



PUC

ISSN 0103-9741

Monografias em Ciência da Computação
n° 06/09

Aprendizado de Máquina para Intraday Traders do Mercado Acionário

**Leandro Guimarães Marques Alvim
Ruy Luiz Milidiú**

Departamento de Informática

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO

RUA MARQUÊS DE SÃO VICENTE, 225 - CEP 22451-900

RIO DE JANEIRO - BRASIL

Aprendizado de Máquina para Intraday Traders do Mercado Acionário

Leandro Guimarães Marques Alvim, Ruy Luiz Milidiú

leandrouff@gmail.com, milidiu@inf.puc-rio.br

Abstract. In this paper we investigate current strategies for Automated Intraday Traders of the Market and the state of the art of Machine Learning techniques. Based on this studies, we propose a new Intraday Trader which buys when prices are beginning to rise and sells when they start to fall, based on VWAP information of market stock prices. The trader strategy uses two market advisors: an interday detector of buy/sell candidate days and an intraday detector of buy/sell instants of candidate days. For the interday detector we use a SVM predictor to classify whether the next day VWAP will rise or fall. For the intraday detector, we use a committee of SVR and PLS to forecast daily VWAP, aiming at buying/selling below/above the VWAP, respectively. For testing the new trader, we use the LearnTrade Framework that supports: financial time series predictors, simulation resources for automated traders and a dataset of Bovespa. Promising results were obtained with maximum profit of 75% for stocks with positive mean returns and 3.7% for stocks with negative mean returns.

Keywords: Intraday Traders, Aprendizado de Máquina, LearnTrade.

Resumo. Neste trabalho, investigamos as recentes estratégias para Automated Intraday Traders do mercado acionário e técnicas do estado da arte de Aprendizado de Máquina. Com base nestes estudos, propomos uma nova estratégia para traders baseada na informação do VWAP. Nesta estratégia uma compra é feita quando o VWAP do ativo está para subir e uma venda é feita quando o VWAP do ativo está para descer. Para esta estratégia, utilizamos dois detectores de entrada e saída do mercado: um detector interday para alertar um dia candidato de compra/venda e um detector intraday para alertar, dentro do dia candidato, um instante candidato de compra/venda. Para o detector interday, utilizamos um preditor SVM para classificar uma alta/queda de VWAPs futuros de forma a escolher o dia candidato. Para o detector intraday utilizamos um comitê do SVR com PLS para predição do VWAP do dia de forma a escolher o instante do dia em que o preço encontra-se abaixo/acima do VWAP para compra/venda, respectivamente. Para os experimentos, utilizamos o framework LearnTrade que provê preditores de séries temporais, um ambiente de simulação para Automated Intraday Traders e um conjunto de dados da Bovespa. Os resultados obtidos são promissores, com ganhos de até 75% para ativos de retorno médio positivo e de até 3.7% para ativos de retorno médio negativo.

Palavras-chave: Intraday Traders, Aprendizado de Máquina, LearnTrade.

Responsável por publicações:

Rosane Teles Lins Castilho
Assessoria de Biblioteca, Documentação e Informação
PUC-Rio Departamento de Informática
Rua Marquês de São Vicente, 225 - Gávea
22451-900 Rio de Janeiro RJ Brasil
Tel. +55 21 3527-1516 Fax: +55 21 3527-1530
E-mail: bib-di@inf.puc-rio.br
Web site: <http://bib-di.inf.puc-rio.br/techreports/>

1 Introdução

O interesse no mercado financeiro vem aumentando nas últimas décadas incluindo operadores de mercado, formuladores de políticas de mercado e acadêmicos. Tal interesse focaliza-se em três domínios: predição, tomada de decisões e gerenciamento de riscos.

A predição do mercado financeiro (preços de ações e índices financeiros) tem sido considerada como uma das aplicações mais desafiadoras e vem sendo estudada extensivamente por pesquisadores e profissionais de mercado. A predição e negociação no mercado de ações obtiveram popularidade nunca antes atingida nos principais mercados financeiros mundiais. Há basicamente duas razões para o sucesso de ferramentas de predição do mercado financeiro. Primeiramente, elas fornecem um meio efetivo de proteção dos investidores contra riscos potenciais do mercado. Em segundo lugar, elas criam novas oportunidades de retorno de lucros para especuladores. A capacidade de predizer com precisão os índices do mercado de ações tem implicações e significados profundos tanto para pesquisadores quanto para os profissionais de mercado [Leung et al., 2000].

É difícil realizar previsões de curto e longo prazo eficientemente. Vasta literatura se acumulou durante vários anos, com relação à validade do EMH (Efficient Market Hypothesis), relativo ao mercado de ações. Uma implicação importante do EMH é que nenhum investidor está em posição de conseguir oportunidades de lucro não exploradas pela predição de preços futuros baseado nos do passado. Por outro lado, um grande número de pesquisadores, investidores, analistas, operadores e outros, usa técnicas diversas para predizer o índice de ações e os preços das mesmas, contrariamente ao que aconselham muitos estudos acadêmicos, que fazem a hipótese de que a previsão de preços de ações do mercado é fútil.

Outro domínio, principal motivador deste trabalho, é o de tomada de decisões para Intraday Traders. Em contraposição às análises técnica e fundamentalista, em que há a necessidade de um acompanhamento e análise de um especialista, na área de Intraday Traders não há a necessidade deste acompanhamento. A partir de dados históricos da bolsa de valores, o Trader deve inferir quais operações (compra e venda) realizar, quanto e quando operar. Nesta área de Intraday Traders, podem ser encontrados trabalhos acadêmicos interessantes em [Silaghi et al., 2005, Sherstov and Stone, 2005] em que são abordados diversos tipos de técnicas utilizadas para tomada de decisões.

Na seção 2, investigamos o estado da arte de Intraday Traders e métodos de Aprendizado de Máquina. Nas seções 3 e 4, estudamos os problemas de classificação, baseado nos artigos [Gunn, 1998] e [Ben-Hur and Weston,], e regressão, respectivamente. Ainda nestas duas seções estudamos os algoritmos de Aprendizado de Máquina: SVM [Boser et al., 1992], baseados também nos artigos [Gunn, 1998] e [Ben-Hur and Weston,], e PLS [Wold et al., 1983a, Wold et al., 1983b]. Na seção 5, são descritos dois preditores de Aprendizado de Máquina para Intraday Traders. Na seção 6, revisamos algumas estratégias para Intraday Traders e propomos uma nova estratégia baseada no VWAP cujo objetivo é a maximização do lucro. Na seção 7, descrevemos a ferramenta utilizada para os experimentos, que contém os preditores e o simulador do Trader. Na seção 8, mostramos experimentos e análises dos resultados obtidos a partir das novas estratégias. Por fim, nas seções 9 e 10 concluímos o trabalho e propomos algumas idéias para trabalhos futuros.

2 Preditores Financeiros e Intraday Traders

Em [Silaghi et al., 2005] são revisadas algumas estratégias de Intraday Traders baseadas em preço e em livro de ordem. O estudo mostra um comparativo das estratégias que foram desenvolvidas e experimentadas no framework PXS (Penn Exchange Simulator). Os melhores resultados obtidos foram de estratégias combinadas e que utilizavam, além de informações do preço, informações do livro de ordens.

Em [Sherstov and Stone, 2005], é feito um estudo comparativo de três estratégias para Intraday Traders que foram desenvolvidas e experimentadas no framework PXS (Penn Exchange Simulator). Dentre as estratégias tem-se: Aprendizado por reforço, MM (Market Making Agent) e Acompanhamento de Tendência de Preços (Trend Following). A estratégia MM demonstrou-se superior e foi a ganhadora da competição PLAT (Penn-Lehman Automated Trading) de 2003.

Ao contrário de outras abordagens, que utilizam previsões de preço, em [Choey and Weigend, 1997] foi proposta uma estratégia não-linear baseada na maximização do Índice de Sharpe. Neste trabalho um novo modelo é expresso por uma rede neural cuja saída binária corresponde a uma posição de risco ou livre de risco. Ainda neste trabalho, o algoritmo de treinamento de gradiente descendente foi modificado de forma a incorporar o risco. Resultados a partir de experimentos com séries artificiais e reais foram expressivos.

Em [Briza and Jr, 2008] é utilizado o método de Otimização por Enxame de Partículas (PSO) a partir de um conjunto de dados de preço diário e índices de análise técnica. O objetivo da estratégia foi o de maximizar o retorno percentual e minimizar o risco. O estudo mostrou que a estratégia superou, na maioria dos experimentos, uma estratégia que utiliza somente índices de análise técnica.

Em [Leung et al., 2000] foi realizado um estudo que indica que o modelo SVM (Support Vector Machine) atinge maior precisão preditiva e melhora a qualidade da previsão em comparação com outros modelos abordados no estudo, tais como redes neurais, regressão de floresta randômica e um modelo ARIMA linear. O modelo SVM pode ser usado como uma ferramenta alternativa de previsão para dividendos do Nifty Index e conduz a melhores resultados, de acordo com medidas tradicionais de precisão para previsões, tais como erros mínimos quadráticos médios e com critérios financeiros, tais como medidas ajustadas baseadas em risco para dividendos.

