15-112 Term Project Documentation: StockGenie

Chaiyatat Chawaldit

May 2, 2019

1 Volatility Calculation

Let N be the number of elements in a column. Let $\bar{X}$ be the mean of all the elements in a column.

$$\text{Volatility} = \text{Sd} = \sqrt{\frac{\sum_{i=1}^{N} (X_i - \bar{X})}{N-1}} \tag{1}$$

2 Relative Strength Index (RSI)

The RSI index measures the momentum of changes in price fluctuations. The RSI process has been adapted to fit the needs of this project. The RSI factor consists of a combination of two components: Average gains and average losses.

2.1 Average Gains

Average of all positive percentage movements over a given time period. Decreases in percentage points are not counted in this calculation.

2.2 Average Losses

Average of all negative percentage movements over a given time period. Increases in percentage points are not counted in this calculation.

2.3 RSI Calculation

$$SI = 100 - \left[\frac{100}{1 - \frac{Avg.Gains}{Avg.Losses}}\right] \tag{2}$$

3 Implementation of Support Vector Machine (SVM)

The sklearn module was used to calculate the dual coefficients and intercept used in the calculations that follow. The modified decision function below was used for this project.

3.1 Decision Function Algorithm

The general form for the decision function is as follows:

$$\mathrm{sgn}\Big(\sum_{i=1}^{N} y_i \alpha_i K(x_i, x_j) + \rho\Big) \tag{3}$$

With the output of the decision function (3) in the set $\{-1, 1\}$.

Definitions

- $\mathrm{sgn}(f(x))$: Returns $+1$ if $f(x) > 0$ and -1 if $f(x) \leq 0$
- N: Given by number of support vectors calculated
- $y_i\alpha_i$: Dual coefficient retrieved from sklearn module
- x_i: i^{th} support vector
- x_j: new vector input into prediction model
- ρ: intercept value retrieved from sklearn module
- $K(x_i, x_j)$: Kernel function which can be tuned to improve the model. For this project the RBF Kernel $K(x_i, x_j) = e^{(-\gamma(||x_i - x_j||)^2)}$ was used.
- Note: Other kernels include:
  - Polynomial kernel: $K(x_i, x_j) = ((x_i)^T(x_j) + 1)^p$
  - Gaussian kernel: $K(x_i, x_j) = e^{\frac{-1}{2\sigma^2}(||x_i - x_j||)^2}$
  - Sigmoid Kernel: $\tanh(\eta(x_i)^T(x_j) + \nu)$

In order to determine the dual coefficients and intercept values, the function below must be minimized (using Lagrange multipliers) with the given constraints:

$$\min\Big(\frac{1}{2}w^T w + C\sum_{i=1}^{n} \zeta_i\Big) \tag{4}$$

Constraints:

$$y_i(w^T\phi(x_i) + b) \geq 1 - \zeta_i \tag{5}$$

$$\zeta_i \geq 0, i = 1, 2, ..., n \tag{6}$$

Definitions

- w: $\sum_n \alpha_n \phi(x_n)$ for some parameters α_n
- C : $\frac{1}{\alpha_n}$
- ζ_i : Slack variable used for optimization
- $y_i \in \{-1, 1\}$

For the purposes of this project, the sklearn module was used to calculate the values required by the decision function.

4 Visualization of SVM

Figure 1 shows a simple representation of the SVM machine learning process. The hyperplane is calculated through equations (4), (5), and (6). Equation (3) is subsequently used to predict the location of a newly introduced data point. The SVM process aims to create an n-dimensional plane between classes while maximizing the margin shown in Figure 1.

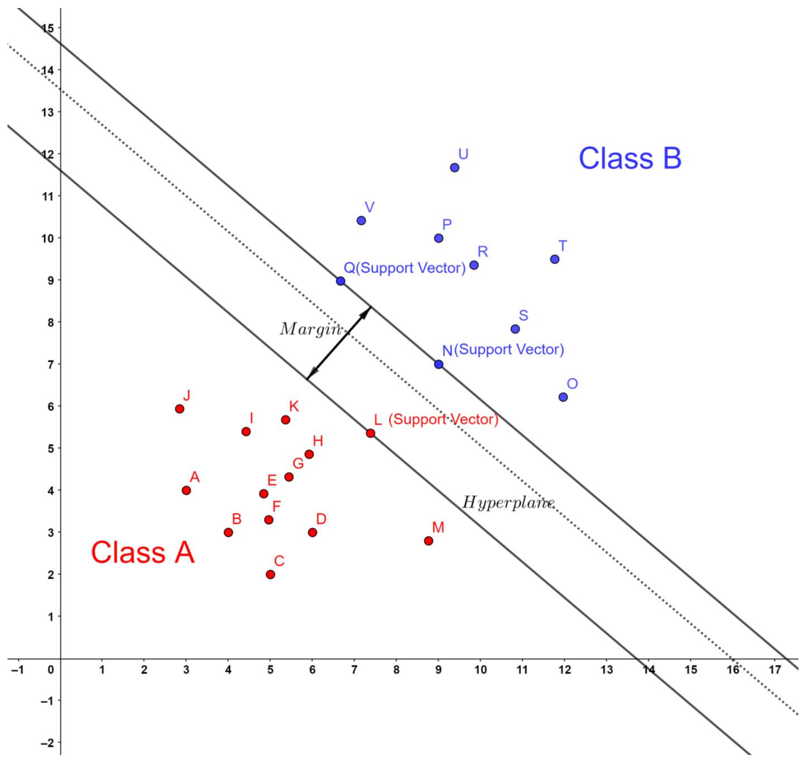

Figure 1: A visual depiction of a support vector machine

The support vectors in Figure 1 (namely L, N, and Q) provide the locations for the boundary lines between class A and class B. These points are important, as they are used in the decision function.

5 References

- https://scikit-learn.org/stable/modules/svm.html
- https://towardsdatascience.com/understanding-support -vector-machine-part-2-kernel-trick-mercers-theorem-e1e6848c6c4d
- https://www.investopedia.com/terms/r/rsi.asp

- http://www.statsoft.com/textbook/support-vector-machines
- A Course in Machine Learning by Hal Daumé III