

Sobre o LAMFO

O Laboratório de Aprendizado de Máquina em Finanças e Organizações (LAMFO) é um centro destinado a desenvolver, aplicar e estudar fenômenos organizacionais nas mais diversas áreas como Marketing, Logística, Administração Pública, Gestão de Pessoas e Finanças.

A equipe é composta por discentes dos cursos de Administração (graduação e pós-graduação), profissionais das mais diversas áreas da Administração e professores da Universidade de Brasília. O uso do Aprendizado de Máquina tem sido cada vez mais utilizado em Ciências Sociais Aplicadas como Economia, Contabilidade e Administração. A proposta é desenvolver algoritmos capazes de reconhecer padrões nos dados visando objetivos estipulados pelos gestores, analistas ou pesquisadores.

O analista em um primeiro momento relaciona as principais informações relevantes para o estudo do fenômeno de interesse, e em seguida por meio de algoritmos estatísticos e econométricos que aprendem como os dados estão relacionados e então reproduzem esse aprendizado por meio de previsões, estimativas e probabilidades para outros conjuntos de dados no qual o comportamento estudado seja similar ao comportamento aprendido por meio dos dados de entrada.

O Laboratório de Aprendizado de Máquina em Finanças e Organizações atua em parceria com entes da Administração Pública federal e regional, iniciativa privada e demais pesquisadores de organizações e finanças na aplicação, compreensão e desenvolvimento de modelos de Aprendizado de Máquina.

O Laboratório de Aprendizado de Máquina em Finanças e Organizações funciona na Universidade de Brasília desde 2016 desenvolvendo pesquisas de Data-Science com foco em aplicação de inteligência artificial. Os projetos de pesquisa do LAMFO seguem calendário próprio de desenvolvimento de cada equipe, e são feitas reuniões de acompanhamento quinzenais gerais do laboratório aos sábados.

Reuniões de desenvolvimento de projeto de pesquisa: De acordo com cada equipe.

Reuniões de acompanhamento e discussão: Quinzenalmente, aos sábados, de 10h:30-11h:00.

Oficinas de programação em Aprendizado de Máquina, Teoria Econômica/Financeira, Métodos de programação: Quinzenalmente, aos sábados, de 11h:15- 12h:00.

A participação nas oficinas é aberta ao público, bastando apenas que seja enviado email para lamfo@unb.br para confirmar a reserva.

Tem alguma dúvida sobre o laboratório? Entre em contato!

O Laboratório recebe novos pesquisadores todos os semestres, ao longo do período letivo da UnB. O LAMFO desenvolve projetos de pesquisa acadêmica, projetos de mercado

privado, bem como desenvolve pesquisa em parcerias com entes externos do setor público e privado.

Laboratório em expansão

Conheça nossos números !

46 Pesquisadores

26 Projetos finalizados

12 Projetos em andamento

Projects Grid

Previsões de Candles e Imagens

Publicado: 09/04/2021

Descrição: Análise de gráficos de cotações com a utilização de Redes Neurais Convolucionais para previsões. Membros: Arthur Nunes Torres Rafael Oliveira Garcêz de Mendonça Vítor Bandeira Borges Vinícius de Oliveira Watanabe Rafael Moraes

Satisfação e Desempenho

Publicado: 09/04/2021

Descrição: Encontrar a relação entre satisfação do trabalhador com desempenho operacional das firmas. Para isso utiliza-se indicadores de satisfação e reviews dos funcionários e indicadores de desempenho econômico da firma. Membros: Pedro Albuquerque Natasha Fogaça Juliana N. Felipe Schiavon João Gabriel Souza

MCTI

Publicado: 09/04/2021

Projeto de ciência de dados do Ministério de Ciência, Tecnologia e Inovações (MCTI), para implantar a automação do portfólio de produtos financeiros (PPF). Membros: Victor Rafael Alixandro Beatriz Simões Fernanda Amorim Hebert Kimura Jairo Alves João Gabriel Lucas Moreira Matheus Kempa Rafael Moraes

Covid Desigualdade

Publicado: 09/04/2021

Em toda crise é comum observar um aumento da distancia da renda entre os mais ricos e mais pobres. O presente projeto propõe analisar a relação e impacto entre desigualdade de renda e Covid. O tamanho do impacto na desigualdade de renda e como cada Estado se comportou (alterando ou não suas posições no rank de desigualdade) [...]

Lamfoot

Publicado: 30/03/2021

Este projeto tem como objetivo prever, por meio de técnicas de Machine Learning, os resultados do Campeonato Italiano para a partir deles construir e validar estratégias de apostas baseadas nas análises e resultados dos modelos rodados. Sua base foi construída com a combinação de resultados e estatísticas reais das equipes, as cotações atualizadas de sites de apostas [...]

University of Essex & Microsoft Ai for Health- NLP para Covid-19

Publicado: 01/10/2020

Misinformation can amplify humanity's greatest challenges. A salient recent example of this is the COVID-19 pandemic, which has bred a multitude of falsehoods even as truth has increasingly become a matter of life and death; people share false claims about COVID-19, in part, because they simply fail to think sufficiently about whether or not content is accurate [...]

Covid-19 Modelling TaskForce: Modelagem Quantitativa para Estudos de Difusão e Impactos da Covid-19, com ênfase em Administração, Economia e Finanças

Publicado: 20/06/2020

Este projeto de pesquisa envolve o desenvolvimento de modelos matemáticos e computacionais para análise quantitativa de fenômenos associados à Covid-19. O projeto explora métodos quantitativos variados para (i) a projeção do número de casos e óbitos por Covid-19 no Brasil e no mundo para (ii) a estimativa da taxa de sub-notificação de casos de Covid-19 no Brasil, [...]