Em [Boyarshinov, 2005] um extenso estudo pode ser encontrado sobre Aprendizado de Máquina em Finanças. O estudo aborda geração de conjunto ótimo para treinamento supervisionado, métricas de avaliação e algoritmos de Aprendizado de Máquina para Intraday Traders

3 Classificação

3.1 Classificadores Lineares

Considere um problema de separar um conjunto de dados $D = \{(x_i, y_i)\}_{i=1}^n$, $x \in \mathbb{R}^d$, $y \in \{-1, 1\}$ em duas classes distintas. Representamos como \mathbf{x} a matriz que contém n exemplos x_i do problema. Cada exemplo x_i é um vetor linha \mathbf{x} com d dimensões e que possui uma resposta y_i associada. Podemos representar um classificador linear pelo produto

escalar $\mathbf{w}^T \mathbf{x}$ em que temos o hiperplano de separação:

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b \quad (1)$$

tal que \mathbf{w} é o vetor de pesos, que em um espaço de duas dimensões é coeficiente angular de uma reta. Já b , ou bias, corresponde ao coeficiente linear da mesma. Exemplos são classificados como positivos quando $f(\mathbf{x}) > 0$. Na figura 1 os pontos que estão acima do hiperplano satisfazem $f(\mathbf{x}) > 0$. Analogamente, quando $f(\mathbf{x}) < 0$ temos a classificação dos exemplos negativos em que os pontos encontram-se abaixo do hiperplano. Quando a divisão dos exemplos de um problema é feita por um hiperplano, então, estamos lidando com um problema de classificação linear. Caso contrário, temos um problema de classificação não-linear.

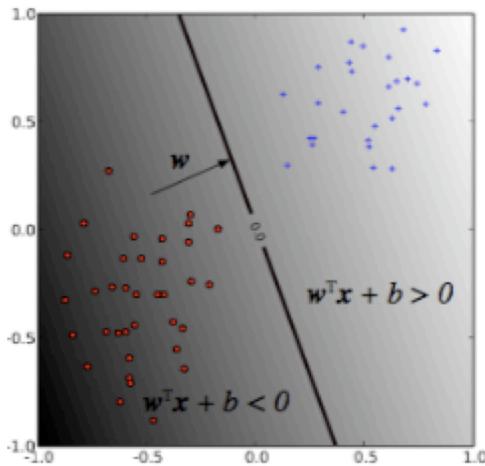


Figure 1: Classificador Linear.

3.2 Classificadores Não-Lineares

Uma das vantagens da utilização de modelagens lineares é a simplicidade do algoritmo de treinamento. Para um conjunto elevado de exemplos, o treinamento é eficiente em comparação a um classificador baseado numa modelagem não-linear. Em vista disto, porque não aproveitar o maquinário de um classificador linear para produzir um classificador não-linear? Veremos a seguir, duas abordagens que calcadas nesta idéia. A primeira abordagem é a de mapear explicitamente os dados para uma dimensão maior, atuando no espaço de características. A seguir temos um exemplo:

$$f(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) + b \quad (2)$$

tal que $\phi(\mathbf{x})$ é uma função não-linear. Por exemplo, se tivéssemos trabalhando em um espaço de duas dimensões teríamos um mapeamento da seguinte forma:

$$\phi(x) = (x_1^2, \sqrt{2}x_1x_2, x_2^2)^T \quad (3)$$

Com este mapeamento teríamos:

$$\mathbf{w}^T \phi(x) = w_1x_1^2 + w_2\sqrt{2}x_1x_2 + w_3x_2^2, \quad (4)$$

tal que, esta seria a nova representação dos dados em uma dimensão superior. Porém, este tipo de prática leva a um aumento quadrático de memória para armazenamento das características. O tempo torna-se quadrático, também, para cômputo das funções não-lineares. Para um problema representado num espaço elevado de atributos, estas duas consequências, muitas vezes, tornam a modelagem impraticável. Uma alternativa a este problema é o emprego de Funções de Núcleo. Ao invés de mapear e computar explicitamente os atributos, as Funções de Núcleo utilizam-se de um artifício bastante interessante. Seja o vetor de pesos representado por uma combinação linear dos exemplos na forma:

$$\mathbf{w} = \sum_{i=1}^n \alpha_i \mathbf{x}_i^T \mathbf{x} \quad (5)$$

e aplicando uma função não-linear no espaço de características:

$$f(\mathbf{x}) = \sum_{i=1}^n \alpha_i \phi(x_i)^T \phi(\mathbf{x}) + b \quad (6)$$

teríamos uma representação em termos da variável α , que é conhecida como Representação Dual. Porém, como descrito anteriormente, este tipo de modelagem é inviável para um elevado número de atributos. Seja a Função de Núcleo:

$$k(\mathbf{x}, \mathbf{x}') = \phi(x)^T \phi(x') \quad (7)$$

teremos:

$$f(\mathbf{x}) = \sum_{i=1}^n \alpha_i k(\mathbf{x}, \mathbf{x}'_i) + b \quad (8)$$

Como veremos a seguir, na equação (9), a grande vantagem da utilização de Funções de Núcleo, além de resolver o problema da não-linearidade, é a eficiência do seu cálculo em comparação ao mapeamento explícito de funções não-lineares. Como exemplo temos:

$$\begin{aligned} \phi(x)\phi(z) &= (x_1^2, \sqrt{2}x_1x_2, x_2^2)(z_1^2, \sqrt{2}z_1z_2, z_2^2) \\ &= x_1^2z_1^2 + 2x_1x_2z_1z_2 + x_2^2z_2^2 \\ &= (\mathbf{x}^T \mathbf{z})^2 \end{aligned} \quad (9)$$

que para resolver $\phi(x)\phi(z)$ basta computar o quadrado do produto interno $(\mathbf{x}^T \mathbf{z})^2$. Existem vários tipos de Funções de Núcleo que podem ser utilizadas. Dentre elas, as funções Polinomial, Gaussiana de Base Radial, Sigmóide e outras. Estudos sugerem a utilização de Funções de Base Radial, que demonstraram acurácia superior às demais Funções de Núcleo. Na equação (10) tem-se a função Gaussiana de Base Radial:

$$k(\mathbf{x}, \mathbf{x}') = \exp(-\gamma \|\mathbf{x} - \mathbf{x}'\|^2) \quad (10)$$

Em que $\gamma > 0$ é o parâmetro que controla a largura da Gaussiana. Cabe ressaltar que, para que seja explorado o potencial das Funções de Núcleo quanto a não-linearidade o treinamento do algoritmo de Aprendizado de Máquina deve ser ajustado para tal.

3.3 Máquina de Vetores de Suporte (SVM)

Máquina de Vetores de Suporte são o estado da arte para algoritmos classificadores e regressivos. O SVM foi desenvolvido por Vapnik [Boser et al., 1992] e colaboradores do laboratório AT&T no final da década de 90. SVM é amplamente utilizado na área de bioinformática e recentemente vem sendo utilizado na área de finanças para predição de séries

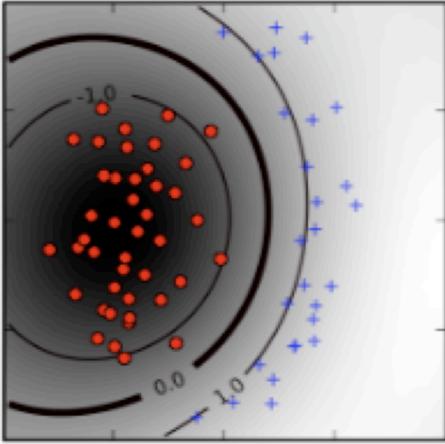


Figure 2: Classificador Não-Linear.

temporais com excelente desempenho Mueller et al.[Mueller et al., 1997]. As principais motivações do desenvolvimento do SVM foram elevada capacidade de generalização, robustez em grandes dimensões e convexidade da função objetivo. Dentre estas três características, devemos ressaltar a terceira. A aplicação de SVM para um dado problema implica na otimização de uma função quadrática que possui apenas um mínimo local, sendo que este é global. Operar em uma função objetivo com este tipo de característica é uma grande vantagem. Isto porque o excesso de mínimos locais dificulta o processo de convergência do algoritmo de aprendizado.

