



Previsão de retornos financeiros: uma abordagem de aprendizado de máquina

Rafael Antônio R. Branco*, Mauricio Zevallos

Resumo

Modelos de aprendizado de máquina vem sendo utilizados cada vez mais em todas as áreas do conhecimento como consequência da notória capacidade preditiva desses métodos, aliada ao crescente desenvolvimento dos processadores e à imensa disponibilidade de dados. Nesse contexto, o objetivo deste projeto é avaliar o desempenho dos modelos mais comuns de aprendizado de máquina aplicados em um problema de previsão de retornos financeiros a partir de dezenas de variáveis explanatórias. Para tal, foram ajustados o modelos de *Gradient boosted regression trees*, florestas aleatórias e redes neurais com base em dados reais de ativos financeiros de mais de 30 anos de diferentes mercados dos Estados Unidos. Os resultados obtidos, no entanto, mostram que os métodos de aprendizado de máquina não melhoram significativamente as previsões obtidas através da média histórica e não foram identificadas variáveis relevantes.

Palavras-chave:

Florestas Aleatórias, Gradient Boosted Regression Trees, Redes Neurais.

Introdução

Predizer os retornos de ativos financeiros com qualidade é um tema clássico em finanças quantitativas, que se caracteriza por ser tão importante quanto complexo, ver por exemplo [2], [4] e [6].

Com os recentes avanços das técnicas de aprendizado de máquina (ver [1]), que se destacam por seu alto poder preditivo, está ocorrendo uma verdadeira corrida para ajustar novos modelos de predição, com poder preditivo maior do que os tradicionais. Nesse sentido é importante estudar se as novas técnicas de predição são melhores do que os modelos tradicionais da literatura.

Assim, o objetivo deste projeto é contribuir com a literatura, aplicando e testando o desempenho de três métodos de aprendizado de máquina utilizados em [5], que já são consolidados e reconhecidos em outras áreas do conhecimento, em um conjunto de dados com mais de trinta anos de retornos mensais de importantes mercados estadunidenses que já foi analisado com técnicas de regressão em [3]. Especificamente, são discutidos em detalhe aspectos relacionados ao ajuste dos modelos (divisão dos dados em treino, teste e validação etc.), à medição de seus desempenhos (métricas de qualidade da predição) e à avaliação da relevância de cada variável para cada método.

Resultados e Discussão.

Todas as previsões calculadas foram comparadas com a média histórica em termos da média dos quadrados dos erros de predição, e foi notório que essa comparação quase sempre favoreceu a média histórica, e não foram identificadas variáveis relevantes em termos da contribuição ao R^2 .

Houveram, entretanto, picos de desempenho isolados nos modelos ajustados, nos quais a predição dos modelos foi consideravelmente melhor do que as da média. Por outro lado, também houveram quedas expressivas na qualidade das previsões durante períodos de crise.

Dentre todos os modelos testados, os que apresentaram o melhor desempenho foram as redes neurais, cujas previsões se aproximaram mais consistentemente da média histórica (sem superá-la), mostrando maior robustez.

Comparando os resultados com os obtidos em [5], obtivemos um desempenho inferior em termos do R^2 *out of sample*. Um dos motivos disso se dá pelo uso de um R^2 diferente por parte desses autores, que assume a média histórica como sendo zero, causando um aumento artificial na medida de poder preditivo.

Conclusão

Os resultados obtidos, mostram que os métodos de aprendizado de máquina testados não melhoram significativamente as previsões obtidas através da média histórica, e não foram identificadas variáveis relevantes.

Frente a esses resultados, são levantadas algumas possibilidades para a construção de um novo modelo preditivo, baseado em redes neurais, que se ajuste melhor à realidade dos dados e traga melhores resultados.

[1] Efron B, Hastie T. Computer age statistical inference. Cambridge University Press; 2016.

[2] Fama E, French KR. Dissecting anomalies. *The Journal of Finance*. 2008 Jul;63(4):1653–1678.

[3] Green J, Hand JR, Zhang XF. The characteristics that provide independent information about average us monthly stock returns. *The Review of Financial Studies*. 2017 Feb;30(12):4389–4436.

[4] Green J, Hand JR, Zhang XF. The supraview of return predictive signals. *Review of Accounting Studies*. 2013 Dec;18(3):692–730.

[5] Gu S, Kelly B, Xiu D. Empirical asset pricing via machine learning. National Bureau of Economic Research; 2018.

[6] Koijen RSJ, Van Nieuwerburgh S. Predictability of returns and cash flows. *Annual Review of Financial Economics*. 2011 Jan;3(1):467–491.