

# Uma Abordagem Probabilística Discreta para Análise Autônoma de Investimentos baseada em sistemas multiagentes e Online learning

Paulo André Lima de Castro

pauloac@ita.br

Professor

Laboratório de Sistemas Computacionais Autônomos (LabSCA)

Divisão de Ciência da Computação

Instituto Tecnológico de Aeronáutica (ITA)

# Sumário

- Robôs podem investir [melhor que nós]?
  - Um ambiente repleto de Desafios e Armadilhas...
  - Overfitting
- Uma Abordagem Probabilística Discreta para Análise Autônoma de Investimentos baseada em sistemas multiagentes e Online learning
  - Discretização e Custos de Erros
  - Resultados
- Conclusões e Futuros Trabalhos

# Robôs podem investir [melhor que nós]?

- A questão rememora a questão colocada por Turing no clássico paper "Can machines think?" [1]
- Muitas objeções podem ser levantadas contra a ideia de máquinas analisando e gerenciando investimentos....porém talvez menos que aquelas imaginadas por Turing para máquinas pensantes..
- Vejamos algumas das objeções mais relevantes a máquinas que podem investir.....

# Objecções a máquinas que podem investir...

- **Objecção Matemática:** Análise de investimentos é mais do que lógica é um tipo de arte, logo está além dos limites de computabilidade...
- **Objecção “Heads in the Sand”:** Máquinas controlando investimentos teria efeitos terríveis, tais como crises econômicas e roubariam empregos de pessoas reais
  - Por outro lado, a ideia de que seria possível obter retornos excepcionais com AAI não parece provável: A ampla adoção de AAI levaria a rápidas correções dos preços a cada nova informação disponibilizada...isto é, o mercado ficaria mais próximo da Hipótese de Mercado Eficiente de Eugene Fama
- Máquinas poderiam fazer grande parte do trabalho, mas nunca seriam capazes de fazer X
  - X: ter intuição, ter bom senso, ser inovador, pensar algo realmente novo, etc.

# Quão difícil é Análise Autônoma de Investimentos?

- Análise Autônoma de Investimentos seria classificada segundo Russel e Norvig como:
  - parcialmente observável, sequencial, estocástico, dinâmico, contínuo e multiagente,
- Que é o caso mais complexo [Russel, Norvig, 2013]
- Entretanto, isso não captura toda a complexidade.
  - Mais que estocástico, AAI é um processo não estacionário (Distribuições de Probabilidade mudam ao longo do tempo) e é também **estratégico** pois há competição entre os agentes
- Há muitas iniciativas relacionadas a AAI (acadêmica e profissionais), mas apesar de todo os esforço realizado, Não há [ainda] um “**Deep blue**” da análise de investimento (ou em investment management)

# Muitos desafios e armadilhas no caminho...

- Entres estes desafios há alguns que são esquecidos ou não entendidos completamente:
  - **Overfitting** tal como em ML pode acontecer e é na verdade pode ser difícil de evitar, a medida que se busca melhorar o desempenho dos algoritmos
  - É difícil determinar quais informações são realmente relevantes e estas provavelmente não são as mesmas para diferentes ativos.
  - Uma mudança no horizonte de investimento pode mudar sensivelmente a lógica de análise!
  - Investidores não tem os mesmos níveis aceitáveis de risco e retorno.  
**Inverstors are not all the same!**
- Estes desafios nos fizeram optar por uma abordagem multiagente e discretizada para o problema, porém cuidados adicionais devem ser utilizados em relação a **overfitting**