Difusão de métodos quantitativos para análise de dados e tomada de decisão em administração e gestão

Publicado: 26/05/2020

A Covid-19 tem gerado notícias e discussões de termos quantitativos como “crescimento exponencial”, “escala logarítmica”, “modelos epidemiológicos”, “estimativa de parâmetros”, “simulação”, “testes”, “hipóteses”, etc. Todavia tais terminologias causam confusões e seus conceitos não são adequadamente compreendidos pela população em geral. O avanço da ciência e da tecnologia depende de um aumento do nível de conhecimento sobre métodos [...]

Covid-19 Databank

Publicado: 22/05/2020

Líder: Prof. Herbert Kimura Matrícula UnB 1063588 Unidade acadêmica da UnB Faculdade de Economia, Administração, Contabilidade e Gestão de Políticas Públicas e-mail herbert.kimura@gmail.com Link Currículo Lattes <http://lattes.cnpq.br/2048706172366367>

Título da Proposta LAMFO Covid-19 Databank Sumário Executivo da Proposta A substancial quantidade de dados sobre a Covid-19 induz um maior nível informacional sobre a doença, sendo benéfica à sociedade. [...]

Covid educação

Publicado: 14/05/2020

Descrição Notícias sobre a Covid-19 estão sendo acompanhadas por termos como “crescimento exponencial”, “escala logarítmica”, “modelos epidemiológicos”, “estimação de parâmetros”, “simulação”, “testes”, “hipóteses”, etc... Apesar de amplamente utilizados pelos cientistas, esses termos, bem como outros conceitos quantitativos usuais em pesquisas científicas, não amplamente dominados pelo público em geral. Neste contexto, este projeto busca desenvolver atividades que possibilitem [...]

Covid subnotificação

Publicado: 14/05/2020

Descrição Estudar os comportamentos de subnotificação de dados relacionados ao Covid-19 no Brasil utilizando metodologia econométricas espaciais como Tobler (1970) e especificamente em epidemiologia em Zhang et. al (2010). Objetivo – Estimar a subnotificação de dados de Covid-19 no Brasil. Líder : Daniel Oliveira Cajueiro Alex Rodrigues do Nascimento Igor Nascimento

Covid Causality isolamento

Publicado: 14/05/2020

Descrição Seguindo a proposta de mensuração de modelos contrafactuais, este estudo irá buscar estimar os efeitos do confinamento/isolamento imposto pela covid-19. Para tal utilizar-se-á a metodologia contrafactual em alta dimensão em painéis de dados de Carvalho et. al (2018). Objetivo – Estimar e dimensionar os impactos gerados pelas políticas de confinamento/isolamento impostas pela Covid-19. Líder: Daniel Oliveira [...]

Covid projeções modelagem

Publicado: 14/05/2020

Descrição Inicialmente, será conduzido um levantamento bibliográfico para identificar as principais modelagens quantitativas associados (i) ao contágio e transmissão do Covid-19 e (ii) aos impactos econômicos e financeiros da pandemia. Técnicas que o LAMFO já vem desenvolvendo especificamente para a análise de Covid-19 incluem ARIMA, modelos SEIR, Lei de Benford, estatística- não-paramétrica, clusterização, etc... Posteriormente, serão implementados [...]

Covid raspagem

Publicado: 14/05/2020

Descrição: Este projeto tecnológico envolve o desenvolvimento de um repositório de dados, relatórios e notícias veiculadas sobre a Covid-19, notadamente publicadas em

fontes confiáveis. Assim, o projeto contribui para a consolidação de dados sobre o Covid-19, constituindo-se em matéria-prima para a realização de pesquisas futuras. O repositório pode ser utilizado para, por exemplo, (i) explorar relacionamentos entre [...]

Consensus

Publicado: 04/05/2020

Descrição: O objetivo do projeto é a construção de um sistema especialista para sugestão de investimentos. Pesquisadores: Pedro Henrique Melo Albuquerque Marcius Stefano

REAL ESTATE

Publicado: 24/04/2020

Pesquisadores: Alfredo Rossi Peng Yao Hao Pedro Henrique Melo Albuquerque Cayan Atreio Portela Bárcena Saavedra Descrição: O projeto tem como objetivo prever os preços dos imóveis por região do Distrito Federal, usando como base dados da Caixa Econômica Federal utilizando técnicas de deep learning (LSTM) e comparando com modelos de regressão tradicional.

FEATURE SELECTION

Publicado: 24/04/2020

Pesquisadores: Cayan Atreio Portela Bárcena Saavedra Peng Yao Hao Herbert Kimura Pedro Henrique Melo Albuquerque Descrição: Desafios inerentes à era do big data, como a alta dimensão, são cada vez mais comuns no aprendizado de máquina e no aprendizado estatístico. No âmbito de finanças, é possível observar esse evento referente ao grande número de indicadores presentes nas análises [...]

TRADING DINÂMICO

Publicado: 24/04/2020

Pesquisadores: Victor Matheus Rodrigues de Carvalho Ennio Ferreira Bastos Junior Leonardo Lacerda Galler Descrição: Apesar de bem explorado em sistemas financeiros mais maduros, o uso de algoritmos para trading tem se tornado cada vez mais comum no mercado nacional. Diante desse cenário, o projeto visa o desenvolvimento de estratégias de investimento baseadas em indicadores da análise técnica [...]

