

PREVISÃO DE MOVIMENTOS FUTUROS DE AÇÕES DA BOLSA DE VALORES



PUCPR
GRUPO MARISTA

Norberto Ritzmann Júnior

Conteúdo

- ❑ Introdução
- ❑ Previsão em Ações
- ❑ SVM
- ❑ Método
- ❑ SVM: Experimentos Realizados
- ❑ Redes Neurais LSTM
- ❑ LSTM: Experimentos Realizados

Introdução

- **Previsão**, em estatística é o processo de estimativas em situações de incertezas. Normalmente se refere a estimação de séries temporais ou dados instantâneos.
- Exemplos:
 - ▣ Previsão de demanda de consumo;
 - ▣ Previsão do clima;
 - ▣ Previsão de produção de fábricas...

Introdução

- ❑ **Mas e ações, é possível prever 100% o preço futuro das ações?** Não! O que é possível é acertar mais do que se fosse aleatoriamente.
- ❑ Taxa de acerto:
 - ❑ Comparado com outras áreas é considerado baixo, por isso este tema é considerado um dos mais desafiadores. Métodos de classificação em previsão de ações acertam bem menos, em torno de 60-65%

Introdução

- Por que há espaço para previsão na bolsa?
 - 70% das transações nos EUA são feitas por software;
 - Morgan Stanley, City Group contratam dezenas de Cientistas de dados por ano;

Introdução

- Por que há espaço para previsão na bolsa?

Valor de Mercado das Ações do iBovespa
R\$ 2.744.866.348.640,00

Valor movimentado diariamente:
R\$ 6.500.000,00

Ou seja:
Apenas 0,23% do valor de mercado é
negociado diariamente

Introdução

□ Por que há espaço para previsão na bolsa?

60,3% das ações são negociadas por estrangeiros, que em sua grande maioria são grandes investidores, fundos de pensão, grandes empresas.

25,7% são fundos, empresas e bancos brasileiros.

Grande Fundo de Investimento



Quanto tempo um grande navio demoraria para mudar 180º na sua direção?

Introdução

□ Por que há espaço para previsão na bolsa?

Pequeno investidor...

Aproximadamente 15% das ações são negociadas por pessoas físicas, e que por terem pequenas quantias das ações podem negociar em instantes todo o seu estoque de ações.



Introdução

- Por que há espaço para previsão na bolsa?

Resultado dos grandes movimentos gerados por grandes fundos: **Tendência.**



Fundamentação Teórica

- Característica dos Dados
 - Dados estruturados
 - Série temporal não estacionária
 - Os valores são afetados por:
 - Ciclos;
 - Tendências;
 - E valores aleatórios.

Fundamentação Teórica

□ Tendências e Ciclos



Fundamentação Teórica

□ Tendências e Ciclos



Fundamentação Teórica

□ Composição de Valor

- Fundamentos da Empresa;
- Eventos políticos;
- Situação econômica e mercadológica;
- Taxas e juros bancários;
- Decisões de investidores institucionais;
- Movimentos de outros mercados de ações;
- E a psicologia dos investidores.

Fundamentação Teórica

□ Dados Iniciais (Não transformados)

Date	Open	High	Low	Close	Volume
2003-06-02	6,92	6,99	6,77	6,78	4874400
2003-06-03	6,79	6,83	6,77	6,79	2041200
2003-06-04	6,75	6,87	6,68	6,8	3501600
2003-06-05	6,76	6,83	6,76	6,78	2508000
2003-06-06	6,8	6,9	6,8	6,9	2706000
2003-06-09	6,9	6,9	6,8	6,8	1426800
2003-06-10	6,82	6,82	6,74	6,75	3268800
2003-06-11	6,76	6,85	6,71	6,73	3192000
2003-06-12	6,74	6,74	6,65	6,66	4872000
2003-06-13	6,7	6,7	6,6	6,6	3304800
2003-06-16	6,6	6,66	6,56	6,62	3888000
2003-06-17	6,67	6,74	6,64	6,67	4522800
2003-06-18	6,73	6,83	6,67	6,75	4756800
2003-06-20	6,8	6,8	6,68	6,71	5431200

Aplicação de mineração de Dados no contexto do trabalho consiste em:



