and rest on Shannon's formidable reputation to validate our inclusion of informativeness as one of the two components of relevance. But it never hurts to appeal to intuition. Therefore, let us consider the following example.

Suppose we would like to measure the relationship between the performance of the stock market and a collection of economic attributes (think variables) such as inflation, interest rates, energy prices, and economic growth. Our initial thought might be to examine how stock returns covary with changes in these attributes. If these economic attributes behaved in an ordinary way, it would be difficult to tell which of the attributes were driving stock returns or even if the performance of the stock market was instead responding to hidden forces. However, if one of the attributes behaved in an unusual way, and the stock market return we observed was also notable, we might suspect that these two occurrences are linked by more than mere coincidence. It could be evidence of a fundamental relationship. We provide a more formal explanation of informativeness in Chapter 2, but for now let us move on to similarity.

[1] Some might prefer to assign this accolade to Albert Einstein, but why quibble? Both were pretty smart.

Similarity

Imagine you are a health care professional charged with treating a patient who has contracted a life-threatening disease. It is your job to decide which treatment to apply among a variety of available treatments. You might consider examining the outcomes of alternative treatments from as large a sample of patients with the same disease as you can find, reasoning that a large sample should produce more reliable guidance than a small sample. Alternatively, you might focus on a subset of the large sample comprising only patients of a similar age, with similar health conditions, and with similar behavior regarding exercise and smoking. The first approach of using as large a sample as possible to evaluate treatments would undoubtedly yield the more robust treatment; that is, the treatment that would help, at least to some extent, the largest number of patients irrespective of each person's specific features. But the second approach of focusing on a targeted subset of similar patients is more likely to identify the most effective treatment for the specific patient under your care.

We contrived these examples to lend intuition to the notions of informativeness and similarity. In most cases, though, informativeness and similarity depend on nuances that we would fail to detect by casual inspection. Moreover, it is important that we combine an observation's informativeness and similarity in proper proportion to determine its relevance. This would be difficult, if not impossible, to do informally.

Fortunately, we have discovered how to measure informativeness, similarity, and therefore relevance, in a mathematically precise way. The recipe for doing so is one of the key insights of this book. However, before we reveal it, we need to establish a new conceptual and mathematical foundation for observing data. By viewing common statistical measures through a new lens, we hope to bring clarity to certain statistical concepts that, although they are commonly accepted, are not always commonly understood. But our purpose is not to present these new statistical concepts merely to enlighten you; rather, we hope to equip you with tools that will enable you to make better predictions.

Roadmap

Here is what awaits you. In Chapter 2, we lay out the foundations of our approach to observing information from data. In Chapter 3, we characterize patterns between multiple attributes. In Chapter 4, we introduce

relevance and show how to use it to form predictions. In Chapter 5, we discuss how to measure confidence in predictions by considering the tradeoff between relevance and noise. In Chapter 6, we apply this new perspective to evaluate the efficacy of prediction models. In Chapter 7, we compare our relevance-based approach to prediction to machine learning algorithms. And lastly, in Chapter 8, we provide biographical sketches of some of the key scientists throughout history who established the theoretical foundation that underpins our notion of relevance.

In each chapter, we first present the material conceptually, leaning heavily on intuition. And we highlight the key takeaways from our conceptual exposition. Then, we present the material again, but this time mathematically. We conclude each chapter with an empirical application of the concepts, which builds upon itself as we progress through the chapters.

If you are strongly disinclined toward mathematics, you can pass by the math and concentrate only on the prose, which is sufficient to convey the key concepts of this book. In fact, you can think of this book as two books: one written in the language of poets and one written in the language of mathematics, although you may conclude we are not very good at poetry.

We expect some readers will view our key insight about relevance skeptically, because it calls into question notions about statistical analysis that are deeply entrenched in beliefs from earlier training. To get the most out of this book, we ask you to suspend these beliefs and give us a chance to convince you of the validity of our counterclassical interpretation of data by appealing to intuition, mathematics, and empirical illustration. We thank you in advance for your forbearance.