QRNN

Publicado: 24/04/2020

Pesquisadores: Stefano Giacomazzi Dantas Leonardo Lacerda Galler Descrição: Métricas de risco são extremamente importantes no contexto financeiro. Dentre várias métricas disponíveis, destaca-se o Value at Risk (VaR), que busca representar os elementos de risco de um portfólio em apenas um número. O objetivo desse projeto é desenvolver um modelo baseado em redes neurais recorrentes para estimar o [...]

BANK DEFAULT CONCENTRATION AND COMPETITION

Publicado: 09/03/2020

Descrição: O Objetivo principal do projeto é analisar os impactos das medidas de concentração e competição bancária no default dessas instituições e em um possível efeito sistêmico. Levando-se em consideração a capacidade de injeção de liquidez do estado, impactos de Too Big to Fail, e as diferenças entre competição e concentração. Pesquisadores: João Gabriel de Moraes Souza [...]

Seminários em IA

Carolina Santos - Ética em ciência de dados

Thiago Marzagão - Ciência de dados & governo

PEDRO ALBUQUERQUE - Let's Data Podcast

Comportamento e Consumo: Seminário e Lançamento do Estudo Anatel / Ibict

INTELIGÊNCIA ARTIFICIAL - 22º BATE-PAPO CIÊNCIA E TECNOLOGIA NO DIA A DIA

Carreira em Data Science e IA - Prof. Pedro Henrique Melo Albuquerque

Machine Learning em Finanças, Banking e Economia - Prof. Daniel Oliveira Cajueiro

Erros que Cometi - Matheus Facure

Fundamentos do Machine Learning, uma introdução sem matemática - Peng Yaohao

Minha experiência em tech companies, vida no exterior e como o LAMFO me ajudou chegar até aqui . - Sarah Sabino

Parcerias

O Laboratório de Aprendizado de Máquina em Finanças e Organizações (LAMFO) atua em parceria com os mais diversos entes da Administração Pública Federal e Regional e iniciativa privada.

Dentre esses acordos de Cooperação Técnica destacam-se os seguintes projetos já realizados ou em fase de realização:

Órgão: Conselho Administrativo de Defesa Econômica

Projeto: BRA/11/008 Economic and Consumer Rights

Desenvolvimento: 08/2014-06/2016

Descrição: A cooperação proveu ao CADE conhecimento técnico especializado consistente no desenvolvimento de modelos de Aprendizado de Máquina aplicados à detecção e repressão a cartéis, aprimorando assim, a atividade de monitoramento de práticas de mercado por parte da Superintendência-Geral (inc. I, art. 13, Lei 12.529/2011). Durante esse projeto desenvolveu-se técnicas de monitoramento de mercados para a

avaliação de risco de colusão e identificação de potenciais infrações a ordem econômica através de métodos estatísticos e econométricos do campo de Aprendizado de Máquina.

Órgão: Confederação Nacional da Indústria

Projeto: Desenvolvimento de sistema georeferenciado para apoio a decisão.

Desenvolvimento: 01/2018-09/2018

Descrição: A cooperação proveu ao CNI um sistema desenvolvido em shiny (<https://lamfo.shinyapps.io/cnifast/>) para apoio a decisão dos gestores quanto a definição dos cursos a serem oferecidos, localidade bem como apresentar as características das empresas e escolas para cada Unidade da Federação.

Empresa: NVIDIA Corporation

Projeto: NVIDIA GPU Grant.

Desenvolvimento: 09/2018

Descrição: O LAMFO foi contemplado pelo NVIDIA com a doação de uma placa aceleradora Titan Xp para execução de projetos que necessitem de GPU no treinamento, validação e teste das máquinas construídas pelo laboratório. Agradecemos o apoio da NVIDIA Corporation pela doação da GPU Titan Xp usada em nossas pesquisas.

Órgão: Laboratório Nacional de Computação Científica

Projeto: SDumont - SINAPAD

Desenvolvimento: 2018-2019

Descrição: O LAMFO foi escolhido, junto com outros laboratórios e grupos de pesquisa do Brasil, para ter acesso ao uso gratuito de processamento em nuvem do computador SDumont do Laboratório Nacional de Computação Científica. Agradecemos o apoio da LNCC Corporation pela doação da GPU Titan Xp usada em nossas pesquisas.

Empresa: Microsoft AI for Health

Projeto: Curbing Misinformation: Information extraction, categorisation and visualisation for COVID-19 news using NLP unsupervised techniques.

Desenvolvimento: 08/2020

Descrição: Research funds to tackle the misinformation around COVID-19 using NPL (US\$ 61,766)

