Modelos de aprendizado de máquina não supervisionado aplicados na construção de portfólios de investimentos

Palavras-chave: Portfólios de investimentos, Seleção de ativos, Fatores de risco, Maximização de retornos

Aluno: Richard Matias de Sousa, IC – UNICAMP
Orientador: Carlos Trucios Maza, IMECC – UNICAMP

Introduction

Tradicionalmente, a construção de portfólios se baseia na teoria média-variância de Markowitz (1952), que busca maximizar o retorno esperado para um dado nível de risco, definindo a chamada fronteira eficiente. Apesar do seu amplo uso, esta abordagem tem sido criticada devido ao baixo desempenho observado fora da amostra, a concentração de pesos em poucos ativos e ao problema de precisar invertar a matriz de covariância, uma tarefa que em dimensões grandes, pode ser bastante desafiadora.

Nesse contexto, técnicas de aprendizado de máquina não supervisionadas têm sido propostas para superar essas dificuldades. Um caso notável é o *Hierarchical Risk Parity* (HRP), introduzido por López de Prado (2016). O HRP aplica o agrupamento hierárquico sobre uma funação da matriz de correlação dos ativos para identificar uma estrutura de grupos de ativos semelhantes. Com base nessa hierarquia, o algoritmo aloca capital de forma ponderada pelo risco dentro de cada grupo, sem requerer a inversão da matriz de covariância. Estudos mostram que o HRP constrói portfólios diversificados que, em testes, superam a otimização média-variância tradicional em desempenho.

A proposta de López de Prado não é isenta de críticas, e diversas abordagens superando algumas das limitações de HRP tem sido propostas (uma da principal critica é não usar toda a informação vinda da clusterização, mas apenas a ordem dos ativos). Entre os métodos que tentam superar HRP, temos o *Hierarchical Clustering-Based Asset Allocation* (HCAA) proposto por Raffinot (2018a). Diferentemente do HRP original, o HCAA explora explicitamente a estrutura hierárquica dos clusters ao alocar capital: após agrupar os ativos, o algoritmo distribui o capital de forma igual entre os diferentes níveis da hierarquia. Em outras palavras, vários ativos altamente correlacionados recebem, no total, a mesma alocação de capital que um único ativo não correlacionado. Essa estratégia de igualar a alocação por cluster reduz a concentração em poucos ativos, resultando em carteiras mais diversificadas.

Outra vantagem de utilizar algoritmos de aprendizado de máquina não supervisionado na construção de portfólios é que eles geram grupos de ativos (clusters) como produto do processo. Essa estrutura de agrupamento pode então ser aproveitada para aplicar diversas técnicas eficientes de alocação de capital. Por exemplo, é possível usar a alocação igualitária (equal weight) entre os clusters formados, ou ainda alocar dentro de cada cluster usando estratégias baseadas em risco, que na prática, isso significa empregar métodos como paridade de risco (risk parity) ou orçamento de risco (risk budget) tanto dentro quanto entre os clusters. Essas abordagens permitem construir carteiras robustas e diversificadas, tirando proveito dos grupos formados pelo algoritmo de agrupamento e otimizando a alocação de capital de forma eficiente.

Além de HCAA, várias outras estratégias baseadas na ideia inicial de HRP foram propostas na literatura, entre elas a estratégia chamada Hierarchical Risk Budgeting proposto por Koumou (2023), Hierarchical Equal Risk Contribution Portfolio que também foi proposto por Raffinot em Raffinot (2018b).

Metodologia

Os modelos aplicados foram as técnicas de agrupamento hierárquico e não hierárquico para construção de portfólios de investimento, com o objetivo de identificar padrões estruturais nos ativos selecionados. No caso do agrupamento hierárquico, foi testada a estratégias de ligação chamada Ward, que funciona juntando clusters buscando minimizar a variância intra cluster dessa forma criando grupos mais equilibrados, podendo favorecer a alocação final dos ativos. Já para o agrupamento não hierárquico, exploramos técnicas como o K-means, visando comparar a estabilidade e a qualidade dos agrupamentos gerados.

Para avaliar o desempenho de cada estratégia, foi realizado um processo de backtesting com janela móvel (Rolling Window), utilizando $T = 120$ observações mensais em cada iteração e calculando os pesos para o

instante $T+1$, uma abordagem amplamente adotada na literatura de econometria financeira. Esse procedimento permite que o desempenho dos modelos seja avaliados em diversos momentos do tempo, trazendo um maior panorama às analises realizadas.