3.3.1 Classificação de Máxima Margem

Seja um Classificador Linear cujo hiperplano de separação é mostrado na figura 3 e os pontos circulados os Vetores de Suporte, denomina-se margem $m(h)$ a distância entre os Vetores de Suporte e o hiperplano h de separação. Numa representação geométrica denotaremos x_+ e x_- como sendo os pontos mais próximos do hiperplano de separação, \mathbf{w} o vetor de suporte com norma $\|\mathbf{w}\| = \sqrt{\mathbf{w}^T \mathbf{w}}$ e o \hat{w} é o vetor unitário na direção de \mathbf{w} . Os hiperplanos de contorno, paralelos a h , são representados da forma:

$$h_+ = \mathbf{w}^T \mathbf{x}_+ + b = a \quad (11)$$

para pontos positivos e

$$h_- = \mathbf{w}^T \mathbf{x}_- + b = -a \quad (12)$$

para pontos negativos, sendo $a > 0$ para as duas equações. Fazendo $a = 1$, somando-se os dois hiperplanos e dividindo por $\|\mathbf{w}\|$ teremos, finalmente:

$$m(h) = \frac{1}{2} \hat{w}^T (\mathbf{x}_+ - \mathbf{x}_-) = \frac{1}{\|\mathbf{w}\|} \quad (13)$$

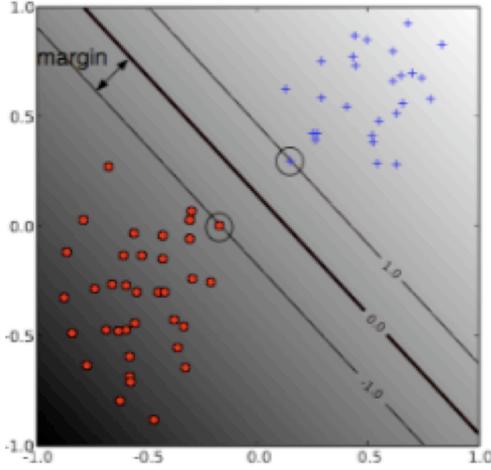


Figure 3: Margem.

3.3.2 O Problema Primal

Para uma classificação de máxima margem o SVM busca maximizar $\frac{1}{\|\mathbf{w}\|}$, ou seja, minimizar $\|\mathbf{w}\|^2$. O que nos leva ao problema:

$$\begin{aligned} & \text{minimizar}_{\mathbf{w}, b} \frac{1}{\|\mathbf{w}\|^2} \\ & \text{sujeito a: } y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, i = 1, \dots, n. \end{aligned} \quad (14)$$

Esta formulação é conhecida como Margem Rígida. Isto porque possui restrições rigorosas, que não consideram os exemplos que estão além de sua margem. Para problemas de classificação não-linear este tipo de formulação acaba por ser um entrave. Em vista disto, queremos que as restrições possam maximizar a margem desclassificando alguns exemplos que estão além da margem. Note que este tipo de flexibilidade na restrição evita que “outliers” atrapalhem a generalização do modelo, pois podem ser os exemplos desclassificados. Assim teremos a nova formulação, adaptada para:

$$\begin{aligned} & \text{minimizar}_{\mathbf{w}, b} h(\mathbf{w}, \epsilon) = \frac{1}{\|\mathbf{w}\|^2} \\ & \text{sujeito a: } y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \epsilon_i, i = 1, \dots, n. \end{aligned} \quad (15)$$

Em que $0 \leq \epsilon \leq 1$ é a margem de erro. Quando $\epsilon < 0$ tem-se uma Variável de Folga, que é quando um exemplo encontra-se na margem de erro. Quando $\epsilon > 1$, então temos uma desclassificação, ou seja, um exemplo que ultrapassou a margem. Com isto, o objetivo agora é o de classificar os exemplos de modo a maximizar a margem minimizando a quantidade de Variáveis de Folga. Para isto, devemos introduzir $C \sum_{i=1} \epsilon_i$ na restrição, levando ao problema de otimização:

$$\begin{aligned} & \text{minimizar}_{\mathbf{w}, b} h(\mathbf{w}, \epsilon) = \frac{1}{\|\mathbf{w}\|^2} + C \sum_{i=1} \epsilon_i \\ & \text{sujeito a: } y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \epsilon_i, \epsilon_i \geq 0 \end{aligned} \quad (16)$$

Este tipo de formulação é chamado de margem flexível e foi introduzido por Cortes em [Cortes and Vapnik, 1995]. A solução para este problema, da equação (16), pode resolvida

na forma:

$$h(w, b, \alpha, \beta) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \epsilon_i - \sum_{i=1}^n \alpha_i (y_i (\mathbf{w}^T \mathbf{x}_i + b) - 1 + \epsilon_i) + \sum_{j=1}^n \beta_j \epsilon_j, \quad (17)$$

tal que, α e β são os Multiplicadores de Lagrange. O Lagrangiano $h(w, b, \alpha, \beta)$ deve ser minimizado com relação a w, b, x e maximizado com relação a α e β .

3.3.3 O Problema Dual

Para resolver o Problema Dual transformamos o Lagrangiano, na equação (17), em:

$$\max_{\alpha} W(\alpha, \beta) = \max_{\alpha, \beta} (\min_{w, b, x} h(w, b, \alpha, \epsilon, \beta)), \quad (18)$$

tal que, que o mínimo de w, b, ϵ com relação ao Lagrangiano h será dado por:

$$\begin{aligned} \frac{\partial h}{\partial b} = 0 &\Rightarrow \sum_{i=1}^n \alpha_i y_i = 0 \\ \frac{\partial h}{\partial w} = 0 &\Rightarrow w = \sum_{i=1}^n \alpha_i y_i x_i \\ \frac{\partial h}{\partial \epsilon} = 0 &\Rightarrow \alpha_i + \beta_i = C \end{aligned} \quad (19)$$

em que $\alpha > 0$ são os os pontos que encontram-se na margem. Estes pontos são conhecidos com os Vetores de Suporte. A partir das equações 17,18 e 19 resolvemos:

$$\begin{aligned} \text{maximizar } & \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i y_j \alpha_i \alpha_j \mathbf{x}_i \mathbf{x}_j \\ \text{sujeito a : } & \sum_{i=1}^n y_i \alpha_i = 0, 0 \leq \alpha_i \leq C. \end{aligned} \quad (20)$$

4 Regressão

4.1 O Problema de Regressão

O problema de regressão consiste em modelar variáveis explicadoras (variáveis independentes) de forma a estimar uma ou mais variáveis contínuas de resposta (variáveis dependentes) de um conjunto de dados do problema. Seja \mathbf{x} conjunto de variáveis independentes e \mathbf{y} as variáveis dependentes, em que \mathbf{x}_i e \mathbf{y}_i correspondem ao i -ésimo vetor linha de \mathbf{x} e \mathbf{y} respectivamente, podemos explicar \mathbf{y}_i a partir de:

$$y_i = \phi(x_i) + \epsilon_i \quad (21)$$

que na verdade que dizer que y é uma função dependente de $\phi(\mathbf{x}_i)$ com adição do que podemos chamar de resíduo de y_i , expresso por ϵ_i . O problema de regressão consiste, fundamentalmente em estimar $\phi(x_i)$, ou seja, encontrar $\hat{\phi}(x_i)$ que satisfaça 21.

4.2 Regressão por Vetores de Suporte (SVR)

O SVM pode ser aplicado para o problema de regressão com o emprego de uma Função de Erro alternativa [Smola and Schölkopf, 2004]. Existem vários tipos de Funções de Erro: Quadrática, Laplaciana, Huber, ϵ -insensível e outras. A função Quadrática é equivalente ao o critério do erro de mínimos quadrados. A função Laplaciana é menos sensível a "Outliers" do que a Quadrática. As funções Quadrática e Laplaciana devem ser utilizadas quando se conhece a distribuição dos dados. Já para uma distribuição desconhecida dos dados, deve ser utilizada a função de Huber. Por fim, a função ϵ -insensível, que foi proposta por Vapnik. Esta função consegue trabalhar com um conjunto de vetores de suporte esparsos. A seguir descreveremos o SVR com o emprego da Função de Erro ϵ -insensível. Seja a função regressão:

$$h(w, \epsilon) = \frac{1}{2} \|w\|^2 + C \sum \epsilon_i^- \epsilon_i^+, \quad (22)$$

em que, C é uma constante e $\epsilon_i^- \epsilon_i^+$ são as Variáveis de Folga. Seja a função ϵ -insensível

$$L(y)_\epsilon = \begin{cases} 0 & \text{se } |f(x) - y| < \epsilon \\ |f(x) - y| - \epsilon & \text{caso contrário} \end{cases} \quad (23)$$

podemos resolver na forma:

$$\begin{aligned} \max_{\alpha, \alpha^*} W(\alpha, \alpha^*) &= \max_{\alpha, \alpha^*} - \\ & - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) x_i x_j + \\ & + \sum \alpha_i (y_i - \epsilon)_{i=1}^l - \alpha_i^* (y_i + \epsilon) \end{aligned} \quad (24)$$

sujeito a $0 \leq \alpha_i, \alpha_i^* \leq C, i = 1, \dots, l$ e

$$\sum (\alpha_i - \alpha_i^*)_{i=1}^l = 0,$$

tal que, α_i e α_i^* são os Multiplicadores de Lagrange. Resolvendo esta equação teremos a solução para a função de regressão dada por:

$$\begin{aligned} \hat{w} &= \sum (\alpha_i - \alpha_i^*) x_{i=1}^l \\ \hat{b} &= -\frac{1}{2} \hat{w} (x_r + x_s) \end{aligned} \quad (25)$$

sujeito a: $\alpha_i \alpha_i^* = 0 \quad i = 1, \dots, l$

4.3 Regressão Baseada em Projeção

Quando a forma funcional ϕ , da equação (21), é conhecida, podemos estimar ϕ a partir do método de mínimos quadrados não-linear ou até mesmo linear, sendo que, quando o número de características de x é elevado, a dificuldade do problema de regressão aumenta demasiadamente e torna-se impraticável a regressão. A solução para este tipo de problema foi subdividir o modelo em modelos menores:

$$\phi(\hat{x}_i) = \sum_{j=1}^n r_j (w_j^T x_i) \quad (26)$$

em que w_j^T é um vetor unitário denominado de Vetor de Carga e r_j é uma função de suavização denominada de Função de Ridge. Podemos chamar este método de Regressão Baseada em Projeção. A seguir, introduziremos dois métodos baseados em projeção que podem ser encontrados em diversos trabalhos na literatura acadêmica.

