

## Projeto de pesquisa de Iniciação Científica

### Insper - Engenharia de Computação

**Aluno:** Phelipe Müller

**Professor:** Fábio Ayres

**Tema:** Aplicação de técnicas de *machine learning* para o mercado financeiro

## Introdução

Na década de 90 houve uma explosão de artigos que estudavam técnicas de inteligência artificial aplicadas ao mercado financeiro (Wei Huang 2005) (Kyoung-Jae Kim 2002). Conforme os resultados foram apresentados, observou-se que a eficiência das metodologias era muito abaixo do esperado, e, portanto, houve uma diminuição da popularidade de tais abordagens. Com a recente popularização das técnicas modernas de *machine learning* houve um ressurgimento de pesquisas sobre o tema (Krollner 2010).

Dentre as técnicas mais utilizadas de *machine learning* estão as máquinas de vetor de suporte (*Support Vector Machines - SVM*) (Vapnik 1998). Huang et al (2005) aplicaram SVM para a previsão de direção de movimento do índice NIKKEI 225, e compararam a performance da SVM com os métodos de Análise Discriminante Linear, Análise Discriminante Quadrática, e Redes Neurais de Elman. Os autores concluíram que SVM apresentou melhor desempenho que as outras técnicas investigadas.

Este trabalho visa investigar a aplicação de técnicas consolidadas de *machine learning*, especificamente SVM e regressão logística, ao mercado brasileiro de ações. Será feito um levantamento de possíveis entradas de indicadores mais adequados à realidade brasileira, e uma comparação de desempenho entre as técnicas de *machine learning* propostas através de figuras de mérito do *backtesting* (Aldridge 2013)

## Metodologia

Máquinas de Vetor de Suporte (do inglês *Support Vector Machines – SVM* [Vapnik 1998]) são técnicas de aprendizado supervisionado que buscam estabelecer a separação entre observações de classes diferentes através da maximização da margem de separação entre classes. Os pontos do conjunto de treinamento que pertencem à fronteira da região de separação são chamados de vetores de suporte. Considere o exemplo abaixo onde amostras fictícias de peixes de duas espécies diferentes (ou seja, duas classes distintas) são exibidas em um gráfico de dispersão (Figura 1):

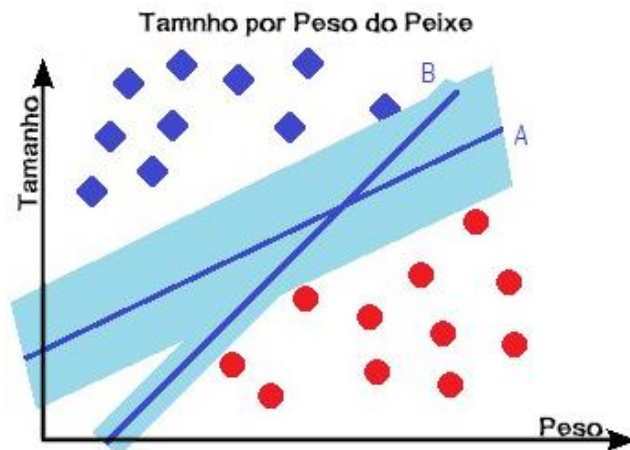


Figura 1: Gráfico de exemplo de SVM

O objetivo do algoritmo de treinamento da SVM é encontrar o hiperplano de separação entre classes que maximize a margem de separação entre os vetores mais próximos ao hiperplano, desta forma minimizando a chance de uma nova amostra seja classificada na classe errada. Na figura 1 observamos que o hiperplano A maximiza a margem de separação, enquanto B é sub-ótimo.

O desempenho da técnica SVM será comparado com uma técnica mais clássica: a regressão logística. Trata-se de uma técnica de predição estatística onde se procura construir um modelo de predição para um problema de classificação binária. O modelo apresenta a forma da equação 1, vista abaixo:

$$F(\mathbf{x}) = \frac{1}{1 + e^{-(\beta_0 + \sum_{i=1}^N \beta_i x_i)}} \quad (1)$$

onde  $\mathbf{x} = (x_1, x_2, \dots, x_N)$  são as entradas do modelo, e  $\beta_i, i \in [0, N]$ , são os seus parâmetros.

Para comparar as duas técnicas serão utilizadas figuras de mérito padrão de análise de *backtesting*, tal qual: *Max Profit*, *Max Drawdown*, *Hitting Ratio*, *Average Gain/Loss*, *Sharpe Ratio*, entre outros.

## Resultados Esperados

Neste trabalho investigaremos o desempenho de uma técnica moderna de *machine learning* e de uma técnica clássica de predição estática no contexto do mercado brasileiro de ações. Com isso esperamos avaliar a efetividade da SVM perante aquela da regressão logística, e como ambas tentam explicar o comportamento do mercado. Outro resultado esperado deste estudo será uma análise da sensibilidade do desempenho dos métodos supracitados em relação às diversas entradas.

## Cronograma

ATIVIDADES	JAN	FEV	MAR	ABR	MAI	JUN	JUL	AGO	SET	OUT	NOV	DEZ
Estudo do Problema	o	o										
Coleta e Pré-processamento de dados			o									
Desenvolvimento				o	o	o	o	o				
Análise dos resultados									o	o	o	
Conclusão e Relatório												o

## Referências

Krollner, Bruce Vanstone, and Gavin Finnie. (2010) "**Financial time series forecasting with machine learning techniques: A survey**" European Symposium on Artificial Neural Networks: Computational and Machine Learning. Bruges, Belgium. Apr. 2010.

Wei Huang, Yoshiteru Nakamori, Shou-Yang Wang. (2005) "**Forecasting stock market movement direction with support vector machine**" Computers & Operations Research 32 (2005) 2513 – 2522

Kyoung-Jae Kim (2002) "**Financial time series forecasting using support vector machines**" Neurocomputing 55 (2003) 307 – 319

Vladimir N. Vapnik. (1998) "**Statistical Learning Theory**". Wiley.

Irene Aldridge (2013) "**High-Frequency Trading**". Wiley.