# Overfitting em ML e Finanças

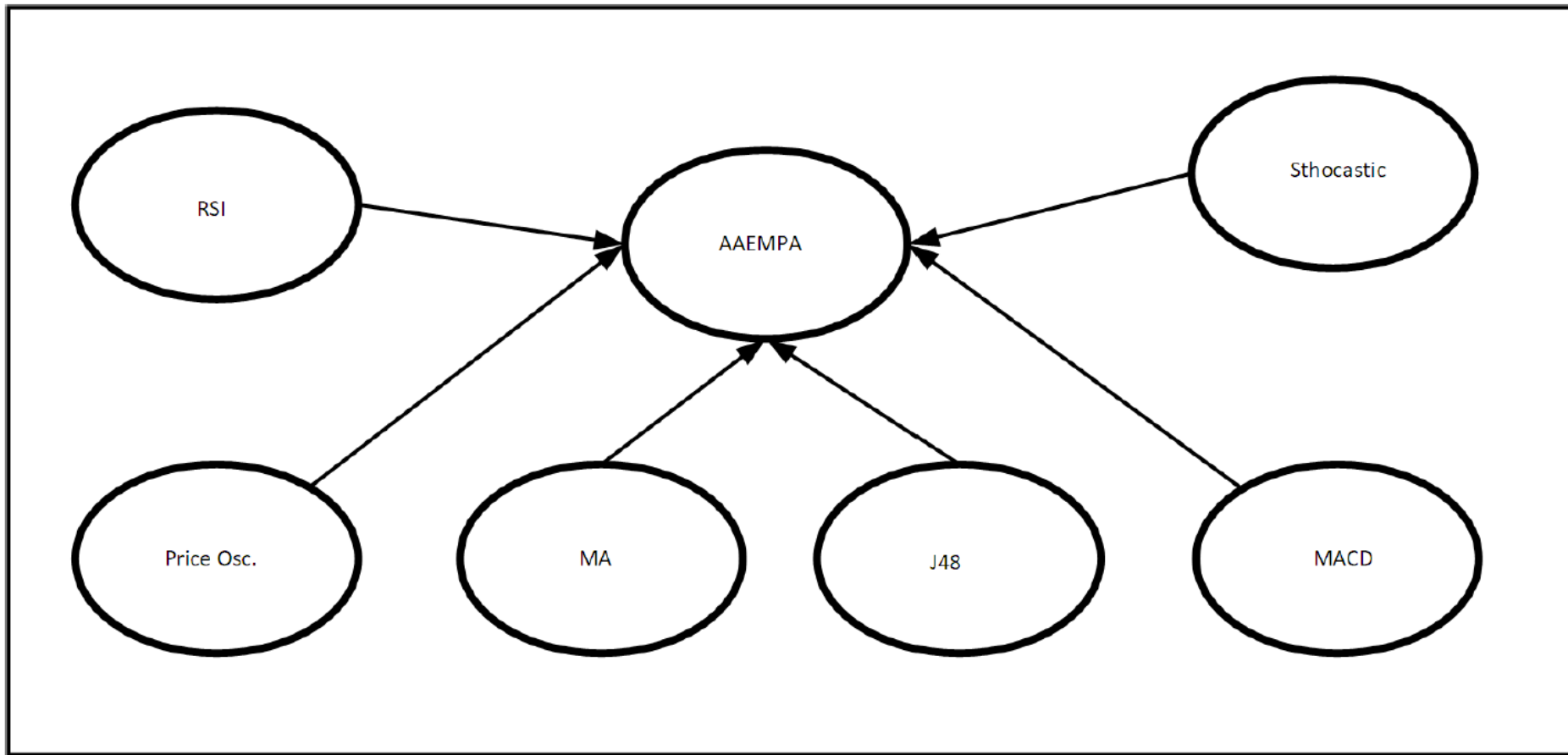
- Dizemos que um Sistema de ML está em “**overfitting**” quando seu desempenho é muito bom em dados de treinamento (ou teste) mas seu desempenho é significativamente pior fora dos dados captados
- De fato, o artigo “**Pseudo-Mathematics and Financial Charlatanism: The Effects of Backtest Overfitting on Out-of-Sample Performance**”, Bailey et al, 2014 faz importantes ( e fortes) afirmações sobre modelos de ML em Finanças:
  - “...given any financial series, **it is relatively simple to overfit an investment strategy so that it performs well...**[in the backtest]”
  - “...Although there are many academic studies that claim to have identified profitable investment strategies, their reported results are almost always based on in-sample statistics....”
  - “**We have shown that backtest overfitting is difficult indeed to avoid...**”
  - “**When financial advisors do not control for overfitting, positive backtested performance will often be followed by negative investment results...**”
- Para evitar overfitting, deve-se utilizar validação cruzada (**cross-validation**) (Witten , 2014), “...**significant amount of historic data..**” e “**model sequestration**” Bailey et al, 2014],
  - *Model sequestration: Freeze the proposed investment strategy and analyze it after some time with pure (yet to be observed) data [Bailey et al, 2014],*

# Nossa Abordagem: Utilizar Múltiplos agentes em um contexto discretizado e Aprendizado Online

- Nossa abordagem baseia-se no uso simultâneo de diversos agentes autônomos e na discretização do problema AAI e sua modelagem como um problema de classificação.
- Aprendizado On-line: Prevendo a partir do conselho de diversos especialistas
- Essa abordagem quebra a complexidade enfrentada pela AAI em problemas que podem ser abordados por um grupo de agentes que trabalham em conjunto para fornecer conselhos de investimento inteligentes e personalizados para indivíduos

