



دانشکده‌ی فیزیک

یادگیری ماشینی در فیزیک

۹ اسفند ۱۳۹۷

پیشنهاد پروژه:

استفاده از تکنیک‌های یادگیری ماشین در تحلیل داده‌های بازار بورس نیویورک

تحت نظارت: دکتر صادق رئیس

نگارنده: کیهان مومنی، علی طه‌ئی، سعید محمودزاده

۱ مقدمه

در طول تاریخ همواره دانشمندان سعی بر توضیح پدیده‌ها به وسیله قوانین ساده داشتند. ولی در بسیاری از موارد وجود پیچیدگی‌های سیستم، بررسی و تحلیل آن‌را دشوار می‌کند. ریاضی‌دان آمریکایی قرن ۱۹، وارن ویور^۱ در مقاله‌ی معروف خود، علم و پیچیدگی [۱]، بیان می‌کند که ۳ دسته از مسائل وجود دارند:

۱. مسائلی که تنها چند متغیر در آن دخیل است (مانند قوانین الکترومغناطیس یا مکانیک) که در قرن ۱۹ توسط علومی مانند فیزیک، شیمی و زیست‌شناسی حل شدند.
 ۲. مسائلی که میلیون‌ها متغیر در آنها دخیل است (مانند مکانیک آماری) که با میانگین‌گیری در طول قرن ۲۰ حل شدند.
 ۳. مسائلی که در آن تعداد قابل توجهی متغیر دخیل است و بین متغیرها روابط غیرخطی حاکم است (مانند الگوهای رفتاری یک گروه انسانی از قبیل یک اتحادیه کارگری).
- روش‌های ما برای حل مسائل دسته سوم (که به سیستم‌های پیچیده^۲ شناخته می‌شوند) تا چنددهه‌ی گذشته چندان کارآمد نبوده است.



شکل ۱: بازار بورس می‌تواند به عنوان یک سیستم پیچیده در نظر گرفته شود.

Warren Weaver^۱
Complex Systems^۲

سیستم‌های پیچیده، ویژگی‌های رفتاری کمابیش مشابهی از خود نشان می‌دهند. مثلاً در بسیاری از آن‌ها خصوصیتی در سطح سیستم ظهور می‌کند که از مطالعه‌ی اجزای سیستم قابل پیش‌بینی نیستند. همچنین بسیاری از آن‌ها دارای خاصیت آشوبناک و نقاط گذار متعدد بوده و به همین خاطر پیش‌بینی پذیری پایینی دارند. دینامیک تکاملی، سازمان‌یافتگی طبیعی و عدم قطعیت ذاتی از دیگر خواصی است که می‌توان برای این سیستم‌ها برشمرد [۲]. استفاده از ابزارهای رایانشی مانند مدل‌سازی، شبیه‌سازی، تحلیل داده و تکنیک‌های یادگیری ماشین در مطالعه این سیستم‌ها در حال فراگیر شدن است. یکی از این سیستم‌های پیچیده بازار بورس است که دارای ویژگی‌های ذکر شده می‌باشد. به همین خاطر می‌توان در پیش‌بینی دینامیک بازار از این روش‌های رایانشی بهره جست. برای مثال می‌توان از روش‌هایی مانند شبکه‌های عصبی^۳ در پیش‌بینی زمان بهینه‌ی خرید یا فروش سهام یک شرکت استفاده کرد [۳].

۲ مسئله

۱.۲ بازار بورس

همانطور که گفته شد، یکی از مثال‌های سیستم‌های پیچیده، بازار بورس است. بازار بورس سیستمی است که شامل مقدار مشخصی سهام بوده و دارای تعدادی خریدار و فروشنده است. در این سیستم تمامی خریداران و فروشنده‌ها، دسترسی برابر به اطلاعات اولیه و اخبار دارند و هرکدام از آنها با توجه به مقدار سرمایه خود توانایی خرید مقداری سهام از فروشنده‌ها را دارند. در نهایت هدف هر شخص در این سیستم، افزایش میزان سود خود می‌باشد. برای افزایش مقدار این سودرسانی، دانستن تغییر قیمت‌ها در آینده لازم و کلیدی است.