Membros

Pedro Albuquerque

Pesquisador Sênior

Herbert Kimura

Daniel Cajueiro

Victor Rafael Celestino

Carlos Henrique Rocha

Silvia Araújo dos Reis

Natasha Fogaça

Coordenador(a) Operacional

Matheus Kempa Severino

Coordenador(a) de Negócios

João Gabriel de Moraes Souza

Coordenador(a) Acadêmico

Yaohao Peng

Coordenador(a) Adjunto Acadêmico

Vinícius Watanabe

Coordenador(a) Adjunto Operacional

Jairo Alves

Vítor Bandeira Borges

Pedro Alexandre

Conselheiro

Sarah Sabino

Alfredo Rossi

Pesquisador

Cayan Portela

Fausto Sá Teles

Fernanda Amorim

Flávio Augusto Rodrigues de Oliveira

Lucas Moreira Gomes

Maisa Melo

Marcus Lima

Mateus Hiro Nagata

Rafael Morais

Vanderson Aparecido Delapedra da Silva

Allan Victor

Assistente de pesquisa

Arthur Nunes Torres

Arthur Nery

Arthur Alves

Bruno Lopes

César Freitas

Ian Barreiro

João Pedro Fontoura da Silva

Joao Pedro Rodovalho Figueiredo

João Victor Maia Costa

Johnathan Milagres

Matteo Kimura

Pedro Zarur Canellas

Pedro Watuhã dos Santos Guerra

Rafael Garcêz

Rafael Lopes dos Santos

Rubens Toledo

Willian Kesley Oliveira Cardoso

Ycleda de Oliveira

OFICINAS

EnANPAD 2015

Getting to the Basics: Dimensão VC e Generalização em Machine Learning

Random Forest e LightGBM

Restricted Boltzmann Machine - RBM

Support Vector Machine - SVM

Teste de Eventos

Generalized Linear Models

Self Organizing Maps

Quantum: Physics & Computation

Crash Course de amostragem

Fatos Estilizados em Finanças

Autoencoder

Análise discriminante linear

Feature Engineering

Bibliometria & Meta-análise

Generalized Method of Moments

PROPENSITY SCORE MATCHING

Modelagem de volatilidade no mercado de opções

Redes Neurais Convolucionais

Blockchain technology

Dados em Painel

Estatística Espacial

Cálculo Estocástico

Programação Dinâmica

Oficina de Medidas Default

COMPLEXIDADE TEMPORAL E ESPACIAL

CLUSTERIZAÇÃO E K-MEANS

Diagnóstico em modelagem

Crash course: Teoria de finanças

Probabilistic Graphical Models

Programação Funcional - Análise de Dados

Introdução a Banco de dados E Linguagem SQL

Bank Stability and Systemic Risk: Competition, Concentration and Bailout Probability Approach

Oficina LAMFO Anova & Manova

Algoritmos evolutivos e programação genética

Previsão de Fraudes em Seguros de Propriedade

Rede Neural Convolucional e Rede Neural de Cápsula

Feature Selection e Deep NNs para previsão dos preços de ações

Modelagem de Volatilidade: ARCH, GARCH e MMAR

Eficiência de Mercado e Random Walk

Difference in differences (DID)

Regression discontinuity design de Mercado e Random Walk

BLOG

Fatos Estilizados em Finanças

Posted by Marcius Lima, Eduardo Rubik, Rafael Moraes on November 7, 2020

Fatos Estilizados em Finanças

Padrões que ocorrem recorrentemente no mercado financeiro são de suma importância para praticantes e estudiosos da área. Eles possibilitam criar e ajustar modelos que ajudarão no melhor entendimento dos movimentos do mercado. Mas quais são alguns desses padrões e como são tratados no campo das Finanças?

Ao observar uma série de elementos em diferentes contextos e períodos, em determinados ocasiões é possível fazer formulações teóricas de fenômenos ou características observadas empiricamente. A essas formulações, dá-se o nome de fatos estilizados.

Em Finanças, os fatos estilizados referem-se a características de ativos financeiros que são consistentes tanto entre diferentes mercados quanto entre diferentes períodos, e que, devido a isso, são generalizadas. Essas generalizações são importantes, visto que norteiam o processo de modelagem estatística dos retornos.

Um ponto importante a se observar é que todas as características referem-se aos retornos dos ativos, e não seus preços. Isso porque uma característica importante nesse processo é a estacionariedade, ou seja, a “consistência” do objeto analisado ao longo do tempo. Os preços são muito afetados por fatores como a conjuntura econômica e o estado da arte tecnológico, por exemplo. Já os retornos costumam ser menos afetados, e portanto, manter uma constância maior.

Nesse texto, abordaremos os seguintes fatos estilizados:

Insignificância de autocorrelações

Caudas grossas

Assimetria de ganhos e perdas

Gaussianidade agregacional

Agrupamento de volatilidade

Efeito alavancagem

Correlação Volume X Volatilidade

Insignificância de Autocorrelações

Na estatística, o conceito de autocorrelação diz respeito a correlação entre uma série temporal e sua versão com lag em diferentes intervalos de tempo. Ou seja, ao medir a correlação entre os retorno de uma série no momento t e no momento $t-1$ para vários valores de t , o resultado é baixíssimo, ao ponto de ser insignificante.

Talvez uma forma mais intuitiva de entender esse conceito seja pelo jargão “retornos passados não garantem retornos futuros”, ou seja, que o retorno de determinada série em determinado momento é insignificamente influenciado pelo anterior.

Insignificância de autocorrelação na prática: Momentum

Lo e MacKinley, 1990

Queda cadenciada de volatilidade das autocorrelações de retornos absolutos ou quadrados

Importante observar que a insignificância das autocorrelações dos retornos é uma propriedade relacionadas ao valor dos retornos. Porém, ao observar os retornos absolutos ou quadrados há uma autocorrelação que decai lentamente.

Ou seja, o retorno em determinado período tem pouca correlação com o valor no período seguinte, mas a “dimensão” desses valores possuem correlações, que vai diminuindo a medida que a diferença temporal aumenta.

Taylor, 2005

Caudas Grossas

Quanto a distribuição de probabilidade dos retornos, é possível observar que eles não apresentam uma curva normal, mas sim uma curva com maior concentração nas caudas da distribuição. Na prática, mais retornos apresentam valores menores e menos retornos apresentam valores maiores (muita gente tem baixos retornos e pouca gente tem altos retornos).

