CHAPTER 12

Matplotlib and Seaborn

It is critical to be able to plot the observations and variables of a dataset before subjecting the dataset to some machine learning algorithm or another. Data visualization is essential to understand your data and to glean insights into the underlying structure of the dataset. These insights help the scientist in deciding with statistical analysis or which learning algorithm is more appropriate for the given dataset. Also, the scientist can get ideas on suitable transformations to apply to the dataset.

In general, visualization in data science can conveniently be split into **univariate** and **multivariate** data visualizations. Univariate data visualization involves plotting a single variable to understand more about its distribution and structure, while multivariate plots expose the relationship and structure between two or more variables.

Matplotlib and Seaborn

Matplotlib is a graphics package for data visualization in Python. Matplotlib has arisen as a key component in the Python data science stack and is well integrated with NumPy and Pandas. The **pyplot** module mirrors the MATLAB plotting commands closely. Hence, MATLAB users can easily transit to plotting with Python.

Seaborn, on the other hand, extends the Matplotlib library for creating beautiful graphics with Python using a more straightforward set of methods. Seaborn is more integrated for working with Pandas DataFrames. We will go through creating simple essential plots with Matplotlib and seaborn.

Pandas Plotting Methods

Pandas also has a robust set of plotting functions which we will also use for visualizing our dataset. The reader will observe how we can easily convert datasets from NumPy to Pandas and vice versa to take advantage of one functionality or the other. The plotting features of Pandas are found in the **plotting** module.

E. Bisong, *Building Machine Learning and Deep Learning Models on Google Cloud Platform*,
https://doi.org/10.1007/978-1-4842-4470-8_12