Resultados e Discussão

O período fora da amostra consiste de 192 meses, coomprendidos entre 1986 e 2024. O backtesting foi realizado utilizando dados mensais das 100 maiores empresas da bolsa brasileira desde dezembro de 1997 e os dados foram obtidos através do Economatica. As estratégias testadas foram Hierarchical Risk Parity (HRP), Hierarchical Clustering-based Asset Allocation (HCAA), Hierarchical Risk Budgeting (HRB) e X-means proposto por Nakagawa et al. (2020). Cada estratégia teve o desempenho avaliado por métricas quantitativas bastante utilizadas na literatura, as quais são brevemente descritas a seguir:

Tabela 1: Desempenho das carteiras fora da amostra

Método	**AV**	**SD**	**SR**	**ASR**	**SO**	**TO**	**SSPW**
X_Means	11.790893	20.578426	0.572974	0.575262	1.028105	0.662684	0.038237
HCAA	11.349303	22.362377	0.507518	0.510565	0.908499	0.628444	0.038892
HRP	11.598831	21.964383	0.528075	0.530079	0.934846	0.514685	0.025408
HRB	10.998324	29.413502	0.373921	0.373777	0.618554	0.142153	0.505496

Os resultados do backtest indicam que a estratégia de alocação baseada no algoritmo X-Means, implementada usando a estratégia de Risk Parity, apresentou o melhor desempenho geral entre as abordagens avaliadas. Essa configuração obteve a maior rentabilidade média anual do estudo, atingindo 11,79%, além de demonstrar os mais elevados indicadores de retorno ajustado ao risco, incluindo um Sharpe Ratio de 0,573, um Adjusted Sharpe Ratio de 0,575 e um Sortino Ratio de 1,028. Em comparação, a estratégia HCAA, que utiliza agrupamento hierárquico com alocação igualitária entre e dentro dos clusters, apresentou uma rentabilidade média ligeiramente inferior de 11,35% e uma volatilidade mais elevada, com desvio padrão de 22,36%. Já a abordagem HRP, atingiu uma rentabilidade média de 11,60% e um desvio padrão de 21,96%, mostrando uma performance relativamente pior em relação ao X-Means. A estratégia HRB, fundamentada na alocação por orçamento de risco, registrou o pior desempenho entre as alternativas testadas, com uma rentabilidade média de 10,99%, acompanhada dos menores indicadores de desempenho ajustado ao risco, com Sharpe Ratio de 0,374, Adjusted Sharpe Ratio de 0,374 e Sortino Ratio de 0,619. Portanto, com base nos resultados obtidos podemos concluir que a abordagem X-Means, ao ser combinada com o critério de alocação Risk Parity, destacou-se como o método de criação de carteira mais eficiente no equilíbrio entre retorno e risco, demonstrando superioridade em todas as métricas de desempenho e consolidando-se como a estratégia de melhor resultado no período analisado.

As métricas analisadas foram as seguintes. O Retorno Médio (AV) é a média aritmética dos retornos periódicos da carteira, refletindo o ganho esperado por período. O Desvio Padrão (SD) quantifica a dispersão estatística desses retornos em torno da média, mensurando a volatilidade do investimento. O Índice de Sharpe (SR) divide o retorno excedente da carteira (sobre a taxa livre de risco) pelo desvio padrão dos retornos, indicando o prêmio de risco obtido por unidade de volatilidade. O Sharpe Ajustado (ASR) amplia esse conceito ao incluir penalizações por assimetria negativa e curtose na distribuição de retornos, recompensando carteiras com maior skew positivo e menor curtose. De forma complementar, o Índice de Sortino (SO) é análogo ao Sharpe, mas substitui a volatilidade total pela volatilidade negativa (de valores abaixo de um retorno de referência), focando apenas nos riscos de queda. O *Turnover* (TO) mensura a rotatividade da carteira: é a fração dos ativos comprados ou vendidos em cada rebalanceamento, sendo indicador de custos de transação potenciais. Por fim, a soma dos quadrados dos pesos (SSPW) mede a concentração do portfólio – valores mais baixos (próximos de zero) sugerem carteiras amplamente diversificadas, enquanto valores mais altos indicam concentração de posições. As implementações podem ser vistas no GitHub.

Referências

Koumou, G. (2023). Hierarchical risk budgeting. *Journal of Financial Data Science*. Available at SSRN: https://ssrn.com/abstract=4821555.

López de Prado, M. (2016). Building diversified portfolios that outperform out of sample. 42(4):59–69.

Markowitz, H. (1952). A large-scale high-resolution wbc image dataset. *SPRINGER NATURE*, 7(1):77–91.

Nakagawa, K., Kawahara, T., and Ito, A. (2020). Asset allocation strategy with non-hierarchical clustering risk parity portfolio. *Journal of Mathematical Finance*, 10:513–524.

Raffinot, T. (2018a). The hierarchical equal risk contribution portfolio. *Available at SSRN 3237540*.

Raffinot, T. (2018b). The hierarchical equal risk contribution portfolio. *Available at SSRN 3237540*.