Caudas Grossas na prática

Silva et al, 2019

Assimetria de ganhos e perdas

Outra característica associada a distribuição dos retornos é a assimetria de ganhos e perdas, representada graficamente por uma das caudas possuir uma maior “inclinação” que a outra. Essa propriedade diz respeito aos movimentos de queda (fundos) serem maiores que os de subida (topos). Utilizando um outro jargão, é como se os retornos “subissem de escada e descessem de elevador”.

Gaussianidade agregacional

A distribuição dos retornos são associadas, porém, a um intervalo de tempo determinado. A gaussianidade agregacional diz respeito a distribuição ir se aproximando de uma normal (também chamada de distribuição Gaussiana, daí o nome) a medida que o horizonte de tempo é expandido.

Visto que o aumento no intervalo de tempo está associado ao tamanho da amostra analisada, o Teorema Central do Limite garante essa propriedade.

Efeito Alavancagem

O Efeito Alavancagem foi explicado pela primeira vez em 1976 (Black, 1976) e atesta que existe uma correlação entre a alavancagem dos ativos da empresa e a volatilidade do preço de suas ações. Empresas com alavancagem financeira e operacional tendem a ficar mais alavancadas com a queda do valor da firma, o que aumenta a volatilidade dos retornos de suas ações.

Correlação Volume X Volatilidade

De acordo com Cont(2001), o volume é correlacionado com todas as medidas de volatilidade. Importante notar que a intensidade da correlação positiva entre os dois fatores dificulta a separação dos efeitos de cada um sobre os ativos (Bianco e Renò, 2006).

<http://businessforecastblog.com>

Como a indústria de gestão de recursos pode utilizar Machine Learning?

A previsão de preços dos ativos talvez seja a aplicação mais popular. No entanto, existem diversas aplicações tão ou mais importantes como:

construção de portfólio; detecção de outliers; análise de sentimento; market making; tamanho de “apostas”; outros.

“Nós precisamos de aprendizado de máquinas para melhorar teorias financeiras, e precisamos de teorias financeiras para restringir a propensão que o aprendizado de máquinas tem de sobreajustar.” Marcos López de Prado

Referências

Empirical properties of asset returns: stylized facts and statistical issues in: Quantitative Finance, Vol 1, No 2, (March 2001) 223-236.

Lewellen, Jonathan. “Momentum and Autocorrelation in Stock Returns.” The Review of Financial Studies, vol. 15, no. 2, 2002, pp. 533–563. JSTOR, www.jstor.org/stable/2696788. Accessed 28 Oct. 2020.

Lo, Andrew MacKinlay, A. (1990). When Are Contrarian Profits Due To Stock Market Overreaction?. Review of Financial Studies. 3. 175-205. 10.1093/rfs/3.2.175.

MALMSTEN, Hans; TERÄSVIRTA, Timo. Stylized Facts of Financial Time Series and Three Popular Models of Volatility. European Journal of Pure and Applied Mathematics, [S.l.], v. 3, n. 3, p. 443-477, may 2010. ISSN 1307-5543.

Girard, Eric Biswas, Rita. (2007). Trading Volume and Market Volatility: Developed versus Emerging Stock Markets. The Financial Review. 42. 429-459.

Sewell, Martin. Characterization of Financial Time Series. Research Note. January, 2011.

Black, Fisher. (1976). Studies in Stock Price Volatility Changes. Proceedings of The American Statistical Association, Business and Economic Statistics Section. 177.

Bianco, Simone & Renò, Roberto. (2006). Dynamics of intraday serial correlation in the Italian futures market. Journal of Futures Markets. 26. 61 - 84. 10.1002/fut.20182.

ARTIGOS CIENTÍFICOS

Título: Between Nonlinearities, Complexity, and Noises: An Application on Portfolio Selection Using Kernel Principal Component Analysis

Referência: Entropy, vol. 21, pp. 376, 2019.

Resumo: This paper discusses the effects of introducing nonlinear interactions and noise-filtering to the covariance matrix used in Markowitz’s portfolio allocation model, evaluating the technique’s performances for daily data from seven financial markets between January 2000 and August 2018. We estimated the covariance matrix by applying Kernel functions, and applied filtering following the theoretical distribution of the eigenvalues based on the Random Matrix Theory. The results were compared with the traditional linear Pearson estimator and robust estimation methods for covariance matrices. The results showed that

noise-filtering yielded portfolios with significantly larger risk-adjusted profitability than its non-filtered counterpart for almost half of the tested cases. Moreover, we analyzed the improvements and setbacks of the nonlinear approaches over linear ones, discussing in which circumstances the additional complexity of nonlinear features seemed to predominantly add more noise or predictive performance.

Autores: Yaohao, P.; Albuquerque, P.H.M; Nascimento, I.F.; Machado, J.V.F.

Título: Non-Linear Interactions and Exchange Rate Prediction: Empirical Evidence Using Support Vector Regression

Referência: Applied Mathematical Finance, vol. 26, no. 1, pp. 1-32, 2019.

Resumo: This paper analysed the prediction of the spot exchange rate of 10 currency pairs using support vector regression (SVR) based on a fundamentalist model composed of 13 explanatory variables. Different structures of non-linear dependence introduced by nine different Kernel functions were tested and the predictions were compared to the Random Walk benchmark. We checked the explanatory power gain of SVR models over the Random Walk by applying White's Reality Check Test. The results showed that the majority of SVR models achieved better out-of-sample performance than the Random Walk, but in overall they failed to achieve statistical significance of predictive superiority. Furthermore, we observed that non-mainstream Kernel functions performed better than the ones commonly used in the machine-learning literature, a finding that can provide new insights regarding machine-learning methods applications and the predictability of exchange rates using non-linear interactions between the predictors.