4.4 Análise Por Componentes Principais (PCA)

O objetivo central do PCA Apud [Shlens, 2005] é o de resumir a informação dos dados entrada. Como consequências deste resumo dos dados temos: eliminação de ruído, redução da dimensionalidade e o fato da informação poder ser melhor entendida. Este resumo é feito decompondo a matriz original em um produto de duas matrizes menores (submodelos) que são chamadas de Carga e Pontuação. Com estes submodelos, a matriz dos dados pode ser expressa por uma combinação linear destes. Sobre uma ótica geométrica, o PCA tenta achar os hiperplanos se “encaixem” melhor entre os exemplos. Estes “encaixes” são obtidos através da minimização da soma dos erros quadráticos de cada exemplo. Seja

$$\mathbf{x}^T \mathbf{x} = Cov(\mathbf{x}, \mathbf{x}) = Var(\mathbf{x}) \quad (27)$$

uma matriz quadrada, simétrica e com suas colunas normalizadas (Z-Score). Através da Decomposição Espectral de Jordan, chegamos a:

$$\mathbf{x}^T \mathbf{x} = \mathbf{w} \Lambda \mathbf{w}^T \quad (28)$$

em que $\mathbf{w}_{(p,p)}$ é uma matriz de autovetores $\mathbf{x}^T \mathbf{x}$ e $\Lambda = diag(\lambda_1, \dots, \lambda_p)$ cujos elementos não nulos são os autovalores de $\mathbf{x}^T \mathbf{x}$ e estão ordenados como $\lambda_1 \geq \lambda_2 \dots \geq \lambda_p \geq 0$. Seja

$$\mathbf{t} = \mathbf{x} \mathbf{w} \quad (29)$$

tal que t é a matriz de Pontuação. Seja $\mathbf{w}_k = (w_1, \dots, w_k)$ a matriz com os k primeiros vetores coluna da matriz \mathbf{w} . Sobre uma ótica geométrica, os autovetores da matriz \mathbf{w} formam uma base no espaço R^p e que combinados com os dados de \mathbf{x} , ou seja, realizando projeções de x em w produzem o que chamamos de Pontuações de Componentes Principais. Com estas componentes, poderemos realizar uma estimativa parcial $\hat{\mathbf{x}}$ para \mathbf{x} da seguinte forma:

$$\hat{\mathbf{x}} = \mathbf{t}_k \mathbf{w}^T \quad (30)$$

Com isto, \mathbf{x} pode ser expresso pela sua projeção como sendo:

$$\mathbf{x} = \hat{\mathbf{x}} + \epsilon \quad (31)$$

tal que ϵ é o erro da estimativa do modelo.

4.5 Regressão Por Mínimos Quadrados Parciais (PLS)

Introduzido por Wold em [Wold et al., 1983b, Wold et al., 1983a], o método de regressão por mínimos quadrados parciais foi inicialmente utilizado na área de quimioterapia para análises de cromatografias e espectrometrias. Além destas áreas, o PLS vem sendo aplicado cada vez mais em outras áreas como: marketing, processamento de imagens e finanças. Para este método, bons resultados podem ser obtidos quando existe uma alta colinearidade entre as características. Além disto, o número de características maior do

que o número de exemplos muitas vezes não é um entrave para o método. O PLS apresenta diversas características interessantes como: a redução da dimensionalidade, número baixo iterações, utilizado para estudo dos dados e utilizado como preditor.

Como o PCA, o PLS também reduz a dimensionalidade do espaço de \mathbf{x} . No entanto, o PLS extrai as Variáveis de Pontuação do modelo a partir da:

$$\mathbf{x}^T \mathbf{y} = Cov(\mathbf{x}, \mathbf{y}) \quad (32)$$

em que o termo à esquerda da igualdade é a matriz de covariância com os exemplos padronizados pela Z-Score. Para assegurar que este termo possui autovalores reais, positivos e seja quadrada utiliza-se como matriz de entrada:

$$\mathbf{x}^T \mathbf{y} \mathbf{y}^T \mathbf{x} \quad (33)$$

Para melhor esclarecimento do algoritmo, dividiremos este em duas etapas: Determinação da Estrutura Latente e Predição.

4.5.1 Determinação da Estrutura Latente

Nesta fase, o PLS aprende com os dados de treinamento. Podemos dividir esta fase em quatro subfases: Inicialização, Carga e Pontuação, Coeficientes de Regressão e Estimativa do Submodelo. Na fase de Inicialização, são efetuadas duas cópias dos dados: $\mathbf{x}_{k=0}$, cópia de \mathbf{x} e o mesmo para \mathbf{y} , em que k o índice da k -ésima iteração do algoritmo. Na fase de Carga e Pontuação, é necessário primeiramente, encontrar os autovetores (vetores de Carga) \mathbf{w} que expliquem a matriz, na equação (33). Para isto, recaímos em dois casos: seja $\mathbf{y}_{(p,q)}$ em que $q = 1$. De acordo com [Renteria, 2003], w pode ser obtido por:

$$\mathbf{w}_k = \mathbf{x}_{k-1}^T \mathbf{y}_{k-1} \quad (34)$$

e neste caso estamos tratando do algoritmo PLS1 [Renteria, 2003]. Já para o segundo caso, algoritmo PLS2 [Renteria, 2003], em que $q > 1$, ou seja, múltiplas variáveis dependentes, deve ser utilizado um método de extração de autovetores. Geralmente utiliza-se o método dos Mínimos Quadrados Parciais Não-Linear (NIPALS) [Golub and Van Loan, 1996] via método das potências.

Após a determinação dos vetores de Carga, determina-se o Vetor de Pontuação t_k . Assim temos:

$$\mathbf{t}_k = \mathbf{x}_{k-1} \mathbf{w}_k \quad (35)$$

Na fase seguinte, determinamos o que chamamos de Coeficientes de Regressão, da seguinte forma:

$$\begin{aligned} \mathbf{b}_k &= [(\mathbf{t}_k^T \mathbf{t}_k)^{-1} \mathbf{t}_k^T \mathbf{x}_{k-1}]^T = \frac{\mathbf{x}_{k-1}^T \mathbf{t}_k}{\mathbf{t}_k^T \mathbf{t}_k} \\ \mathbf{p}_k &= (\mathbf{t}_k^T \mathbf{t}_k)^{-1} \mathbf{t}_k^T \mathbf{y}_{k-1} = \frac{\mathbf{y}_{k-1}^T \mathbf{t}_k}{\mathbf{t}_k^T \mathbf{t}_k} \end{aligned} \quad (36)$$

e por fim teremos a fase de Estimativa do Submodelo, que procede da seguinte forma:

$$\begin{aligned} \mathbf{x}_k &= \mathbf{x}_{k-1} - \mathbf{t} \mathbf{b}_k^T \\ \mathbf{y}_k &= \mathbf{y}_{k-1} - \mathbf{t} \mathbf{p}_k^T \end{aligned} \quad (37)$$

Após a fase de estimativas, o processo se repete até o número de iterações desejadas. Note que, realizar k iterações é o mesmo que utilizar k variáveis latentes.

4.5.2 Predição

Após a determinação da estrutura latente, ou melhor, a fase de aprendizado, deve-se, para a fase de Predição, capturar os vetores de carga e os coeficientes de regressão aprendidos. Com estes dados, poderemos estimar cada y_i a partir dos submodelos aprendidos. Assim, a fase de predição pode ser descrita em três etapas: determinação da Vetor de Pontuação, na equação (35), com o Vetor de Carga aprendido, predição de y na forma:

$$\mathbf{y}_k = \mathbf{y}_{k-1} + \mathbf{tp}^T \quad (38)$$

e por fim o cálculo do resíduo, como na equação (37), mas somente para x_k . Este processo se repete para o mesmo número k de fatores aprendidos na fase de regressão.

5 Preditores Financeiros

Preditores de séries temporais financeiras são de fundamental importância para Intraday Traders. O conhecimento acurado do valor do preço de um determinado ativo em instantes futuros ou até mesmo sua variação, pode ser de grande valia para áreas como a de tomada de decisões e a de gerenciamento de risco.

