Bài tập áp dụng

Phân tích xu hướng thị trường chứng khoán bằng giải thuật SVM

Giới thiệu:

Dự đoán chứng khoán là cách thức cố gắng xác định giá trị của một cổ phiếu (stock) trong tương lai

Mục tiêu của bài toán (project) là dùng các chỉ số phân tích kỹ thuật để dự đoán sự tăng hay giảm giá cổ phiếu và sàn giao dịch.

Phương pháp học máy sử dụng thuật phân lớp SVM (Support Vector Classifier)

Các bước cơ bản thực hiện project:

1. Thu thập và xử lý dữ liệu đầu vào (dataset input)
2. Cài đặt giải thuật cho bài toán dự đoán
3. Đánh giá kết quả bài toán sau khi thực nghiệm và kiểm thử
4. Thu thập và xử lý dữ liệu:

* Thu thập dữ liệu (data collection):

+ Nhóm sử dụng dữ liệu quá khứ (historical data) từ trang web Yahoo Finance: <https://finance.yahoo.com/>

(Ví dụ chỉ số index sàn giao dịch NASDAQ: https://finance.yahoo.com/quote/NDAQ?p=NDAQ&.tsrc=fin-srch-v1)

+ Thời gian khảo sát là 5 năm, từ 31/07/2013 đến 01/08/2018

+ Dữ liệu bao gồm cố phiếu 5 công ty (Stock) thuộc Nhóm ngành công nghệ và dịch vụ trực tuyến (Sector) được niêm yết trên sàn giao dịch chứng khoán điện tử NASDAQ (Index)

+ Cổ phiếu 5 công ty gồm:

- Amazon (AMZN.csv)

- Apple (AAPL.csv)

- Google (GOOG.csv)

- Intel (INTC.csv)

- Microsoft (MSFT.csv)

+ Chỉ số sàn NASDAQ (NDAQ.csv)

Tham khảo:

Yahoo Finance: <https://finance.yahoo.com/>

Chỉ số index sàn giao dịch NASDAQ: <https://finance.yahoo.com/quote/NDAQ?p=NDAQ&.tsrc=fin-srch-v1>

Sàn giao dịch NASDAQ: <https://www.saga.vn/thuat-ngu/nasdaq-san-giao-dich-chung-khoan-nasdaq~501>

+ Dữ liệu chứng khoán (dataset original) được lấy từ web Yahoo Finance gồm các tham số sau:

Date (Ngày)

Open (Giá mở cửa)

High (Giá cao nhất)

Low (Giá thấp nhất)

Close (Giá đóng cửa)

Adj Close (Giá đóng cửa đã điều chỉnh

Volume (Số lượng cổ phiếu)

- Xử lý dữ liệu:

+ Nhóm sử dụng tham số **Close** là giá của cổ phiếu tại ngày đang xét.

+ Giả định cơ bản của thị trường chứng khoán là xu hướng thay đổi tăng giảm của cổ phiếu phụ thuộc vào sự biến động của chính cổ phiếu đó trong quá khứ (những giao dịch trước ngày hiện tại đang xét)

Vì vậy, dựa trên giá Close và những tiêu chí được sử dụng trong lý thyết phân tích kỹ thuật thị trường chứng khoán, nhóm thực hiện xử lý, tính toán, sử dụng thêm 6 tham số sau:

* Index\_Momentum
* Index\_Volatility
* Sector\_Momentum
* Stock\_Momentum
* Stock\_Volatility
* Change (Ouput)

Ý nghĩa các tham số trên:

- Momentum (chỉ số của tốc độ thay đổi giá): chỉ báo phân tích kỹ thuật, phản ánh xu hướng và đánh giá tốc độ thay đổi giá dựa trên sự so sánh các giá trị hiện tại và quá khứ. Nếu giá cổ phiếu hay chỉ số index sàn giao dịch cao hơn ngày hôm trước thì momentum bằng 1 (tăng), ngược lại bằng 0 (giảm hoặc giữ nguyên)

- Volatility (mức biến động): chỉ số thể hiện giá trị tăng giảm nhiều hay ít của stock/index, đặc trưng cho độ bất ổn hoặc là mức rủi ro trong giao động của giá trị Chứng khoán. Được tính như sau: (Closei – Closei-1)/ Closei-1

**- Index\_Momentum (chỉ số thay đổi của Index): được tính dựa trên trung bình hiệu suất của thị trường trong 5 ngày gần nhất.**

**- Index\_Volatility: trung bình mức biến động (volatility) của index trong 5 ngày gần nhất.**

**- Sector\_Momentum: trung bình chỉ số thay đổi (momentum) năm ngày gần nhất của 5 công ty trong Nhóm ngành (sector) đã chọn ở trên.**

**- Stock\_Momentum: trung bình chỉ số thay đổi (momentum) cổ phiếu của mỗi công ty trong 5 ngày gần nhất.**

**- Stock\_Volatility: trung bình mức biến động giá cổ phiếu (volatility) của mỗi công ty trong 5 ngày gần nhất**

Cài đặt giải thuật cho bài toán dự đoán (Algorithm)

SVM (Support vector machine) là một trong những mô hình thuật toán có giám sát (supervi được sử dụng phổ biến cho các bài toán phân lớp

SVM (Support vector machine) is one popular algorithm used for many classification problems. It is one of the supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier.