Autores: Yaohao, P.; Albuquerque, P.H.M;

Título: Na era das máquinas, o emprego é de quem? Estimção da probabilidade de automação de ocupações no Brasil

Referência: Texto para Discussão - Instituto de Pesquisa Econômica Aplicada, n. 2457 - Março, 2019.

Resumo: O objetivo deste trabalho foi reproduzir a metodologia de Carl Benedikt Frey e Michael Osborne, de 2017, para estimação das probabilidades de automação das ocupações no Brasil. Essas estimativas são de potencial importância para os formadores de políticas públicas e profissionais por ser passível de nortear a carreira de trabalhadores, bem como definir cursos prioritários que as instituições de ensino deveriam oferecer visando maximizar as oportunidades de emprego no país. A opinião especializada de 69 acadêmicos e profissionais atuantes em aprendizado de máquinas foi levantada para embasar a estimação dessas probabilidades. Os achados apontam que boa parte das ocupações pode ser automatizada nos próximos anos. Ademais, percebe-se que essas profissões

com maior risco de automação apresentam uma tendência de crescimento ao longo do tempo, o que poderá resultar em um elevado nível de desemprego nos próximos anos caso os profissionais e o Estado não se preparem para esse cenário. Os resultados do trabalho, bem como a implementação computacional e as tabelas de conversão utilizadas para o tratamento dos dados, estão disponíveis abertamente em <https://lamfo.shinyapps.io/automacao>

Autores: Albuquerque, P.H.M; Saavedra, C.A.P.B.; Morais, R.L.; Alves, P.F.; Yaohao, P.

Título: The best of two worlds: Forecasting high frequency volatility for cryptocurrencies and traditional currencies with Support Vector Regression.

Referência: Expert Systems with Applications, vol. 1, no. 1, pp. 177-192, 2018.

Resumo: This paper provides an evaluation of the predictive performance of the volatility of three cryptocurrencies and three currencies with recognized stores of value using daily and hourly frequency data. We combined the traditional GARCH model with the machine learning approach to volatility estimation, estimating the mean and volatility equations using Support Vector Regression (SVR) and comparing to GARCH family models. Furthermore, the models' predictive ability was evaluated using Diebold-Mariano test and Hansen's Model Confidence Set. The analysis was reiterated for both low and high frequency data. Results showed that SVR-GARCH models managed to outperform GARCH, EGARCH and GJR-GARCH models with Normal, Student's t and Skewed Student's t distributions. For all variables and both time frequencies, the SVR-GARCH model exhibited statistical significance towards its superiority over GARCH and its extensions.

Autores: Yaohao, P.; Albuquerque, P.H.M; Camboim de Sá, J.M.; Padula, A.J.A.; Montenegro, M.R.

Título: Stock price prediction using support vector regression on daily and up to the minute prices.

Referência: The Journal of Finance and Data Science, vol. 4, no. 3, pp. 183-201, 2018.

Resumo: The purpose of predictive stock price systems is to provide abnormal returns for financial market operators and serve as a basis for risk management tools. Although the Efficient Market Hypothesis (EMH) states that it is not possible to anticipate market movements consistently, the use of computationally intensive systems that employ machine learning algorithms is increasingly common in the development of stock trading mechanisms. Several studies, using daily stock prices, have presented predictive system applications trained on fixed periods without considering new model updates. In this context, this study uses a machine learning technique called Support Vector Regression (SVR) to predict stock prices for large and small capitalisations and in three different markets, employing prices with both daily and up-to-the-minute frequencies. Prediction errors are measured, and the model is compared to the random walk model proposed by the EMH. The results suggest that the SVR has predictive power, especially when using a

strategy of updating the model periodically. There are also indicative results of increased predictions precision during lower volatility periods.

Autores: Bruno Miranda Henrique; Vinicius Amorim Sobreiro; Herbert Kimura.

Título: Segurança Pública e Inteligência Artificial: um estudo georreferenciado para o DF.

Referência: Texto para Discussão - CODEPLAN, n. 33 - Dezembro, 2017.

Resumo: O presente estudo realizou um estudo de georreferenciamento para a ocorrência de crimes violentos e contra o patrimônio com base em dados do ano de 2013 da Pesquisa Distrital por Amostra de Domicílios (PDAD), mediante metodologias inovadoras de aprendizado de máquinas – uma área do conhecimento recente e que vem, cada vez mais, sendo explorada por pesquisas científicas de ponta em nível global em aplicações diversas. Realizou-se um levantamento bibliográfico sistemático da literatura acadêmica correlata, culminando na seleção e fundamentação de 50 variáveis explicativas presentes na base de dados que subsidiam a previsão de ocorrências futuras de crimes nas 31 Regiões Administrativas do Distrito Federal. Para o cômputo das previsões, foram consideradas as metodologias de regressão logística, classificador Bayes ingênuo (Naive-Bayes), árvore de decisão random forest e máquina de vetores de suporte (Support Vector Machine). Cada conjunto de previsões foi avaliado mediante três métricas de desempenho (acurácia, precisão e revocação) e georreferenciadas com a malha digital do Distrito Federal, gerando mapas que ilustram diferentes gradações da propensão à criminalidade de cada Região Administrativa. Os resultados mostram que as previsões atingiram alta aderência aos dados reais, evidenciando que, com base nas variáveis explicativas escolhidas, a ocorrência de crimes no Distrito Federal possui um grau razoável de previsibilidade, visto pelo mapeamento de zonas de alta propensão a crimes. Os resultados obtidos são de grande valia para a sociedade civil e gestores de políticas públicas, os quais podem utilizá-los como insumo para fundamentar a formulação e execução de medidas de segurança pública de intervenção ex-post e políticas de desenvolvimento social de intervenção ex-ante para combater a criminalidade e promover o bem-estar da população nessa região.