5.1 Preditor Classificador de Comportamento de Séries Financeiras

A classificação do comportamento de séries financeiras é bastante utilizada para Intraday Traders. Por exemplo, na área de tomada de decisões, o conhecimento de uma possível alta do preço do ativo em um instante futuro confirma uma decisão de compra da ação para o instante de negociação. De forma análoga, tem-se para a venda, em que uma possível queda do preço do ativo ratifica a decisão de compra. Para este trabalho, utilizamos um classificador SVM para detectar altas e quedas de preço médio do dia seguinte.

No campo de gerenciamento de risco, classificadores podem atuar como seletores de ativos de baixo risco que compõem uma carteira. Além disto, podem atuar como detectores de situações de risco, indicando momentos oportunos de entrada e saída no mercado acionário.

5.2 Preditor Regressivo de Séries Financeiras

Assim como a classificação do comportamento de séries financeiras, a predição de valores futuros pode ser bastante útil para uma boa estratégia de tomada de decisão. Por exemplo, veja o caso da predição de preço médio futuro: se em determinado instante deste mesmo dia o preço do ativo encontra-se abaixo de seu preço médio, uma operação de compra neste instante pode ser considerada uma barganha se a predição do preço médio for acurada. Isto porque estamos comprando por um valor abaixo do preço de mercado do ativo. Analogamente, uma venda acima do preço médio do dia é uma boa operação, considerando-se que se no início do dia, compra-se abaixo do preço médio, existe grande chance de se vender por um preço acima do preço médio dentro deste mesmo dia. Para

este trabalho, desenvolvemos três preditores regressivos: SVR, PLS e um comitê de SVR e PLS. Todos os três preditores são para preço médio do dia.

Além da área de tomada de decisão, devemos mencionar a área de negociação algorítmica. Nesta, são desenvolvidas estratégias de execução de ordens para as corretoras. Estas estratégias objetivam, além de maximizar o lucro com a taxa de corretagem por execução de ordem, minimizar o impacto de mercado causado por estas execuções. O impacto de mercado é causado quando um volume financeiro negociado está acima do volume médio de mercado para um dado ativo. As consequências são variações do preço do ativo tanto para alta quanto para queda. Predições de volume financeiro, quantidade e preço médio do ativo auxiliam estas estratégias de execução.

6 Intraday Traders

Nesta seção introduziremos alguns conceitos básicos de Intraday Traders tanto para execução de uma ordem quanto para uma estratégia de negociação.

6.1 Algoritmo de Envio de Ordens do Trader

Antes de entrarmos em detalhes das estratégias de negociação devemos conhecer o funcionamento o algoritmo de execução de ordens do Trader. Para isso, consideremos o seguinte pseudo-código:

Algorithm 1 : sendOrders

```
while marketIsOpen do
  state ← updateMarket, updateWallet
  orders ← computeStrategy(state)
  orders ← valid(orders)
  if orders ≠ void then
    (matchedOrders, unmatchedOrders) ← sendOrders(orders)
    storeOrders(matchedOrders, unmatchedOrders)
  end if
end while
```

No algoritmo 1, marketIsOpen é a condição necessária que faz com que o Trader opere na horas em que o mercado está aberto. State representa todas as informações de mercado necessárias para que o Trader efetue sua estratégia. As funções updateMarket e updateWallet atualizam, respectivamente, as informações do mercado e da carteira do Trader. A função computeStrategy é responsável pela estratégia de negociação do Trader. A função valid é responsável por verificar se as ordens são válidas e estão de acordo com os dados da carteira. Por fim, temos a função sendOrders que envia as ordens a uma corretora ou a um simulador. A função sendOrders também retorna as ordens executadas. StoreOrders é um log de execução das ordens do Trader. É importante considerar que, como estamos trabalhando com dados de histórico, não será considerado o casamento das ordens de entrada e saída de mercado.

6.2 Revisão de Estratégias para Intraday Traders

Baseado em estudos de [Silaghi et al., 2005, Sherstov and Stone, 2005] faremos uma revisão de estratégias para Intraday Traders. As estratégias que serão vistas podem ser divididas em três classes. A primeira classe é a das estratégias baseadas no comportamento do preço das ações, que levam em consideração os dados do passado e presente para inferir algo com relação à decisão futura a ser tomada. A segunda classe de estratégias é a baseada na informação do livro de ordens. Este livro de ordens contém informações como a quantidade de compradores e vendedores para uma dada ação. Por fim, a terceira classe de estratégias, que contém tanto a informação do livro de ordens quanto a informação do comportamento do preço da ação. Veremos a seguir algumas estratégias que se enquadram em uma das duas primeiras classes descritas aqui.

6.2.1 Acompanhamento de Tendência de Preços

Esta estratégia, da classe 1, é baseada exclusivamente na variação do preço da ação ao longo do tempo funcionando do seguinte modo: primeiramente, calculam-se duas retas de regressão com janelas de tempo distintas. A primeira reta representa uma tendência a longo prazo. Já a segunda reta representa uma tendência a curto prazo. A partir da informação do coeficiente angular das duas retas de tendência, verifica-se o sinal destes coeficientes. Caso ambos os coeficientes sejam positivos, então é efetuada uma ordem de compra. Caso ambos sejam negativos, é feita uma venda. Caso contrário o Trader mantém-se inoperante. A seguir temos o algoritmo da estratégia.

Algorithm 2 : computeStrategy

```
slope1 ← makeRegression(prices,longWindow)
slope2 ← makeRegression(prices,shortWindow)
if slope1 > 0 and slope2 > 0 then
    "placeBuyOrder"
else
    if slope1 < 0 and slope2 < 0 then
        "placeSellOrder"
    end if
end if
```

6.2.2 Acompanhamento Reverso de Tendência de Preços

Esta estratégia é semelhante à estratégia de Acompanhamento de Tendência de Preços. No entanto, a tomada de decisão é reversa. Isto significa que, o Trader entra no mercado quando há uma suspeita de queda de preços. De forma análoga, o Trader sai do mercado quando há uma suspeita de alta de preços. À princípio, a tomada de decisão deste tipo de estratégia parece pouco intuitiva. No entanto, segundo [Silaghi et al., 2005], este tipo de estratégia explora as microvariações do preço do ativo. Os resultados obtidos em [Silaghi et al., 2005] demonstram que este tipo de estratégia muitas vezes acaba por superar a estratégia de Acompanhamento de Tendência de Preços.

6.2.3 Preço Médio Ponderado pela Quantidade (VWAP)

O VWAP é uma métrica muito utilizada pelas corretoras que desejam executar suas ordens de forma a causar o mínimo impacto de mercado possível. O VWAP corresponde ao volume financeiro transacionado em um determinado período dividido pelo número de ações daquele mesmo período. Além das corretoras, o VWAP também pode ser útil para Intraday Traders. Uma boa estimativa do preço médio de um ativo ou até mesmo sua variação ao longo do tempo ajuda o Trader a comprar abaixo e vender acima do preço médio de mercado. O VWAP pode ser expresso por:

$$VWAP = \frac{\sum_{i=1}^n Q_i P_i}{\sum_{i=1}^n Q_i} \quad (39)$$

tal que Q_i é a quantidade de ações negociadas no instante i e P_i é o preço da ação no instante i .

Nesta estratégia, da classe 2, o VWAP é utilizado como o preço de equilíbrio do mercado. O funcionamento desta estratégia se dá da seguinte forma: primeiramente, calcula-se o VWAP do dia para todo o mercado. Se o preço das primeiras ordens de entrada de mercado estiverem acima do VWAP, então o Trader sai do mercado. Se o preço das primeiras ordens de saída de mercado estiverem abaixo do VWAP, então é feita uma entrada de mercado. Caso contrário, o Trader mantém-se inoperante.

Algorithm 3 : computeStrategy

```
vwap ← computeVWAP(orders)
if buyPrice > vwap then
  "placeBuyOrder"
else
  if sellPrice < vwap then
    "placeSellOrder"
  end if
end if
```

6.2.4 Livro de Ordens Estáticas Desbalanceadas (SOBI)

SOBI é da classe 2 de estratégias e apresenta resultados muito bons. Em [Silaghi et al., 2005, Sherstov and Stone, 2005] podem ser encontrados resultados interessantes deste tipo de estratégia que, muitas vezes, acaba por superar as demais estratégias descritas. Nesta estratégia calcula-se, separadamente, dois VWAPs. Um para o livro de compras e outro para o livro de vendas. Com a informação do último preço transacionado, calculam-se duas diferenças com relação aos VWAPs de compra e venda. Com estas diferenças, compara-se as duas e fazem-se as seguintes verificações: caso a diferença do VWAP de venda seja maior do que a de compra, o que quer dizer que o preço subirá, então faz-se uma compra; caso a diferença do VWAP de compra seja maior do que a de venda, o preço descera, então faz-se uma venda; caso contrário, mantenha-se inoperante.

