

Introduction

An overview of the main models used in scikit-learn.

In the **Data Modeling** section, you will be creating a variety of models for linear regression and classifying data. You will also learn how to perform hyperparameter tuning and model evaluation through cross-validation.

A. Creating models for data

The main job of a data scientist is analyzing data and creating models for obtaining results from the data. Oftentimes, data scientists will use simple statistical models for their data, rather than machine learning models like neural networks. This is because data scientists tend to work with smaller datasets than machine learning engineers, so they can quickly extract good results using statistical models.

The scikit-learn library provides many statistical models for [linear regression](#). It also provides a few good models for classifying data, which will be introduced in later chapters.

When creating these models, data scientists need to figure out the optimal [hyperparameters](#) to use. Hyperparameters are values that we set when creating a model, e.g. certain constant coefficients used in the model's calculations. We'll talk more about hyperparameter tuning, the process of finding the optimal hyperparameter settings, in later chapters.