*What magical trick makes us intelligent? The trick is that there is no trick. The power of intelligence stems from our vast diversity, not from any single, perfect principle.” Marvin Minsky, AI pioneer*





Agentes utilizados na implementação nomeados segundos suas técnicas de análise

# Aprendizado On-line: Prevendo a partir do conselho de especialistas

- Os algoritmos clássicos para resolver o problema de aprendizado on-line [Blum, 98]. O primeiro, bastante simples e direto, estabelece pesos iguais para cada especialista inicialmente e reduz o peso de qualquer especialista que cometa um erro `a metade. A saída escolhida é aquela com maior peso. Blum provou que os erros cometidos por este algoritmo nunca são superiores a  $2,41 (m + \log n)$ , onde  $m$  é o número de erros cometidos pelo melhor especialista e  $n$  é o número de especialistas.
- O segundo algoritmo é uma versão melhorada do primeiro. Blum demonstrou que o número de erros deste segundo algoritmo  $M$  é limitado por ( $\beta$  é um parâmetro escolhido pelo projetista):

$$M \leq \frac{m * \ln(1/\beta) + \ln(n)}{\beta}$$

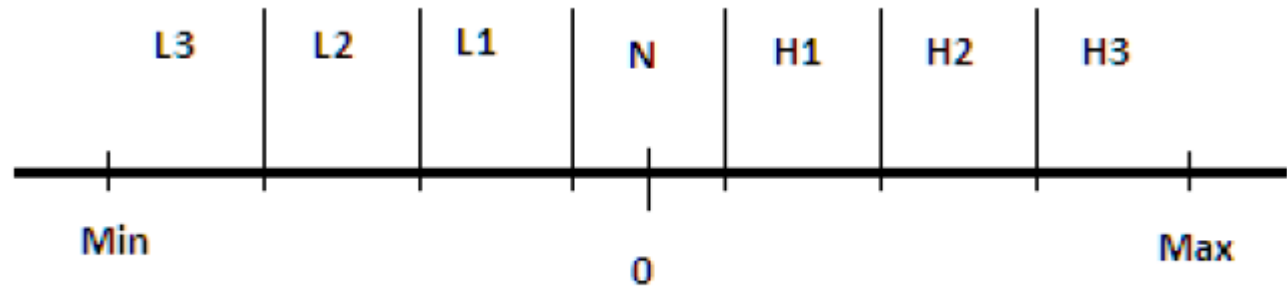
# Erros e Seus Custos...

Um aspecto importante a se observar em AAI é que os erros podem ter custos muito distintos.....

**“Não importa se você está certo ou errado, mas sim quanto dinheiro você faz quando está certo e quanto você perde quando está errado...”** George Soros, Investidor Internacional

# Erros e Custo de Erros

- Se o analista indica um retorno de 85% e a realidade apresenta um valor de 84,9% ou 85,1%, o investidor provavelmente interpretará isso como um acerto na análise, não um pequeno erro.
- Além disso, pequenas variações no retorno previsto, dificilmente mudariam a decisão de investimento, sendo portanto insignificantes para o desempenho do investimento
- A discretização de retornos torna indistinguíveis, valores de retorno próximos, o que é compatível com as preferências gerais de um investidor.
- Seria inadequado avaliar um analista através de simples taxa de acerto, devido a diferença de custos de Erros. Uma forma mais adequada seria avaliar por taxa de acerto ponderada pelo custo do erro.



Exemplo de discretização com cinco classes.

$$AdjustedCostMatrix = \begin{bmatrix} 0 & c_{1,2} & \dots & c_{1,n} \\ c_{2,1} & 0 & \dots & c_{2,n} \\ c_{3,1} & c_{3,2} & 0 & c_{3,n} \\ c_{4,1} & c_{4,2} & \dots & 0 \end{bmatrix}$$

DistribuiçãoPorClasses =

$$\begin{bmatrix} L2 & L1 & N & H1 & H2 & \\ d_{1,1} & d_{1,2} & d_{1,3} & d_{1,4} & d_{1,5} & L2 \\ d_{2,1} & d_{2,2} & d_{2,3} & d_{2,4} & d_{2,5} & L1 \\ d_{3,1} & d_{3,2} & d_{3,3} & d_{3,4} & d_{3,5} & N \\ d_{4,1} & d_{4,2} & d_{4,3} & d_{4,4} & d_{4,5} & H1 \\ d_{5,1} & d_{5,2} & d_{5,3} & d_{5,4} & d_{5,5} & H2 \end{bmatrix}$$

$$TaxaAcertoAjustada = 1 - \frac{1}{\tau} * \frac{\sum_i \sum_j c_{i,j} * d_{i,j}}{c_{max}}$$

