Python for Finance and Algorithmic Trading

Machine Learning, Deep Learning, Time Series Analysis, Risk and Portfolio Management for Metatrader™ 5 Live Trading

POWERED BY

jupyter

Quantreo

Lucas Inglese

Book Link : https://www.amazon.com/gp/product/B09HG18CYL

Extract Chapter 10: Ensemble methods and decision tree

In this chapter, we are going to see many algorithms of machine learning. First, we will see the decision tree, then the random forest, and some other ensemble methods instrumental in finance. All algorithms in this chapter will be applied to the Google stock price.

10.1. Decision tree

In this section, we will study the decision tree algorithm, which is very useful in finance.

10.1.1. Decision Tree classifier

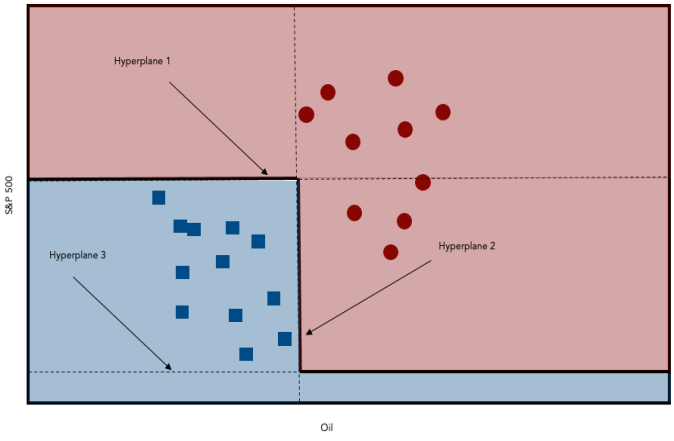
In this subsection, we are beginning with the decision tree[1] classifier. Before code something, we are going to explain how a decision tree classifier works.

Mathematically, when you have a space with groups of points, the decision tree will cut the space with hyperplanes to separate the groups. A hyperplane is a line in a 2d space, so if you do not know what this means, just remember it is a line in our example.

Moreover, each hyperplane is orthogonal to the others. What does it mean? In our case, each hyperplane is perpendicular to the others. It is essential to know that to understand the behavior of the algorithm. I know it can be hard to know like this. In the following figure, you will see how works a decision tree and realize that it is straightforward to understand.

[1] **Additional lecture** : Decision Trees in Machine Learning, Prashant Gupta

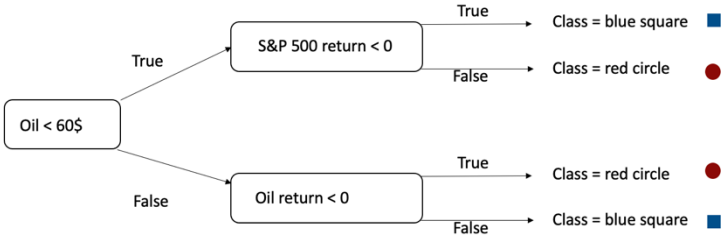
Figure 10.1: How works graphically a decision tree



In this figure, we can see the different hyperplanes of the decision tree. In this example, there are three hyperplanes which means the depth of this decision tree is 3.

We will schematize the functioning of a decision tree to be sure we have understood what a decision tree classifier does is. The decision tree works with conditions, let us see in figure 10.2.

Figure 10.2: Decision tree classifier



In this figure, we can see how a decision tree works.

10.2. Random Forest

Book Link : https://www.amazon.com/gp/product/B09HG18CYL

In this section, we are going to speak about the random forest. First, we will see how to compute a random forest classifier, then a random forest regressor, and how to optimize the hyperparameters of a random forest.

10.2.1. Random Forest classifier

In this part, we will learn how to implement a random forest classifier. The random forest is the first ensemble method that we will learn. An ensemble method is a method that regroups algorithms to have a better prediction.

Naturally, as his name says, the random forest is an ensemble of the decision trees. Indeed, the random forest is just an algorithm that combines the predictions of many decision trees to have better predictions.

We are not going deep in explaining the random forest because intuition is easy to understand, but if we want to go deeper, we need much math, which is irrelevant for this book. Let us see the code of a random forest classifier.

Besides that, it is a very interpretable model, which is interesting in finance. The decision trees do not need normalization to have good results. Moreover, it is a perfect algorithm when you have many dummy variables in features.

Now, we are going to see how to implement a decision tree classifier with Python. To do it, we import the DecisionTreeClassifier class of scikit-learn, and we will use the same syntax as for the previous algorithms.