1.8 Exercises

1. What is **predictive data analytics**?

2. What is **supervised machine learning**?

3. Machine learning is often referred to as an **ill-posed problem**. What does this mean?

4. The following table lists a dataset from the credit scoring domain we discussed in the chapter. Underneath the table we list two prediction models that are consistent with this dataset, **Model 1** and **Model 2**.

ID	Occupation	Age	Loan-Salary Ratio	Outcome
1	industrial	39	3.40	default
2	industrial	22	4.02	default
3	professional	30	2.7 0	repay
4	professional	27	3.32	default
5	professional	40	2.04	repay
6	professional	50	6.95	default
7	industrial	27	3.00	repay
8	industrial	33	2.60	repay
9	industrial	30	4.5 0	default
10	professional	45	2.78	repay

Model 1

if Loan-Salary Ratio > 3.00 **then**

Outcome = *default*

else

Outcome = *repay*

Model 2

if Age= 50 **then**

Outcome = *default*

else if AGE= 39 **then**

OUTCOME = *default*

else if AGE= 30 **and** OCCUPATION = *industrial* **then**

OUTCOME = *default*

else if AGE= 27 **and** OCCUPATION = *professional* **then**

OUTCOME = *default*

else

OUTCOME = *repay*

a. Which of these two models do you think will generalise better to instances not contained in the dataset?

b. Propose an inductive bias that would enable a machine learning algorithm to make the same preference choice as you did in part (a).

c. Do you think that the model that you rejected in part (a) of this question is overfitting or underfitting the data?

✻ 5. What is meant by the term **inductive bias**?

✻ 6. How do machine learning algorithms deal with the fact that machine learning is an **ill-posed problem**?

✻ 7. What can go wrong when an inappropriate **inductive bias** is used?

✻ 8. It is often said that 80% of the work done on predictive data analytics projects is done in the Business Understanding, Data Understanding, and Data Preparation phases of **CRISP-DM**, and just 20% is spent on the Modeling, Evaluation, and Deployment phases. Why do you think this would be the case?

---

1 Other types of machine learning include **unsupervised learning**, **semi-supervised learning**, and **reinforcement learning**. In this book, however, we focus exclusively on supervised machine learning and use the terms supervised machine learning and machine learning interchangeably.

2 This dataset has been artificially generated for this example. Siddiqi (2005) gives an excellent overview of building predictive data analytics models for financial credit scoring.