In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces. As for given problem our data was not linear separable so I implemented a “RBF” kernel which gives better result for nonlinear kernel.

For our problem I implemented SVM using Scikit Learn (sklearn) Library. Using python code, I import library, first I try to for SVM on train dataset and then I try to predict values on that machine for test dataset.

EFFICIENT MARKET HYPOTHESIS - EMH / THUYẾT THỊ TRƯỜNG HIỆU QUẢ

ĐỊNH NGHĨA

EMH là một giả thuyết đầu tư cho rằng nhà đầu tư sẽ ít có cơ hội vượt trội hơn so với những người khác vì hiệu quả của thị trường chứng khoán sẽ khiến cho giá của các cổ phiếu hiện có trên thị trường luôn gắn kết và cung cấp các thông tin tương thích, tức là mọi người đều có khả năng tiếp cận thông tin như nhau. Theo thuyết này thì các chứng khoán sẽ được mua bán trao đổi ngang giá trên các sàn giao dịch chứng khoán và do đó người chơi sẽ không có cơ hội để mua ép giá cổ phiếu hoặc thổi phồng mức giá khi bán. Điểm mấu chốt của EMH là các nhà đầu tư sẽ không có khả năng thành công trên thị trường chứng khoán nếu chỉ dựa vào việc lựa chọn chính xác loại chứng khoán và định đúng thời gian tiến hành giao dịch, cách duy nhất để có được lợi nhuận cao trên thị trường này là tiến hành các đầu tư có mức độ rủi ro cao hơn.

Mặc dù là nền tảng của lý thuyết tài chính hiện đại, EMH vẫn là giả thuyết gây nhiều tranh cãi. Những người ủng hộ thuyết này thì cho rằng việc tìm kiếm các chứng khoán được định giá thấp hơn so với giá trị thực hoặc cố gắng dự đoán các xu hướng biến động trên thị trường thông qua các phân tích cơ sở hoặc phân tích kĩ thuật đều là vô ích.

 EMH không có khả năng lý giải sự sụp đổ cũng như phát triển nở rộ của thị trường như đã nói, tuy nhiên các lý thuyết kinh tế học hiện đại lại có thể giải thích được các hành vi trên. Tại sao thị trường lại luôn biến động? Có thể giải thích điều này là do sự tồn tại của hai loại đại lý trong hệ thống thị trường là : "đại lý chính thống" (fundamentalists) và các "đại lý theo phong trào" (chartists)\_nhân tố mà các nguyên lý kinh tế học tân cổ điển chưa bao giờ xét tới. Các đại lý chính thống sẽ tiến hành mua bán trên cơ sở mức giá phản ánh giá trị thực tế của chúng. Trong khi đó các đại lý phong trào sẽ phán đoán dựa trên các dữ liệu quá khứ và tin rằng chúng sẽ ảnh hưởng đến giá cả tương lai của tài sản. Sẽ có một bộ phận trên thị trường tin theo đại lý chính thống và bộ phận còn lại theo cách thức của đại lý phong trào. Tỉ lệ này thay đổi ngẫu nhiên và do đó sự biến động của giá cả tài sản trên thị trường chính là sự phản ánh biến động của tỉ lệ này\_tỉ lệ đối trọng giữa hai loại đại lý hoạt động trên thị trường.

Mục đích chính của thuyết EMH là nhằm giải thích cách thức vận hành thực tế của thị trường tuy nhiên sự đơn giản về mặt toán học của nguyên lý kinh tế học tân cổ điển đã xoá nhoà đi tác dụng của nó. Mặc dù mô hình kinh tế hiện đại còn ít nhiều lộn xộn và chưa được hoàn thiện song nó đã tính đến sự tương tác giữa các mô hình đại lý và có cơ sở thực nghiệm vững chắc hơn.

**Giả thuyết thị trường hiệu quả** (hay lý thuyết thị trường hiệu quả) (en: efficient market hypothesis) là một giả thuyết của lý thuyết [tài chính](https://vi.wikipedia.org/wiki/T%C3%A0i_ch%C3%ADnh) khẳng định rằng *các thị trường tài chính là hiệu quả* (efficient), rằng giá của chứng khoán (securities) trên thị trường tài chính, đặc biệt là [thị trường chứng khoán](https://vi.wikipedia.org/wiki/Th%E1%BB%8B_tr%C6%B0%E1%BB%9Dng_ch%E1%BB%A9ng_kho%C3%A1n), phản ánh đầy đủ mọi thông tin đã biết. Do đó không thể kiếm được [lợi nhuận](https://vi.wikipedia.org/wiki/L%E1%BB%A3i_nhu%E1%BA%ADn) bằng cách căn cứ vào các thông tin đã biết hay những hình thái biến động của [giá cả](https://vi.wikipedia.org/wiki/Gi%C3%A1_c%E1%BA%A3) trong quá khứ. Có thể nói một cách ngắn gọn là các nhà đầu tư *không thể khôn hơn thị trường*