# Algoritmo para Aprendizado On-line ajustado a Custo de erro

- O algoritmo de Blum não leva em consideração os custos de Erros, por isso Propomos uma adaptação do Algoritmo de Blum (AAEMPA) e demonstramos que [Castro, 18] ( $m$  é o custo do melhor analista isolado):

Teorema . Em qualquer sequência de ensaios, o custo esperado de erros  $M$  feito pelo Algoritmo Ajustado ao Erro de Maioria Ponderada Aleatória (AAEMPA) satisfaz o seguinte:

$$M \leq \frac{m * \ln(1/\beta_{min}) + \ln(n)}{\beta_{min}}$$

# Resultados obtidos com simulações com 18 ativos mais negociados da B3 [Castro 18]

Taxa de acerto ajustada (%) para horizontes de 1 dia (D), 1 semana (S) e 1 mês(M) com respectivas médias e desvios padrões

Ativo	J48			MA			MACD			PRIOSC			RSI			Stochastic			AAEMPA			Média.	Desv. Pad.
	D	S	M	D	S	M	D	S	M	D	S	M	D	S	M	D	S	M	D	S	M		
BBAS3	0,484	0,531	0,443	0,489	0,402	0,354	0,484	0,554	0,532	0,467	0,552	0,524	0,459	0,403	0,522	0,474	0,458	0,414	0,497	0,420	0,568	0,478	0,058
BBDC4	0,477	0,622	0,499	0,431	0,383	0,326	0,548	0,654	0,586	0,544	0,632	0,578	0,378	0,339	0,429	0,416	0,411	0,330	0,393	0,352	0,447	0,465	0,106
BVMF3	0,533	0,470	0,410	0,342	0,370	0,282	0,559	0,647	0,566	0,566	0,637	0,548	0,365	0,326	0,369	0,396	0,442	0,348	0,414	0,343	0,438	0,446	0,108
CIEL3	0,462	0,752	0,659	0,304	0,361	0,290	0,579	0,771	0,677	0,578	0,756	0,668	0,343	0,306	0,293	0,445	0,516	0,442	0,351	0,314	0,295	0,484	0,174
CMIG4	0,367	0,506	0,440	0,406	0,384	0,313	0,441	0,635	0,557	0,434	0,624	0,546	0,459	0,422	0,463	0,408	0,406	0,349	0,477	0,434	0,499	0,456	0,083
CSNA3	0,415	0,473	0,451	0,443	0,467	0,326	0,427	0,515	0,474	0,430	0,498	0,481	0,498	0,456	0,508	0,450	0,514	0,360	0,489	0,473	0,566	0,463	0,053
GGBR4	0,394	0,465	0,458	0,451	0,409	0,460	0,441	0,498	0,404	0,425	0,489	0,408	0,514	0,491	0,688	0,507	0,537	0,432	0,525	0,487	0,681	0,484	0,079
GOAU4	0,278	0,343	0,456	0,461	0,469	0,551	0,413	0,469	0,450	0,403	0,474	0,454	0,509	0,493	0,716	0,490	0,502	0,469	0,413	0,525	0,756	0,481	0,105
HYPE3	0,492	0,467	0,393	0,365	0,408	0,378	0,590	0,715	0,688	0,581	0,707	0,670	0,372	0,304	0,297	0,446	0,522	0,496	0,391	0,314	0,336	0,473	0,138
ITSA4	0,492	0,597	0,658	0,398	0,356	0,354	0,578	0,696	0,621	0,556	0,677	0,600	0,380	0,352	0,393	0,426	0,483	0,358	0,383	0,354	0,397	0,481	0,123
ITUB4	0,463	0,604	0,643	0,428	0,346	0,347	0,614	0,642	0,642	0,609	0,639	0,637	0,415	0,328	0,423	0,419	0,355	0,346	0,421	0,344	0,440	0,481	0,124
JBSS3	0,416	0,470	0,501	0,367	0,332	0,257	0,531	0,675	0,474	0,528	0,660	0,470	0,410	0,367	0,404	0,427	0,430	0,329	0,417	0,369	0,404	0,440	0,101
PETR3	0,411	0,565	0,610	0,475	0,492	0,470	0,482	0,610	0,567	0,470	0,586	0,549	0,436	0,376	0,509	0,464	0,427	0,480	0,448	0,395	0,519	0,492	0,068
PETR4	0,470	0,495	0,513	0,488	0,487	0,464	0,457	0,569	0,501	0,447	0,554	0,500	0,475	0,429	0,614	0,474	0,495	0,577	0,477	0,429	0,614	0,501	0,054
SUZB5	0,417	0,584	0,494	0,457	0,404	0,332	0,498	0,634	0,592	0,489	0,633	0,593	0,424	0,346	0,397	0,400	0,372	0,282	0,432	0,346	0,397	0,454	0,103
USIM5	0,335	0,420	0,648	0,478	0,517	0,444	0,375	0,479	0,481	0,354	0,470	0,488	0,480	0,448	0,589	0,385	0,510	0,428	0,462	0,485	0,686	0,474	0,087
VALE3	0,351	0,536	0,467	0,438	0,413	0,362	0,448	0,559	0,562	0,439	0,547	0,566	0,476	0,468	0,634	0,450	0,477	0,355	0,452	0,468	0,634	0,481	0,081
VALE5	0,370	0,449	0,468	0,490	0,452	0,435	0,461	0,526	0,514	0,458	0,507	0,518	0,471	0,477	0,668	0,474	0,463	0,474	0,469	0,510	0,688	0,492	0,070
Média	0,424	0,519	0,512	0,428	0,414	0,375	0,496	0,603	0,549	0,488	0,591	0,544	0,437	0,396	0,495	0,442	0,462	0,404	0,440	0,409	0,520	0,474	0,065
Desv. Pad.	0,066	0,092	0,090	0,054	0,055	0,079	0,070	0,086	0,078	0,072	0,085	0,073	0,053	0,066	0,132	0,034	0,053	0,076	0,046	0,070	0,135	0,074	0,025

