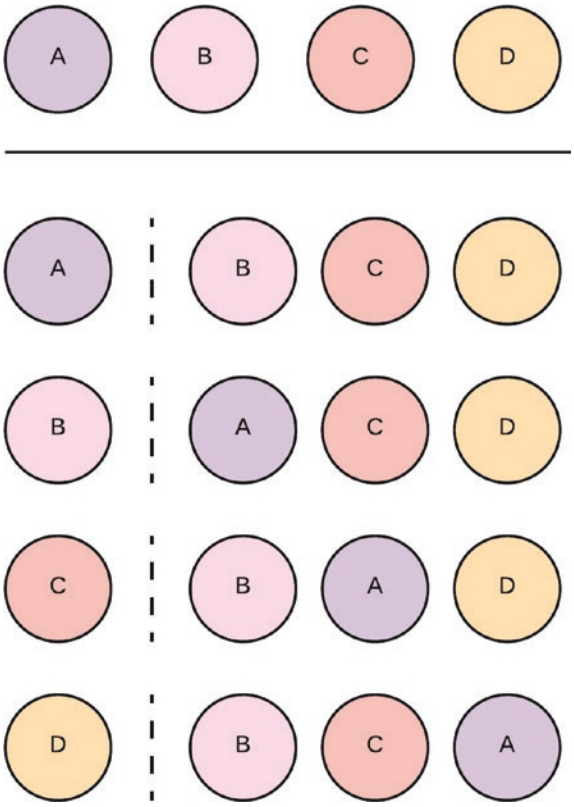

Figure 22-9. *Given four classes in a dataset, we construct four classifiers, with each class fitted against the rest*

The classifiers are evaluated by comparing a test example to each fitted classifier. The classifier for which the margin of the hyperplane is the largest is chosen as the predicted classification target because the classifier margin size is indicative of high confidence of class membership.

The Kernel Trick: Fitting Non-linear Decision Boundaries

Non-linear datasets occur more often than not in real world scenarios.

Technically speaking, the name support vector machine is when a support vector classifier is used with a non-linear kernel to learn non-linear decision boundaries.

SVM uses an essential technique for extending the feature space of a dataset to construct a non-linear classifier. This technique is called kernel and is popularly known as the kernel trick. Figure 22-10 illustrates the kernel trick as an extra dimension is added to the feature space.

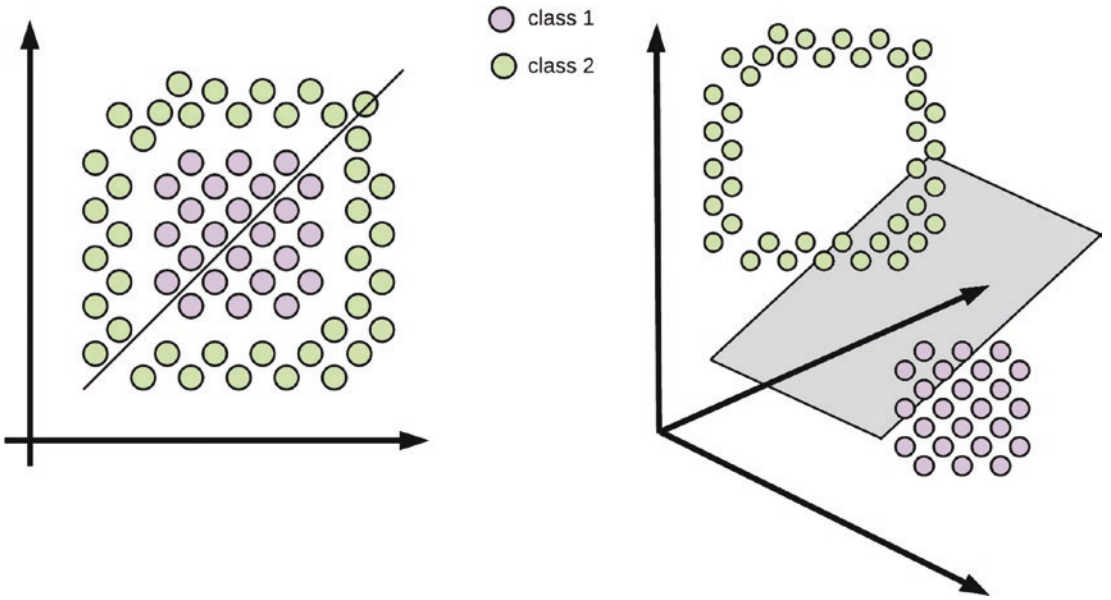

Figure 22-10. *Left: Linear discriminant to non-linear data. Right: By using the kernel trick, we can linearly separate a non-linear dataset by adding an extra dimension to the feature space.*

Adding Polynomial Features

The feature space of the dataset can be extended by adding higher-order polynomial terms or interaction terms. For example, instead of training the classifier with linear features, we can add polynomial features or add interaction terms to our model.

Depending on the dimensions of the dataset, the combinations for extending the feature space can quickly become unmanageable, and this can easily lead to a model that overfits the test set and also become expensive to compute with a larger feature space.