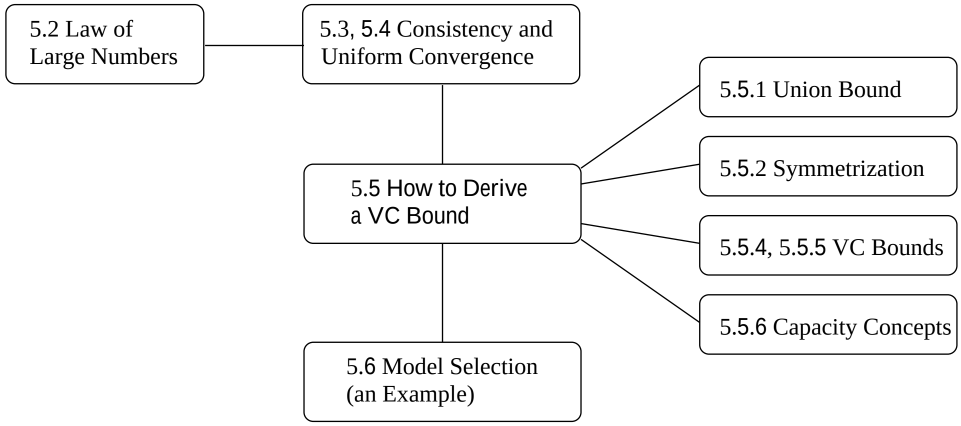

Regression Example

where for simplicity we take $\mathcal{X} = \mathcal{Y} = \mathbb{R}$. Figure 5.1 shows a plot of such a dataset, along with two possible functional dependencies that could underlie the data. The dashed line represents a fairly complex model, and fits the training data perfectly. The straight line, on the other hand, does not completely "explain" the data, in the sense that there are some residual errors; it is much "simpler," however. A physicist measuring these data points would argue that it cannot be by chance that the measurements almost lie on a straight line, and would much prefer to attribute the residuals to measurement error than to an erroneous model. But is it possible to *characterize* the way in which the straight line is simpler, and why this should imply that it is, in some sense, closer to an underlying true dependency?

Bias-Variance Dilemma

In one form or another, this issue has long occupied the minds of researchers studying the problem of learning. In classical statistics, it has been studied as the *bias-variance dilemma*. If we computed a linear fit for every data set that we ever encountered, then every functional dependency we would ever "discover" would be linear. But this would not come from the data; it would be a *bias* imposed by us. If, on the other hand, we fitted a polynomial of sufficiently high degree to any given data set, we would always be able to fit the data perfectly, but the exact model we came up with would be subject to large fluctuations, depending on

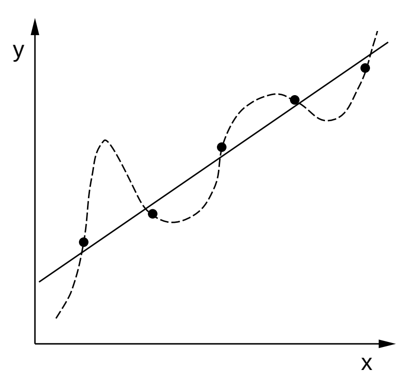

Figure 5.1 Suppose we want to estimate a functional dependence from a set of examples (black dots). Which model is preferable? The complex model perfectly fits all data points, whereas the straight line exhibits residual errors. Statistical learning theory formalizes the role of the *complexity* of the model class, and gives probabilistic guarantees for the validity of the inferred model.