ARIMA

Lý Thuyết:

* **AutoRegression (AR)**

Mô hình AR là một trong những mô hình dự báo thời gian phổ biến trong kinh tế học và thống kê. AR viết tắt cho “Autoregressive” (tự hồi quy), và mô hình AR giả định rằng giá trị của chuỗi thời gian tại một thời điểm phụ thuộc vào các GIÁ TRỊ TRƯỚC ĐÓ của cùng chuỗi thời gian.

Sử dụng mô hình AR để dự báo giá trị tại một thời điểm trong tương lai, ta có thể sử dụng giá trị thực tế của chuỗi thời gian trước đó để ước tính các tham số hồi quy của mô hình và sau đó áp dụng mô hình để dự báo giá trị tại thời điểm mong muốn. Mô hình AR có thể được mở rộng để bao gồm các biến giải thích bổ sung hoặc để xử lý các vấn đề phức tạp hơn, như mô hình ARIMA (Autoregressive Integrated Moving Average) hay mô hình VAR (Vector Autoregression).

Biễu diễn toán học:

Trong đó:

* **Moving Average (MA)**

MA viết tắt cho “Moving Average” (trung bình di động), và mô hình MA giả định rằng giá trị của chuỗi thời gian tại một thời điểm phụ thuộc vào CÁC SAI SỐ NGẪU NHIÊN của chuỗi thời gian tại các thời điểm trước đó.

Biểu diễn toán học:

Trong đó:

* **Autoregressive Integrated Moving Average (ARIMA)**

Mô hình ARIMA kết hợp cả hai mô hình AR (tự hồi quy) và MA (trung bình di động) với một thành phần tích hợp (I) để xử lý các vấn đề phức tạp hơn, như chuỗi thời gian không ổn định hay xu hướng phức tạp của dữ liệu.

ARIMA model là viết tắt của cụm từ Autoregressive Intergrated Moving Average. Mô hình sẽ biểu diễn phương trình hồi qui tuyến tính đa biến (multiple linear regression) của các biến đầu vào (còn gọi là biến phụ thuộc trong thống kê).

**Intergrated**: Là quá trình đồng tích hợp hoặc lấy sai phân. Yêu cầu chung của các thuật toán trong time series là chuỗi phải đảm bảo tính dừng. Hầu hết các chuỗi đều tăng hoặc giảm theo thời gian. Do đó yếu tố tương quan giữa chúng chưa chắc là thực sự mà là do chúng cùng tương quan theo thời gian. Khi biến đổi sang chuỗi dừng, các nhân tố ảnh hưởng thời gian được loại bỏ và chuỗi sẽ dễ dự báo hơn. Để tạo thành chuỗi dừng, một phương pháp đơn giản nhất là chúng ta sẽ lấy sai phân. Một số chuỗi tài chính còn qui đổi sang logarit hoặc lợi suất. Bậc của sai phân để tạo thành chuỗi dừng còn gọi là bậc của quá trình đồng tích hợp (order of intergration). Qúa trình sai phân bậc của chuỗi được thực hiện như sau:

* Sai phân bậc 1:
* Sai phân bậc :

Mô hình ARIMA được biểu diễn bằng phương trình:

Trong đó:

Như vậy về tổng quát thì ARIMA là mô hình kết hợp của 2 quá trình tự hồi qui và trung bình trượt. Dữ liệu trong quá khứ sẽ được sử dụng để dự báo dữ liệu trong tương lai. Trước khi huấn luyện mô hình, cần chuyển hóa chuỗi sang chuỗi dừng bằng cách lấy sai phân bậc 1 hoặc logarit.

* CODE:

Tiếp theo chúng ta sẽ sử dụng package vnquant, một package được tôi viết để hỗ trợ cộng đồng khai thác dữ liệu chứng khoán thuận tiện hơn.