Algorithm 4 : computeStrategy

```
buyVwap ← computeBuyVWAP(buyOrders)
sellVwap ← computeSellVWAP(sellOrders)
if buyVwap - lastTradePrice > sellVwap - lastTradePrice then
  "placeBuyOrder"
else
  if buyVwap - lastTradePrice < sellVwap - lastTradePrice then
    "placeSellOrder"
  end if
end if
```

6.3 Estratégia Proposta: Estratégia de Reversão de VWAP

Para esta estratégia utilizamos dois do que podemos chamar de detectores. Detectores são responsáveis por alertar ao Trader quando é o momento de entrar e sair do mercado. Dentre os dois detectores, o primeiro, detector interday, é responsável por alertar qual dia deve-se entrar ou sair do mercado. O segundo, é um detector intraday, que é responsável por avisar qual é o instante do dia em que o Trader deve entrar ou sair do mercado.

O detector interday, é um preditor classificador binário de variações de VWAP entre dias. Este detector classificará uma alta ou queda entre os VWAPs do dia atual e seguinte. Com este tipo de classificação, podemos implementar uma estratégia que opere somente em reversões de classificação. Por exemplo, se do dia anterior para o atual houve uma queda do VWAP e se o detector classifica como uma alta de VWAP para o dia posterior, então temos uma reversão. Esta reversão indica um dia de entrada no mercado. De forma inversa temos uma reversão para uma saída do mercado.

O detector intraday, é um preditor regressivo de VWAP do dia atual. Com esta informação, podemos escolher um instante de entrada ou saída de mercado apropriados. Um ponto de entrada apropriado seria quando o preço do ativo estivesse abaixo do VWAP estimado do dia. De forma análoga, um ponto de saída do mercado seria quando o preço do ativo estivesse acima do VWAP predito do dia.

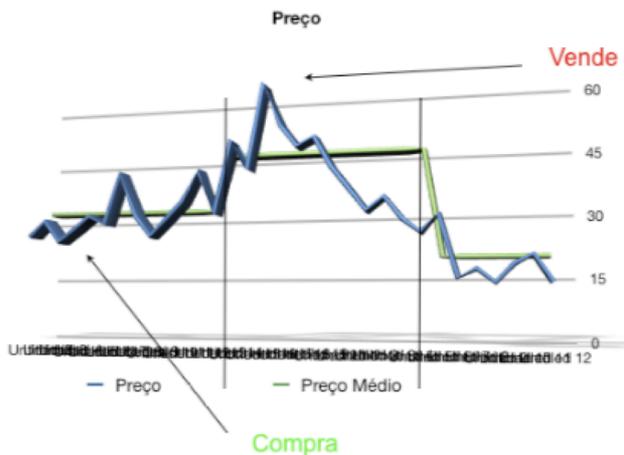


Figure 4: Estratégia 2.

Na figura 4, a linha azul representa o preço do ativo, a linha verde o VWAP estimado

do dia e o intervalo de barras verticais representa um dia de negociação. Para o primeiro dia, o detector interday avisa que é o dia atual é um dia apropriado para entrada no mercado. Dentro deste mesmo dia, o detector intraday escolhe um instante dentro do dia para fazer uma compra por um preço abaixo do VWAP do dia. Da mesma forma ocorre para o segundo dia. No entanto, os detectores informam o momento de saída do mercado. Como podemos notar, a compra é feita abaixo da linha verde e a venda acima da linha verde.

Algorithm 5 : computeStrategy

```
vwapSeries ← makeSeries(orders)
currChange ← predictVWAPChange(vwapSeries)
vwap ← predictVWAP(vwapSeries)
if (currChange = 1 and lastChange = -1) and (currPrice < vwap) then
    "placeBuyOrder"
else
    if (currChange = -1 and lastChange = 1) and (currPrice > vwap) then
        "placeSellOrder"
    end if
end if
lastChange ← currChange
```

7 O Framework LearnTrade

Para implementação e experimentos dos Intraday Traders e dos preditores de Aprendizado de Máquina foi desenvolvido, em paralelo, o framework LearnTrade. O objetivo central deste framework é dar suporte às áreas de predição, tomada de decisões e gerenciamento de riscos. Desenvolvido com técnicas de padrões de projetos e com enfoque no fácil manuseio, LearnTrade permite desenvolvimento ágil de experimentos. Em uma combinação da linguagem Python, a biblioteca numérica Numpy e a linguagem C, o framework atinge, além do desenvolvimento ágil, bom desempenho. Atualmente, o LearnTrade está dividido em duas camadas: uma para Preditores de Séries Financeiras e outra para Intraday Traders.

Na camada de preditores, o LearnTrade provê algoritmos de regressão e classificação do estado da arte de Aprendizado de Máquina para Séries Temporais Financeiras. Podem ser encontrados, também, uma série de recursos para realização de experimentos como: métodos de validação, construtores de modelos, ferramentas de pré-processamento e pós-processamento dos dados, um pacote para uma análise estatística dos dados, medidores de performance, mapeadores de resultados e gráficos de qualidade para visualização de séries temporais e erros de validação.

Na camada de Intraday Traders, o LearnTrade fornece um simulador com carteira, log de transações e alguns exemplos de estratégias de negociação. O simulador trabalha com dados tanto em tempo real quanto em modo simulado. Para os dados em tempo real, estes são obtidos do repositório da Yahoo Finanças. Já para dados em modo simulado, o framework provê um histórico de dados de ativos de alta liquidez e média liquidez da Bovespa.

8 Experimentos

Para este trabalho, foi selecionado um conjunto de dados de transações intraday de dez ativos da bovespa. Os ativos são: PETR3, PETR4, VALE5, BBDC4, GOAU5, ITAU4, ITSA4, USIM5, UBBR11 e CSNA3. Os dados encontram-se num intervalo de dois anos com início em 2006 e fim no início de 2008, totalizando 460 dias. Dentre estes 460 dias, foram utilizados 360 dias para treino e 100 dias para teste.

A partir desta seção, utilizaremos algumas definições e métricas para avaliação de ativos e Intraday Traders que são comumente encontradas na literatura. Seja uma sequência de preços $S = s, i = 1, n$ em que s_i é o preço da sequência no tempo discreto i , podemos expressar o retorno como sendo:

$$r_i = \frac{s_i - s_{i-1}}{s_i} \quad (40)$$

tal que r_i é o retorno da sequência no tempo i , s_i e s_{i-1} correspondem, respectivamente, aos valores da sequência nos instantes i e $i - 1$. A segunda é o retorno médio que é:

$$r_m = \frac{1}{n} \sum_{i=1}^n \frac{s_i - s_{i-1}}{s_i} \quad (41)$$

tal que r_m corresponde a média de todos os retornos da sequência. A terceira métrica é o Índice de Sharpe, que pode ser expresso por:

$$i_s = \frac{r_m - fr_m}{\sigma} \quad (42)$$

tal que r_m corresponde a média de todos os retornos da sequência, fr_m o retorno médio de um capital livre de risco no mesmo período e σ é o desvio padrão da sequência de retorno S . O Índice de Sharpe, proposto por William F. Sharpe, tem como objetivo avaliar o ganho sobre a volatilidade do ativo. Esta métrica foi aplicada, inicialmente, para avaliar o risco de ativos. No entanto, este tipo de métrica pode ser utilizado, também, para avaliar o desempenho de Intraday Traders como em [Silaghi et al., 2005, Sherstov and Stone, 2005]. O Índice de Sharpe está diretamente relacionado a confiabilidade da estratégia de negociação do Trader. Quanto maior for este índice, maior a confiabilidade do Trader para o ativo.

8.1 Análise dos Dados

Baseado em [Varga, 2001] podemos avaliar os ativos pelo retorno médio e pelo Índice de Sharpe. Com relação a essas duas métricas podemos separar os ativos em três grupos distintos. O primeiro grupo, que contém os ativos PETR3, PETR4 e CSNA5, é o grupo dos ativos de alto retorno e baixo risco. O segundo grupo, que contém VALE5, USIM5, BBDC4 e UBBR11, é o grupo dos ativos de médio retorno e risco moderado. Já para o último grupo, temos os ativos GOAU4, ITAU4 e ITSA4, que contém os ativos de retorno baixo ou negativo e com um risco alto de investimento. A primeira coluna da tabela 1, corresponde apenas ao retorno obtido pela compra do ativo no primeiro dia e venda no último dia de operação.

