```
# create polynomial features
polynomial_features = PolynomialFeatures(2)
data = polynomial_features.fit_transform(data)
data
'Output':
array([[ 1.,  5.,  8., 25., 40., 64.],
       [ 1.,  9.,  3., 81., 27.,  9.],
       [ 1.,  8.,  6., 64., 48., 36.],
       [ 1.,  5.,  2., 25., 10.,  4.],
       [ 1.,  3.,  9.,  9., 27., 81.],
       [ 1.,  8.,  7., 64., 56., 49.],
       [ 1.,  1.,  5.,  1.,  5., 25.]])
```

Machine Learning Algorithms

This chapter provides an introduction to working with the Scikit-learn library for implementing machine learning algorithms.

In the next chapters, we'll implement supervised and unsupervised machine learning models using Scikit-learn. Scikit-learn provides a consistent set of methods, which are the **fit()** method for fitting models to the training dataset and the **predict()** method for using the fitted parameters to make a prediction on the test dataset. The examples are geared at explaining working with Scikit-learn; hence, we are not so keen on the performance of the model.

CHAPTER 19

Linear Regression

The fundamental idea behind the linear regression algorithm is that it assumes a linear relationship between the features of the dataset. As a result of the pre-defined structure that is imposed on the parameters of the model, it is also called a parametric learning algorithm. Linear regression is used to predict targets that contain real values. As we will see later in Chapter 20 on logistic regression, the linear regression model is not adequate to deal with learning problems whose targets are categorical.

The Regression Model

In linear regression, the prevailing assumption is that the target variable (i.e., the unit that we want to predict) can be modeled as a linear combination of the features.

A linear combination is simply the addition of a certain number of vectors that are scaled (or adjusted) by some arbitrary constant. A vector is a mathematical construct for representing a set of numbers.

For example, let us assume a randomly generated dataset consisting of two features and a target variable. The dataset has 50 observations (see Figure 19-1).

E. Bisong, *Building Machine Learning and Deep Learning Models on Google Cloud Platform*,
https://doi.org/10.1007/978-1-4842-4470-8_19