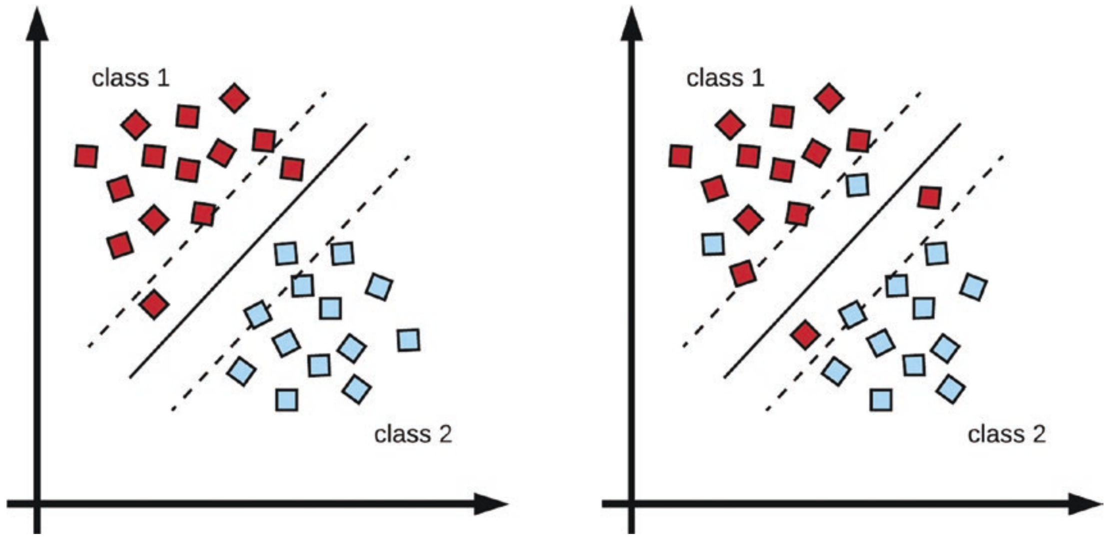

Figure 22-6. *Left: An example of a soft margin with points allowed to violate the margin. Right: An example with some points intentionally misclassified.*

The C Parameter

The C parameter is the hyper-parameter that is responsible for controlling the degree of violations to the margins or the number of intentionally misclassified points allowed by the support vector classifier. The C hyper-parameter is a non-negative real number. When this C parameter is set to 0, the classifier becomes the large margin classifier.

In a soft margin classifier, the C parameter is tuned by adjusting its values to control the tolerance of the margin. With larger values of C, the classifier margins become wider and more tolerant to violations and misclassifications. However, with smaller values of C, the margins become narrower and are less tolerant of violations and misclassified points.

Observe that the C hyper-parameter is vital for regulating the bias/variance trade-off of the support vector classifier. The higher the value of C, our classifier is more prone to variability in the data points and can under-simplify the learning problem. Also, if C is set closer to zero, it results in a much narrower margin, and this can overfit the classifier, leading to high variance – and this will likely fail to generalize to new examples (see Figure 22-7).

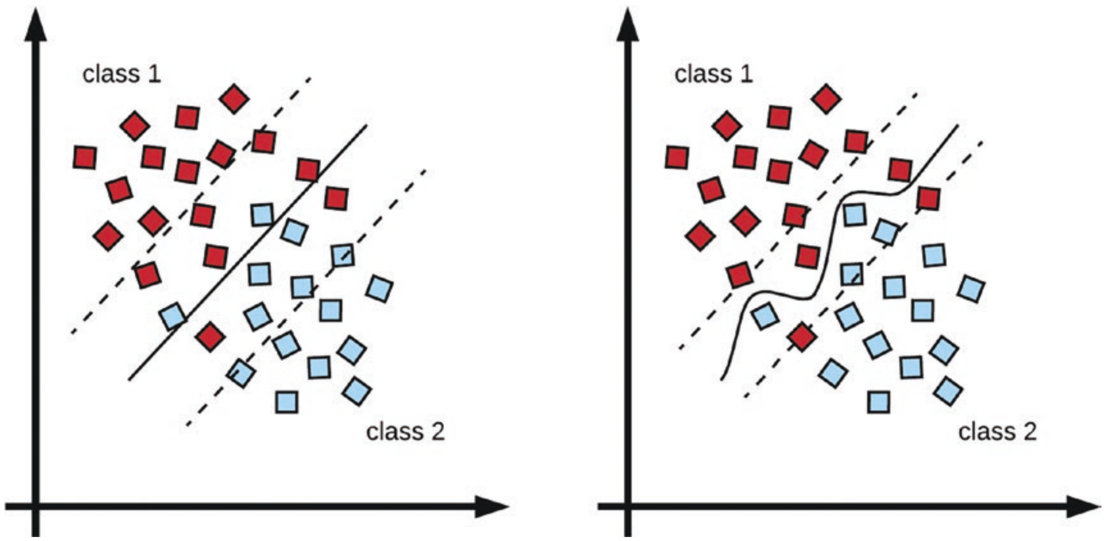

Figure 22-7. *Left: Higher values of C result in wider margins with more tolerance. Right: Lower values of C result in narrower margins with less tolerance*

Multi-class Classification

Previously, we have used the SVC to build a discriminant classifier for binary classes. What happens when we have more than two classes of outputs in the dataset, which is often the case in practice? The SVM can be extended for classifying k classes within a dataset, where k > 2. This extension is, however, not trivial with the SVM. There exist two standard approaches for addressing this problem. The first is the one-vs.-one (OVO) multi-class classification, while the other is the one-vs.-all (OVA) or one-vs.rest (OVR) multi-class classification technique.

One-vs.-One (OVO)

In the one-vs.-one approach, when the number of classes, k, is greater than 2, the algorithm constructs "k combination 2", $\binom{k}{2}$ classifiers, where each classifier is for a pair of classes. So if we have 10 classes in our dataset, a total of 45 classifiers is constructed or trained for every pair of classes. This is illustrated with four classes in Figure 22-8.

After training, the classifiers are evaluated by comparing examples from the test set against each of the $\binom{k}{2}$ classifiers. The predicted class is then determined by choosing the highest number of times an example is assigned to a particular class.