پیش‌بینی تغییر قیمت‌ها و مقدار آن بسیار دشوار است. علت این امر وجود متغیرهای مختلف (از قبیل متغیرهای اقتصادی، روان‌شناختی، جامعه‌شناختی، منطقی، عاطفی و ...) می‌باشد. این متغیرها گاهی برهم تأثیر هم می‌گذارند و پیچیدگی سیستم را بیشتر می‌کنند. این امر موجب می‌شود تا پیش‌بینی رفتار بازار سهام با دقت بالا بسیار دشوار باشد. استفاده از یادگیری ماشین در این شاخه می‌تواند موجب شود

^۳Neural Networks

تا الگوهایی که به طور مکرر (روزانه، هفتگی، ماهانه، فصلی، سالانه و بلندمدت) در بازار بورس رخ می‌دهند را شناسایی کرده و ابزارهای آماری و محاسباتی برای پیش‌بینی قیمت‌ها توسعه دهیم.

۲.۲ بازار بورس نیویورک

بازار بورس نیویورک^۴ یا با اختصار *NYSE* بزرگ‌ترین بازار بورس از نظر میزان معاملات و ارزش بازار سرمایه در ایالات متحده آمریکا و جهان می‌باشد [۴]. این بازار بورس که در خیابان شماره ۱۱ وال استریت، منهن جنوبی در شهر نیویورک واقع شده‌است، داده‌های خود را به طور مکرر در وبسایت خود قرار می‌دهد.

۳ داده‌ها

وبسایت *Kaggle* داده‌های روزانه مربوط به تمام شرکت‌های این بازار را منتشر کرده است [۶]. در این مجموعه داده، هر شرکت یک فایل جداگانه دیتابیس دارد که در آن *Open* و *High* و *Low* و *Close* و *Volume* و *AdjClose* هر شرکت به ازای هر روز آورده شده است. همچنین زبان *R* دارای کتابخانه‌ای به نام *quantmod* می‌باشد که با استفاده از آن می‌توان داده‌های به‌روز برای یک شرکت خاص را از سورس *Yahoo* دانلود کرد. ما در این پروژه از هر دوی این منابع استفاده خواهیم کرد.

۴ تحلیل داده‌ها

ما در این پروژه از امکانات هر دو زبان *R* و *Python* استفاده خواهیم کرد. هدف اصلی و نهایی، ساخت یک تخمین‌گر خوب برای پیش‌بینی قیمت سهام یک شرکت می‌باشد. تکنیک‌های یادگیری ماشین از آن جهت مفیدند که می‌توانند با بررسی حجم زیادی از داده، به صورت خودکار به یادگیری، طبقه‌بندی و پیش‌بینی بپردازند. در داده‌هایی که الگوهای پیچیده در کار است، تکنیک‌های بسیار کارآمدی در حوزه یادگیری ماشین توسعه‌یافته‌اند که ما در این پروژه به بررسی آن‌ها خواهیم پرداخت و متدها و *Classifier* های مختلفی مانند *Moving Average*، مدل‌های رگرسیونی، *KNN*، آریما^۵، *Prophet*، *SVC*، *LSTM* و ... را مورد بررسی قرار خواهیم داد. در نهایت بررسی می‌کنیم که استفاده از مدل‌های یادگیری عمیق^۶ تا چه میزان می‌تواند عملکرد پیش‌بینی ما را بهبود ببخشند. این پروژه به تدریج در طول ترم تکمیل شده و بر روی گیت‌هاب [۷] قرار خواهد گرفت.

مراجع

- [1] Weaver, Warren (1948). Science and Complexity. *American Scientist* 36 (536–544).
- [2] <https://www.futurelearn.com/courses/complexity-and-uncertainty/0/steps/1836>
- [3] T. Kimoto, K. Asakawa, M. Yoda and M. Takeoka, "Stock market prediction system with modular neural networks," 1990 *IJCNN International Joint Conference on Neural Networks*, San Diego, CA, USA, 1990, pp. 1-6 vol.1.
- [4] https://en.wikipedia.org/wiki/New_York_Stock_Exchange
- [5] <https://www.nyse.com/data/transactions-statistics-data-library>
- [6] <https://www.kaggle.com/borismarjanovic/price-volume-data-for-all-us-stocks-etfs>
- [7] <https://github.com/wwwKAYHANwww/ML>