Summary of Chapter Two from

Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow

By

Mohamed Abdelbaset Abdelsamie Bakr

Contents

This chaper illistrate End-to-End Machine Learning Project the goal is to illustrate the main steps of a Machine Learning project Here are the main steps we will walk through:

1.Look at the big picture.

➢ Frame the Problem

- what exactly the business objective is? Knowing the objective is important because it will determine how you frame the problem, which algorithms you will select, which performance measure you will use to evaluate your model, and how much effort you will spend tweaking it.

- what the current solution looks like? The current situation will often give you a reference for performance, as well as insights on how to solve the problem.

➢ Select a Performance Measure

It gives an idea of how much error the system typically makes in its predictions.

➢ Check the Assumptions

it is good practice to list and verify the assumptions that have been made so far (by you or others); this can help you catch serious issues early on.

2.Get the data.

It is of the utmost importance to collect reliable data so that your machine learning model can find the correct patterns. The quality of the data that you feed to the machine will determine how accurate your model is. If you have incorrect or outdated data, you will have wrong outcomes or predictions which are not relevant.

3.Explore and visualize the data to gain insights.

to understand how it is structured and understand the relationship between various variables and classes present.

4.Prepare the data for Machine Learning algorithms.

Putting together all the data you have and randomizing it. This helps make sure that data is evenly distributed, and the ordering does not affect the learning process.Cleaning the data to remove unwanted data, missing values, rows, and columns, duplicate   values, data type conversion, etc. You might even have to restructure the dataset and change the rows and columns or index of rows and columns. Splitting the cleaned data into two sets - a training set and a testing set. The training set is the set your model learns from. A testing set is used to check the accuracy of your model after training.