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.

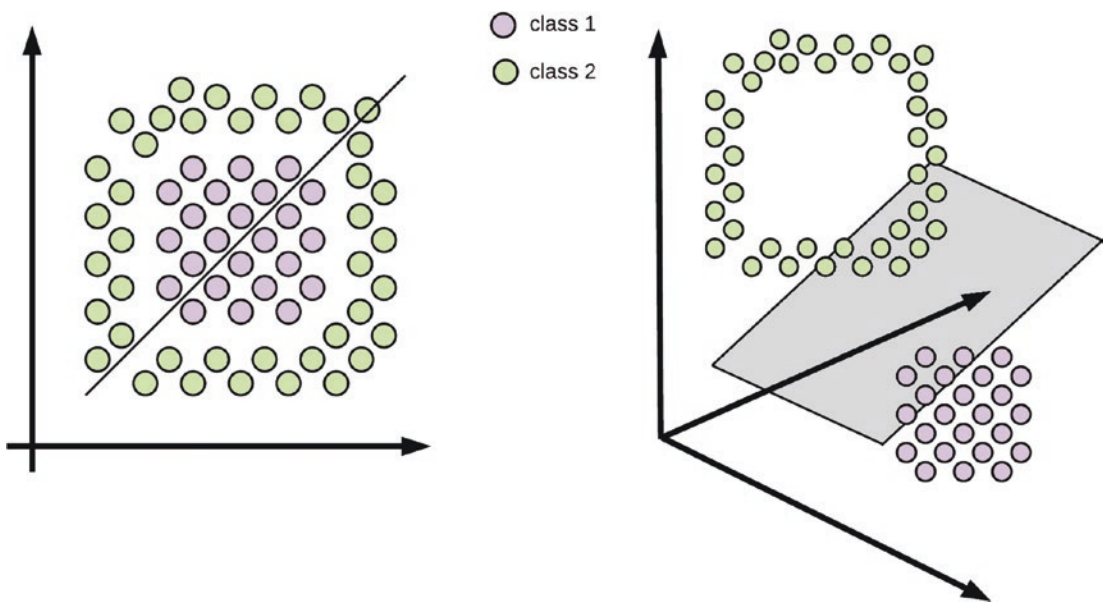

Figure 22-10. *Left: Linear discriminant to non-linear data. Right: By using the kernel trick, we can linearly separate a non-linear dataset by adding an extra dimension to the feature space.*

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.

Kernels

Kernel is a mathematical procedure for extending the feature space of a dataset to learn non-linear decision boundaries between different classes. The mathematical details of kernels are beyond the scope of this text. Suffice to say that a kernel can be seen as a mathematical function that captures similarity between data samples.

Linear Kernel

The support vector classifier is the same as a linear kernel. It is also known as a linear kernel because the feature space of the support vector classifier is linear.

Polynomial Kernel

The kernel can also be expressed as a polynomial. With this, a support vector classifier is trained on higher-dimensional polynomial features without manually adding an exponential number of polynomial features to the dataset. Adding a polynomial kernel to the support vector classifier enables the classifier to learn a non-linear decision boundary.

Radial Basis Function or the Radial Kernel

The radial basis function or radial kernel is another non-linear kernel that enables the support vector classifier to learn a non-linear decision boundary. The radial kernel is similar to adding multiple similarity features to the space. For the radial basis function, a hyper-parameter called gamma, γ, is used to control the flexibility of the non-linear decision boundary. The smaller the gamma value, the less complex (or flexible) the non-linear discriminant becomes, but a larger value for gamma leads to a more flexible and sophisticated decision boundary that tightly fits the non-linearity in the data, which can inadvertently lead to overfitting. This is illustrated in Figure 22-11. RBF is a popular kernel option used in practice.