# Conclusões e Trabalhos Futuros

- Neste trabalho, discutimos a viabilidade de realizar precificação de investimentos e ativos de forma automatizada, campo que chamamos aqui de análise autônoma de investimentos (AAI)
- Propomos uma abordagem para análise autônoma de ativos baseada na discretização de tempo e retorno. Desta forma, AAI pode ser tratado como um problema de classificação, porém é importante observar que os custos podem ser significativamente diferentes entre erros distintos.
- Além disso, observamos que a ***não estacionaridade*** pode fazer com que alguns agentes possam ter bom desempenho em determinado período e depois serem suplantados por outros agentes. Este fato é similar ao enfrentado no contexto de aprendizado on-line, todavia os algoritmos clássicos para lidar com aprendizado on-line não tratam de situações de custo não uniforme
- Nós elaboramos uma versão adaptada do algoritmo proposto por Blum, para trabalhar com custos não uniformes e demonstramos um limite de custo de erro para o novo algoritmo



# Conclusões e Trabalhos Futuros - 2

- Os resultados obtidos apontam uma variação significativa de desempenho entre os analistas, destacam-se os desempenhos do MACD e AAEMPA, mas há ainda bastante espaço para melhoramentos pois as melhores taxas ficam apenas em torno de 60%. A variação de desempenho por ativo é relativamente pequena.
- Acreditamos que uma forma de tentar melhorar o desempenho do sistema é a introdução de novas técnicas de análise, especialmente algoritmos baseados em análise fundamentalista.
- No entanto, está claro que existe um longo caminho para alcançar um analista de investimentos autônomo eficiente. É necessária uma melhor compreensão de como configurar os vários parâmetros nesta arquitetura multiagentes.
  - De fato, como observado por LeBaron, uma crítica comum sobre os mercados baseados em agentes é que eles costumam ter muitos parâmetros e o impacto desses parâmetros não é bem entendido ([29], pp.1222)
- Também acreditamos que uma evolução significativa seria uma modelagem formal das expectativas, que são extremamente importantes no raciocínio financeiro, bem como uma validação do sistema através de “*model sequestration*”

# Referências

- CASTRO, P. A. L.; ANNONI JUNIOR, R. . Towards Autonomous Investment Analysts - helping people to make good investment decisions. In: Future Technologies Conference (FTC) 2016, 2016, San Francisco. Proceedings of the Future Technologies Conference (FTC) 2016. San Francisco: IEEE Press, 2016
- CASTRO, P. A. L.; ANNONI JUNIOR, R. **Análise Autônoma de Investimento: Uma Abordagem Probabilística Discreta** .Revista de Informática Teórica e Aplicada (RITA). vol. 25. num. 1. (Jan,2018).pp. 23-38. 2018
- *David H. Bailey, Jonathan M. Borwein, Marcos López de Prado, and Qiji Jim Zhu. **Pseudo-Mathematics and Financial Charlatanism: The Effects of Backtest Overfitting on Out-of-Sample Performance**. Notices of AMS. V. 61. N.5. 2014*