Autores: Padula, A.J.A. ; Amorim, F.S. ; Pereira, G.M; Camboim de Sá, J.M.; Oliveira, M.F.F; Montenegro, M.R.; Alves, M.F.; Yaohao, P.

Título: Portfolio Selection with Support Vector Machines in Low Economic: Perspectives in Emerging Market.

Referência: Economic Computation and Economic Cybernetics Studies and Research., 49(4):261-278, nov./dez, 2015.

Resumo: Some of the innovations are presented here are the use of the ranking probability for classification of assets and the balance of the cost by the type of forecast error. The study was conducted on two different samples. The first sample consists on assets that are part of the Ibovespa, considering the portfolio valid from January to April 2015 and this sample was compared to the portfolio generated by SVM and the BOVA11 index fund. The second sample consists of the assets comprising the IBrX-100 index, the portfolio valid from January to April 2015 which similarly was compared with the portfolio generated by SVM and the BRAX11 index fund. In order to evaluate the proposed strategy results were also compared to the free return indicator CDI risk. The portfolio cumulative return of the sample selected by SVM was 94.15%, compared to -14.42% of BOVA11 that replicates the Ibovespa. While the portfolio selected on the portfolio of IBrX100 was 38.25% against 13.86% of BRAX11 index fund. For the period rated the CDI presented a return of 53.75%. The average cumulative return of assets in the study period was in the case of portfolio of the Ibovespa 57.1% and 34.4% for IBrX100 portfolio.

Autores: Marcelino, S.; Henrique, P.A.; Albuquerque, P.H.M.

An empirical overview of nonlinearity and overfitting in machine learning using COVID-19 data

Abstract

In this paper, we applied support vector regression to predict the number of COVID-19 cases for the 12 most-affected countries, testing for different structures of nonlinearity using Kernel functions and analyzing the sensitivity of the models' predictive performance to different hyperparameters settings using 3-D interpolated surfaces. In our experiment, the model that incorporates the highest degree of nonlinearity (Gaussian Kernel) had the best in-sample performance, but also yielded the worst out-of-sample predictions, a typical example of overfitting in a machine learning model. On the other hand, the linear Kernel function performed badly in-sample but generated the best out-of-sample forecasts. The findings of this paper provide an empirical assessment of fundamental concepts in data analysis and evidence the need for caution when applying machine learning models to support real-world decision making, notably with respect to the challenges arising from the COVID-19 pandemics.

Yaohao Peng Mateus Hiro Nagata

Practical machine learning: Forecasting daily financial markets directions

Authors Bruno Miranda Henrique, Vinicius Amorim Sobreiro, Herbert Kimura

Publication date 2023/6/25

Journal Expert Systems with Applications

Pages 120840

Publisher Pergamon

Description

Financial time series prediction has many applications in economics, but producing profitable strategies certainly has a special place among them, a daunting challenge. Statistical and machine learning techniques are intensively researched in the search for a holy grail of stock markets forecasting. However, it is not clear to prospecting researchers how good those popular models are regarding useful predictions on a real scenario. This paper contributes to that discussion, providing decisive evidences contrary to the use of basic out-of-the-box models, specifically Artificial Neural Networks (ANN), Support Vector Machines (SVM), Random Forest (RF) and Naive-Bayes (NB). Results consider optimistic and unreal variables often found in literature, as well as a more close-to-real simulation of the models usage. Specifically, current day closing prices direction forecasting results are contrasted with those on next day ...

Credit and liquidity risk in the determination of spread over business cycles

Authors Eduardo Medeiros Rubik, Vinicius Amorim Sobreiro, Herbert Kimura

Publication date 2021

Journal International Journal of Financial Innovation in Banking

Volume 3

Issue 1

Pages 26-48

Publisher Inderscience Publishers (IEL)

Description

The difference between the amount paid to depositors and the amount charged from borrowers, defined as bank spread, is determined by several factors. Two of them, according to theoretical models, are credit risk and liquidity risk. In this context, we aim to analyse whether these factors change their influence on spread over different periods of business cycles. Brazilian banks were used to evaluate the behaviour of spread at various periods during business cycles by applying the modelling suggested by Aydemir and Guloglu (2017). However, the results show that the influence of credit risk and liquidity risk on the average conditional spread does not change at different periods over business cycles. Furthermore, credit risk and liquidity risk adversely affect spread, contradicting studies on the Brazilian economy, such as Almeida and Divino (2015). Finally, we conclude that spread is procyclical and has a high speed ...

