

Michael Lipschultz

Assignment 5 Proposal: Applying Machine Learning to Predict the Stock Market

Introduction

Investing in the stock market seems like a great way to earn money fairly quickly. Unfortunately, it is hard to predict the actions of the stock market. However, I would like to try by using the Dow Jones Index Data Set from UCI

(<http://archive.ics.uci.edu/ml/datasets/Dow+Jones+Index>). This data set contains stocks and some information about those stocks and the time period being considered. It also includes how well the stock did in the next time period. My goal is to predict how well the stock will do in that next time period. I will use scikit-learn to do the machine learning and predictions. Machine learning is a method for having the computer identify patterns in the data. In this case, it will be used to identify what things seem to suggest that the stock market will go up.

Description

My program will start by asking for a comma-separated file holding the information (basically, the format of the file from UCI). A function **will handle asking the user for the filename and making sure it exists**. This function will return a valid filename.

Another function will take a valid filename as a parameter and will open the file and load the contents (into **numpy arrays**?). This data will be returned.

A **third function** will be in charge of running machine learning using **scikit-learn**. I will use cross-validation to train and evaluate my model. I will then get the evaluation results. I will use two different classifiers to see which has better predictions (by comparing their evaluation results). These evaluation results will be displayed to the user.

Topics and Libraries Possibly Used

In the description above, I bolded the topics being used to show some context.

Here, I show them again so they're all in one place:

- Functions -- I will use a method to ask for a valid filename from the user, another to load in the contents, and a third to perform the machine learning. I'm using these functions to separate the tasks into smaller, more specific areas.
- Files - The data I'm working with is coming from a file. I will need to make sure the file exists and I'll need to open the file and read in the contents into the program.
- Lists or numpy - I will need to store the data from the file into something. I'll probably use lists or numpy arrays, depending on which is easier.
- Scikit-learn - I will use this library to perform the actual machine learning