Introduction

An overview of the development process for machine learning models.

In this section you will be learning about the model execution process in TensorFlow. This involves model training, evaluation, and making predictions. The techniques and APIs introduced in the upcoming chapters are used by top industry professionals to easily create organized and efficient machine learning models.

The first model you create is going to be a classification model. Classification models take data observations and label them with a certain category. For example, classification models can used for tasks like fraud detection, where something is categorized as either fraud or legitimate.

The second model you create is going to be a regression model. Regression models take data observations and output a number. which which can be used for predicting sales with this course's supermarket dataset.

A. Creating a model

Machine learning models all follow the same format; they require an input layer, apply some computations, and produce an output. In the following chapters, we will cover this entire process using a generic multilayer perceptron (MLP) neural network.

B. Development process

The development process for a machine learning model involves continuous training and evaluation. Many machine learning models take an incredibly long time to train. Some take days, or even weeks, to finish training.

Because of this long training time, it is useful to periodically save the current model state while training. This allows us to stop training if we need to and resume training from the same spot at a later date. It also lets us restore previous versions of a trained model, in case something goes wrong or we

overtrain.

Being able to save and restore models in a streamlined fashion is one of TensorFlow's best features. Another strength of the TensorFlow library is the ability to easily log certain values, such as a model's accuracy or loss. This is important in making sure we notice any strange occurrences during training.