Panic buying and fake news in urban vs. rural England: A case study of twitter during COVID-19

Autores Maged Ali, Lucas Moreira Gomes, Nahed Azab, João Gabriel de Moraes Souza, M Karim Sorour, Herbert Kimura

Data de publicação 2023/8/1

Publicações Technological Forecasting and Social Change

Volume 193

Páginas 122598

Editora North-Holland

Descrição

This paper explores the potential association between the spread of fake news and the panic buying behavior, in urban and rural UK, widely accessible on Twitter since COVID 19 was announced by the WHO as a global pandemic. It describes how consumer's behavior is affected by the content generated over social media and discuss various means to control such occurrence that results in an undesirable social change. The research methodology is based on extracting data from texts on the subject of panic buying and analysing both the total volume and the rate of fake news classification during COVID-19, through crowdsourcing techniques with text-mining and Natural Language Processing models. In this paper, we have extracted the main topics in different phases of the pandemic using term frequency strategies and word clouds as well as applied artificial intelligence in exploring the reliability behind online ...

Are fintechs worth investing? case study for startup using real option theory

Autores João Carlos Félix Souza, João Gabriel De Moraes Souza, Peng Yao Hao

Data de publicação 2021

Publicações International Journal of Development Research

Volume 11

Edição 04

Páginas 45999-46007

Descrição

A startup, according to the Brazilian Startup Association (2015), is defined as “a technology-based company with a repeatable and scalable business model, which has elements of innovation, and operates under conditions of extreme uncertainty”. According to Corder and Salles-Filho (2006, p. 37) “venture capital is a financing mechanism that proposes to finance innovation and, in this sense, differs from traditional finance options”. Typical characteristics of startups include the high-risk nature of the business, coupled with the difficulty in securing loans from the financial market. As a result, financing does not normally involve banks but venture capital instead (Chorev, 2006). Startup founders do not

seek debt-generating financing, but rather investors who want to participate in the business, due to the perceived attractiveness of the company (Colombo, 2002). According to the Brazilian Association of Private Equity and Venture Capital ABVCAP (2014), the national market has seen remarkable contributions from angel investors, these being autonomous individuals who choose to invest their own financial resources into private companies (Osnabrugge, 2000; Bacher & Guild, 1996). The low survival rate for this type of company has been statistically demonstrated time and again (Arruda et al. 2014), as they are typically short-lived.

For Rencher (2012), startups will normally undergo a period known as the “Valley of Death”, which is the period between the creation of the company and the moment when its cash flow becomes positive. Venture capital is one of the main contributing factors in aiding the company to cross this threshold. Among the ...

A comprehensive review of automatic text summarization techniques: method, data, evaluation and coding

Autores Daniel O Cajueiro, Arthur G Nery, Igor Tavares, Maísa K De Melo, Silvia A dos Reis, Li Weigang, Victor RR Celestino

Data de publicação 2023/1/4

Fonte arXiv preprint arXiv:2301.03403

Descrição

We provide a literature review about Automatic Text Summarization (ATS) systems. We consider a citation-based approach. We start with some popular and well-known papers that we have in hand about each topic we want to cover and we have tracked the "backward citations" (papers that are cited by the set of papers we knew beforehand) and the "forward citations" (newer papers that cite the set of papers we knew beforehand). In order to organize the different methods, we present the diverse approaches to ATS guided by the mechanisms they use to generate a summary. Besides presenting the methods, we also present an extensive review of the datasets available for summarization tasks and the methods used to evaluate the quality of the summaries. Finally, we present an empirical exploration of these methods using the CNN Corpus dataset that provides golden summaries for extractive and abstractive methods.

Machine learning models for forecasting power electricity consumption using a high dimensional dataset

Autores Pedro C Albuquerque, Daniel O Cajueiro, Marina DC Rossi

Data de publicação 2022/1/1

Publicações Expert Systems with Applications

Volume 187

Páginas 115917

Editora Pergamon

Descrição

We use regularized machine learning models to forecast Brazilian power electricity consumption for short and medium terms. We compare our models to benchmark specifications such as Random Walk and Autoregressive Integrated Moving Average. Our results show that machine learning methods, especially Random Forest and Lasso Lars, give more accurate forecasts for all horizons. Random Forest and Lasso Lars managed to keep up with the trend and the seasonality for various time horizons. The gain in predicting PEC using machine learning models relative to the benchmarks is considerably higher for the very short-term. Machine learning variable selection further shows that lagged consumption values are extremely important for very short-term forecasting due to the series high autocorrelation. Other variables such as weather and calendar variables are important for longer time horizons.

Few-shot Approach for Systematic Literature Review Classifications.

Autores Maísa Kely de Melo, Allan Victor Almeida Faria, Li Weigang, Arthur Gomes Nery, Flávio Augusto R de Oliveira, Ian Teixeira Barreiro, Victor Rafael Rezende Celestino

Data de publicação 2022

Fonte WEBIST

Páginas 33-44

Descrição

Systematic Literature Review (SLR) studies aim to leverage relevant insights from scientific publications to achieve a comprehensive overview of the academic progress of a specific field. In recent years, a major effort has been expended in automating the SLR process by extracting, processing, and presenting the synthesized findings. However, implementations capable of few-shot classification for fields of study with a smaller amount of material available seem to be lacking. This study aims to present a system capable of conducting automated systematic literature reviews on classification constraint by a few-shot learning. We propose an open-source, domain-agnostic meta-learning SLR framework for few-shot classification, which has been validated using 64 SLR datasets. We also define an Adjusted Work Saved over Sampling (AWSS) metric to take into account the class imbalance during validation. The initial results show that AWSS@ 95% scored as high as 0.9 when validating our learner with data from 32 domains (just 16 examples were used for training in each domain), and only four of them resulted in scores lower than 0.1. These findings indicate significant savings in screening time for literature reviewers.