Nas figuras 5,6 e 7 nos mostram o retorno acumulado dos dez ativos em um período de 100 dias. Podemos notar que muitos dos ativos estão fortemente correlacionados. Em

Table 1: Métricas por Ativo

Ativo	Retorno	Retorno Médio	Índice de Sharpe
PETR3	67.81 %	0.46 %	2.61
PETR4	63.96 %	0.43 %	2.49
CSNA3	52.08 %	0.35 %	2.42
VALE5	14.32 %	0.06 %	0.36
UBBR11	11.27 %	0.07 %	0.45
USIM5	11.90 %	0.03 %	0.21
BBDC4	3.23 %	0.01 %	0.06
GOAU4	7.04 %	-0.01 %	-0.07
ITAU4	-0.39 %	-0.03 %	-0.23
ITSA4	-9.18 %	-0.14 %	-1.04

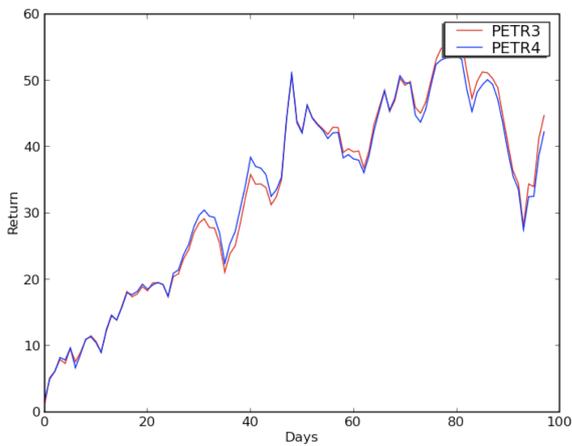


Figure 5: Retorno Acumulado dos Ativos

termos dessa correlação entre ativos, podemos dividir os ativos em dois grupos: PETR3 e PETR4 no grupo 1 e os demais ativos no grupo 2. O importante deste tipo de análise é notar que, construir modelos especializados para cada grupo, pode levar a construção de modelos multiativos.

8.2 Preditor Classificador da Variação do VWAP

Como métrica de qualidade para o problema de classificação, foram utilizados dois tipos de métricas: Acurácia

$$AC = \frac{(tp + tn)}{(tp + tn + fp + fn)} \quad (43)$$

tal que tp são os verdadeiros positivos, tn são os verdadeiros negativos, fp são os falsos positivos e fn são os falsos negativos. A acurácia mede o número de acertos com relação ao total de respostas possíveis. Uma outra métrica, muito utilizada para problemas em que se tem classificação binária na área de Aprendizado de Máquina é o MCC (Coefi-

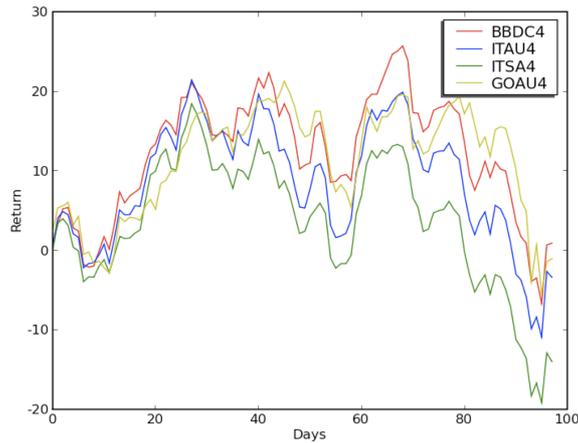


Figure 6: Retorno Acumulado dos Ativos

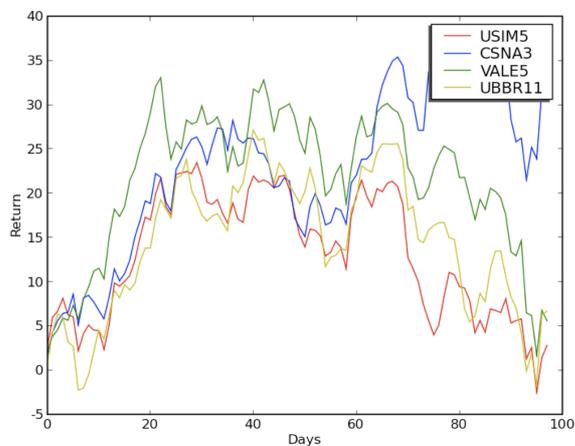


Figure 7: Retorno Acumulado dos Ativos

ciente de Correlação de Mathews) que pode ser expresso por:

$$MCC = \frac{tp * tn - fp * fn}{\sqrt{(tp + fp) + (tp + fn) + (tn + fp) + (tn + fn)}} \quad (44)$$

Um dos pontos fortes do MCC é que, mesmo com classes de tamanhos bem distintos pode-se ter uma estimativa boa da qualidade da classificação. O MCC retorna valores entre -1 e 1, em que 1 representa uma predição perfeita, 0 um passeio aleatório e -1 uma predição inversa. Em nosso caso, foi multiplicado o resultado do MCC por 100 de forma a ficar mais claro para análise.

Como podemos observar o único resultado desfavorável foi para o ativo USIM5, os demais ativos apresentam um MCC maior do que zero o que representa um classificador melhor do que um passeio aleatório. Da mesma forma, podemos dizer sobre a acurácia, que teve seu único resultado abaixo de 50% para a USIM5. Sobre estes resultados, tanto do MCC quanto da acurácia, ainda há muito a melhorar com outras modelagens, técnicas de pre-processamento e classificadores auxiliares. No entanto, como veremos adiante, podemos chegar a resultados bastantes expressivos com este preditor.

Table 2: Preditor Classificador da Variação do VWAP

Ativo	Acurácia	MCC
ITAU4	59.60%	23%
ITSA4	59.60%	18.30%
BBDC4	55.56%	10.02%
UBBR11	57.58%	18.73%
CSNA3	56.57%	9.63%
PETR3	57.58%	8.42%
GOAU4	57.58%	6.62%
VALE5	52.53%	4.94%
PETR4	53.54%	3.97%
USIM5	43.43%	-10.41%

8.3 Preditor Regressivo do VWAP

Para avaliar a qualidade da predição do preditor de preço médio do dia, foi utilizada a métrica de erro relativo, dada por:

$$\epsilon_m = \sum_{i=1}^n \left\| \frac{\hat{y}_i - y_i}{y_i} \right\| \quad (45)$$

tal que y_i é a resposta certa, \hat{y}_i é a resposta estimada e ϵ_m é o erro relativo médio para todas as respostas estimadas. Em contrapartida a métricas de erro quadrático, como a média do erro quadrático (MSE), que são estimativas de erro global extremamente sensíveis à ordem de grandeza dos valores de saída, o erro relativo não sofre este tipo de problema, pois é uma métrica percentual de erro.

Table 3: Preditor PLS do VWAP

Ativo	início do dia	meio do dia	média global
PETR3	2.18%	0.67%	0.82%
USIM5	1.73%	0.46%	0.55%
CSNA3	1.80%	0.53%	0.75%
VALE5	2.04%	0.36%	0.56%
UBBR11	1.83%	0.50%	0.70%
BBDC4	1.79%	0.44%	0.53%
ITAU4	1.80%	0.39%	0.56%
ITSA4	1.65%	0.53%	0.55%
PETR4	2.18%	0.40%	0.54%
GOAU4	1.80%	0.80%	0.93%

Como podemos observar, os resultados de erro relativo apresentam-se baixos. Isto porque a variância do VWAP ao longo do dia é muito baixa, o que requer um preditor com erro relativo muito baixo. Dentro de um dia, a variação do VWAP é em centavos de real. Podemos notar, também que, ao passar do dia, o erro dos preditores cai bastante e no meio do dia a predição já abaixo de 1% de erro relativo, o que já nos dá um preditor regressivo de qualidade.

Table 4: Preditor SVR do VWAP

Ativo	início do dia	meio do dia	média global
PETR3	2.16%	0.54%	0.80%
USIM5	1.73%	0.46%	0.58%
CSNA3	1.83%	0.52%	0.78%
VALE5	2.04%	0.36%	0.58%
UBBR11	1.80%	0.52%	0.73%
BBDC4	1.83%	0.46%	0.59%
ITAU4	1.80%	0.40%	0.60%
ITSA4	1.66%	0.50%	0.55%
PETR4	2.15%	0.57%	0.68%
GOAU4	1.81%	0.63%	0.78%

Analisando os preditores regressivos PLS e SVR, podemos chegar a conclusão de que uma combinação dos dois preditores pode vir a melhorar o resultado global do erro relativo.

Table 5: Preditor SVR+PLS do VWAP

Ativo	início do dia	meio do dia	média global
PETR3	2.16%	0.53%	0.77%
USIM5	1.73%	0.44%	0.54%
CSNA3	1.80%	0.51%	0.73%
VALE5	2.04%	0.35%	0.54%
UBBR11	1.81%	0.50%	0.69%
BBDC4	1.81%	0.43%	0.54%
ITAU4	1.79%	0.38%	0.56%
ITSA4	1.65%	0.48%	0.52%
PETR4	2.16%	0.45%	0.59%
GOAU4	1.81%	0.69%	0.82%

Como podemos observar na tabela 6, obtivemos uma melhoria dos resultados com a utilização do comitê de SVR e PLS. Os únicos resultados que não obtiveram melhora foram para GOAU4 e BBDC4.

8.4 Estratégia de Reversão de VWAP

Os experimentos da tabela 7, descrevem a atuação do Trader em um período de 100 dias com um capital inicial de R\$ 10.000. Como podemos observar, na coluna Lucro das tabelas, os experimentos acabam por refletir a realidade sobre os melhores ativos a investir. De acordo com o esperado, os ativos com maior retorno, na tabela 1, acabam por gerar melhores resultados ao Trader em comparação aos ativos de baixo retorno. Da mesma forma podemos analisar sobre o ponto de vista do risco. Os ativos com menor risco, na tabela 1, proporcionaram maior retorno para o Trader, dado o menor risco que os ativos possuíam. Mesmo que o leitor não recorra a utilização de Intraday Traders, este estudo pode orientar na seleção de ativos com carteira mais rentável.

