PART

DISCRIMINATIVE APPROACH TO STATISTICAL MACHINE LEARNING

4

As discussed in Chapter 11, the problem of statistical pattern recognition is formulated as the problem of estimating the class-posterior probability $p(y|\boldsymbol{x})$. In the generative approach explored in Part 3, the problem of estimating the class-posterior probability $p(y|\boldsymbol{x})$ is replaced with the problem of estimating the joint probability $p(\boldsymbol{x}, y)$ based on the following equality:

$$\underset{y}{\operatorname{argmax}}\, p(y|\boldsymbol{x}) = \underset{y}{\operatorname{argmax}}\, p(\boldsymbol{x}, y).$$

Generative model estimation is the most general approach in statistical machine learning, because knowing the data generating model is equivalent to knowing everything about the data. Therefore, any kind of data analysis is possible through generative model estimation.

When a good parametric model is available, *maximum likelihood estimation* or *Bayesian methods* will be highly useful for generative model estimation (see Part 3). However, without strong prior knowledge on parametric models, estimating the generative model is statistically a hard problem. Non-parametric methods introduced in Chapter 16 could be used if prior knowledge on generative models is not available, but non-parametric methods tend to perform poorly when the dimensionality of data is not small.

In Part 4, an alternative approach to generative model estimation called the *discriminative approach* is explored. In the discriminative approach, the class-posterior probability $p(y|\boldsymbol{x})$ is directly modeled. More specifically, $p(y|\boldsymbol{x})$ is regarded as the sum of a function $f(\boldsymbol{x})$ and some noise, and the problem of *function approximation* from input–output paired samples $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^n$ is considered. Such a problem formulation is called *supervised learning*, where output y is regarded as supervision from the (noisy) oracle. The supervised learning problem is called *regression* if the output y is continuous, and is called *classification* if the output y is categorical.

After introducing standard function models used for regression and classification in Chapter 21, various regression and classification techniques will be introduced. The most fundamental regression technique called the *least squares* method is introduced in Chapter 22, its constraint (or regularized) variants for avoiding overfitting are introduced in Chapter 23. In Chapter 24 and Chapter 25, more advanced regression techniques considering *sparsity* and *robustness* will be discussed, respectively.

In Chapter 26, it is shown that the least squares regression method can also be used for classification, and various issues specific to classification will be discussed. In Chapter 27, a powerful classification algorithm based on the *maximum margin principle* called the *support vector machine* and its robust variant is introduced. In Chapter 28, a probabilistic pattern recognition method that directly learns the class-posterior probability $p(y|\boldsymbol{x})$ called *logistic regression* is introduced. Finally, in Chapter 29, classification of sequence data is discussed.