Table 6: Erro Relativo Global dos Preditores Regressivos

Ativo	SVR	PLS	PLS+SVR
PETR3	0.80%	0.82%	0.77%
USIM5	0.58%	0.55%	0.54%
CSNA3	0.78%	0.75%	0.73%
VALE5	0.58%	0.56%	0.54%
UBBR11	0.73%	0.70%	0.69%
BBDC4	0.59%	0.53%	0.54%
ITAU4	0.60%	0.56%	0.56%
ITSA4	0.55%	0.55%	0.52%
PETR4	0.68%	0.54%	0.59%
GOAU4	0.78%	0.93%	0.82%

Table 7: Estratégia de Reversão de VWAP

Ativo	Lucro	Índ. Sharpe	Par Compra/Venda	Operações
PETR4	75.00%	2.45	4.68%	31%
PETR3	66.75%	2.70	6.67%	19%
CSNA3	63.36%	4.03	4.52%	27%
USIM5	27.52%	1.83	2.75%	19%
VALE5	15.53%	0.75	1.29%	23%
GOAU4	5.17%	0.09	0.25%	39%
ITSA4	3.73%	0.02	0.20%	35%
ITAU4	1.87%	-0.08	0.12%	30%
UBBR11	-3.56%	-0.87	-0.16%	42%
BBDC4	1.62%	-0.11	0.07%	43%

9 Conclusões

Neste trabalho foi proposta uma nova estratégia de VWAP para Intraday Traders do mercado acionário. Esta estratégia utilizou dois detectores automáticos de VWAP. Um detector classificador SVM para a escolha do dia de negociação e um segundo detector com objetivo de escolher um instante dentro dia que fosse apropriado para negociação. Os resultados obtidos são significativos e com ganhos de até 75% para ativos de retorno médio positivo e de até 3.7% para ativos de retorno médio negativo.

10 Trabalhos Futuros

Como trabalhos futuros, pretende-se aplicar novos tipos de preditores que auxiliem ao Trader à tomada de decisões mais seguras. Uma opção seria incorporar o Índice de Sharpe na modelagem do preditor. Com isto, teríamos detectores do retorno dado o risco de entrar e sair do mercado. Modelagens deste tipo podem ser encontradas em [Boyarshinov, 2005, Choey and Weigend, 1997].

Investir em melhores preditores regressivos de VWAP do dia pode melhorar a qualidade do comitê de preditores. Além disso, a adição de um número maior de preditores pode contribuir, também, para a melhora do comitês.

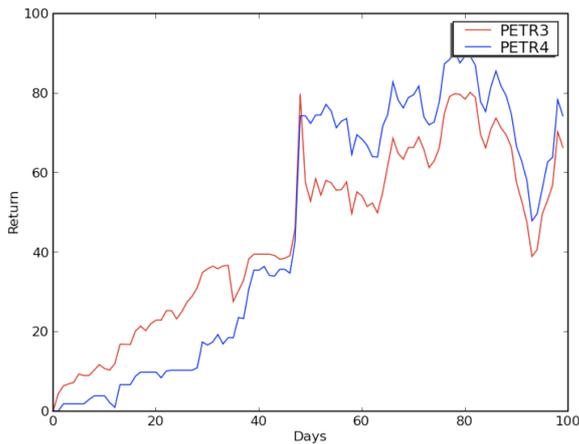


Figure 8: Retorno Acumulado do Trader.

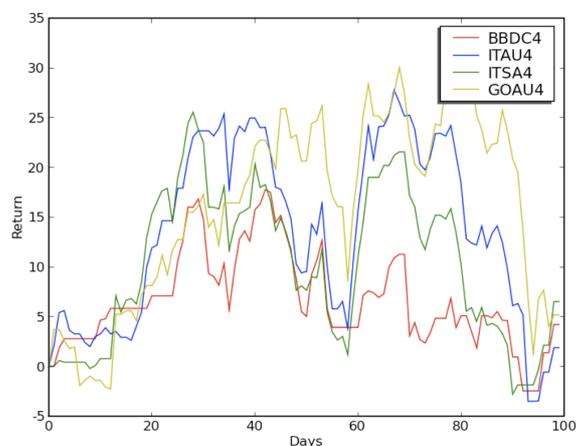


Figure 9: Retorno Acumulado do Trader.

Outro trabalho que pode ser explorado, é o da criação detectores de médio e longo prazo. Estes detectores indicariam tendências futuras do comportamento do ativo. Por exemplo, o Trader poderia investir em uma estratégia mais arriscada para tendências, a longo prazo, de alta do preço do ativo. Da mesma forma, o Trader pode utilizar uma estratégia conservadora quando detectar uma tendência de queda a longo prazo.

References

- [Ben-Hur and Weston,] Ben-Hur, A. and Weston, J. *A User's Guide to Support Vector Machines*.
- [Boser et al., 1992] Boser, B. E., Vapnik, V. N., and Guyon, I. M. (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*, pages 144–152. ACM Press.
- [Boyarshinov, 2005] Boyarshinov, V. (2005). Machine learning in computational finance. *cs.rpi.edu*.

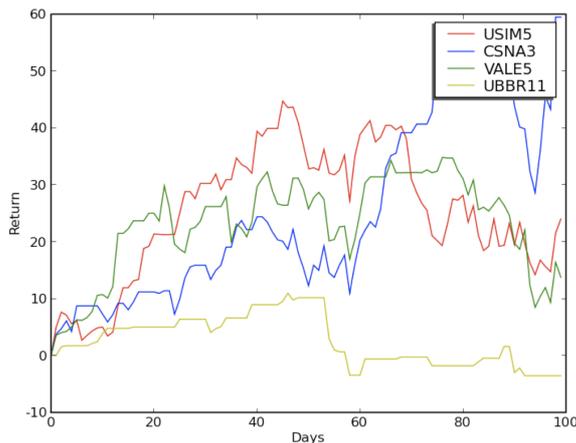


Figure 10: Retorno Acumulado do Trader.

- [Briza and Jr, 2008] Briza, A. and Jr, P. N. (2008). ... for historical market data using multiobjective particle swarm optimization of technical indicators. *Proceedings of the 2008 GECCO conference companion on ...*
- [Choey and Weigend, 1997] Choey, M. and Weigend, A. (1997). Nonlinear trading models through sharpe ratio maximization. *International Journal Of Neural Systems*.
- [Cortes and Vapnik, 1995] Cortes, C. and Vapnik, V. (1995). Support vector networks. In *Machine Learning*, pages 273–297.
- [Golub and Van Loan, 1996] Golub, G. H. and Van Loan, C. F. (1996). *Matrix computations (3rd ed.)*. Johns Hopkins University Press, Baltimore, MD, USA.
- [Gunn, 1998] Gunn, S. (1998). Support vector machines for classification and regression. *ISIS Technical Report*.
- [Leung et al., 2000] Leung, M. T., Daouk, H., and Chen, A. (April 2000). Forecasting stock indices: A comparison of classification and level estimation models. *International Journal of Forecasting*, 16:173–190(18).
- [Mueller et al., 1997] Mueller, K., Smola, A., Raetsch, G., and Schoelkopf, B. (1997). Predicting time series with support vector machines. *Lecture Notes In Computer Science*.
- [Renteria, 2003] Renteria, R. (2003). *Algoritmos para Regressão por Mínimos Quadrados Parciais*. PhD thesis, Pontifícia Universidade Católica do Rio de Janeiro.
- [Sherstov and Stone, 2005] Sherstov, A. and Stone, P. (2005). Three automated stock-trading agents: A comparative study. *Lecture Notes In Computer Science*.
- [Shlens, 2005] Shlens, J. (2005). A tutorial on principal component analysis. <http://www.sn1.salk.edu/shlens/pub/notes/pca.pdf>.
- [Silaghi et al., 2005] Silaghi, G., Cluj-Napoca, R., and Robu, V. (2005). An agent strategy for automated stock market trading combining price and order book information. *Computational Intelligence Methods and Applications*.

- [Smola and Schölkopf, 2004] Smola, A. and Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*.
- [Varga, 2001] Varga, G. (2001). Índice de sharpe e outros indicadores de performance aplicados a fundos de ações brasileiros. *Revista de Administração Contemporânea*.
- [Wold et al., 1983a] Wold, S., Albano, C., Esbensen, K., Hellberg, S., Johansson, E., and Sjostrom, H. (1983a). Pattern recognition: Finding and using regularities in multivariate data. In *Proceedings Iufost Conf Food Research and Data Analysis*. Applied Science Publications.
- [Wold et al., 1983b] Wold, S., Martens, H., and Wold, H. (1983b). The multivariate calibration problem in chemistry solved by the pls method. In *Proceedings Conf. Matrix Pencils*, pages 286–293. Springer Verlag.