Limpeza

Ordenação

Transformação

Seleção de Atributos

Classificação

Fundamentação Teórica

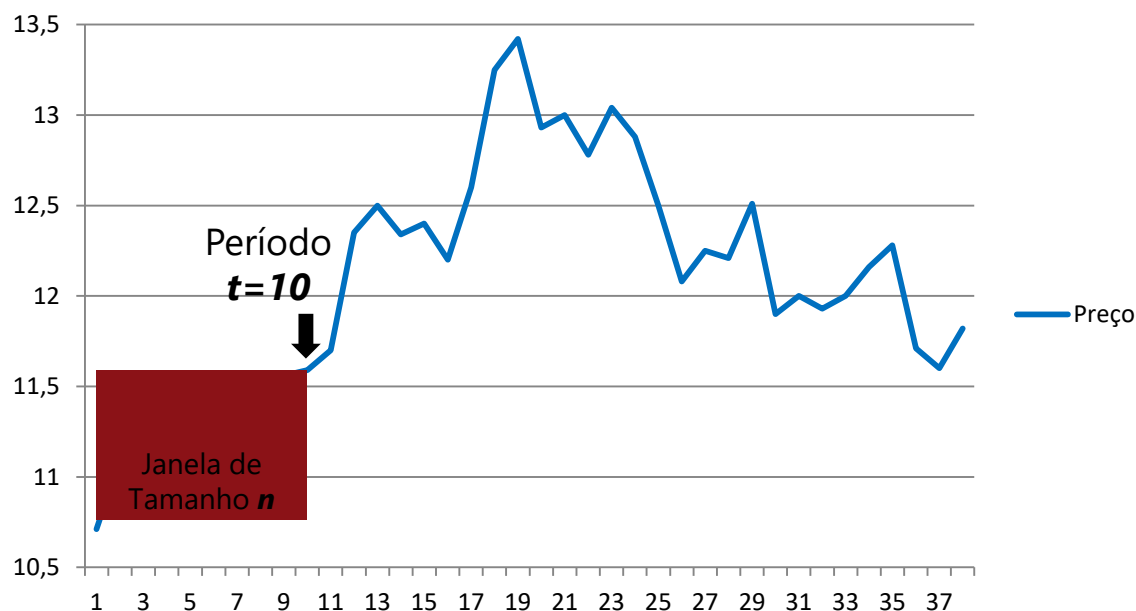
□ Transformação

- Fase de extração de características e criação de novos atributos. Esses atributos podem refletir características como:
 - Tendência
 - Força compradora / vendedora;
 - Esgotamento de tendência.

Fundamentação Teórica

□ Transformação – Extração de Indicadores Técnicos

□ Médias Móveis



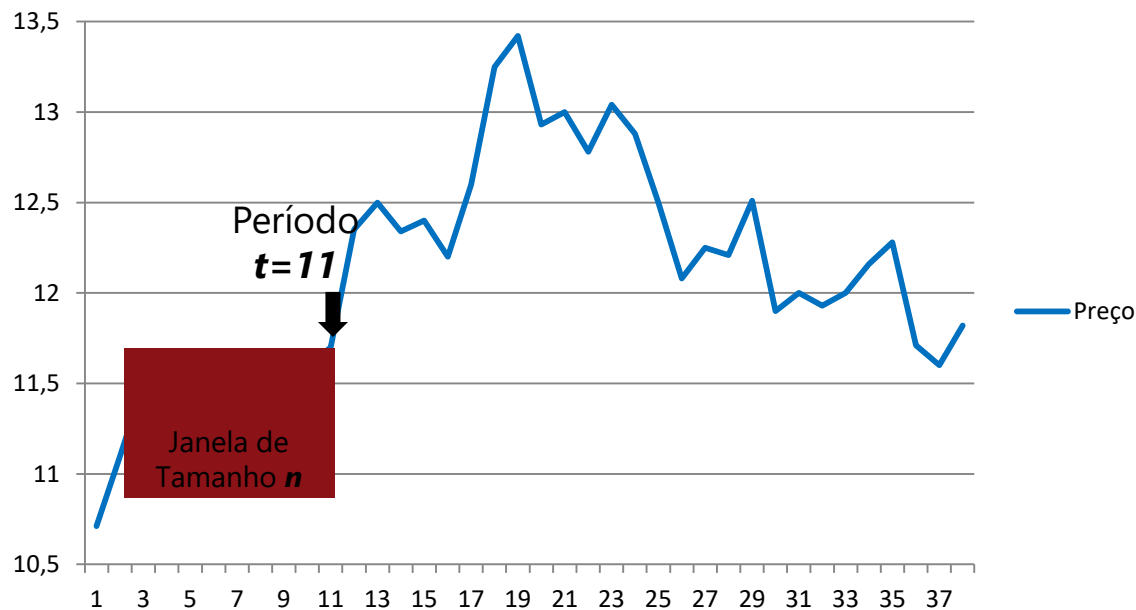
□ Simples (SMA):

$$SMA_t = \frac{\sum_{i=0}^n C_{t-i}}{n}$$

Fundamentação Teórica

□ Transformação – Extração de Indicadores Técnicos

□ Médias Móveis



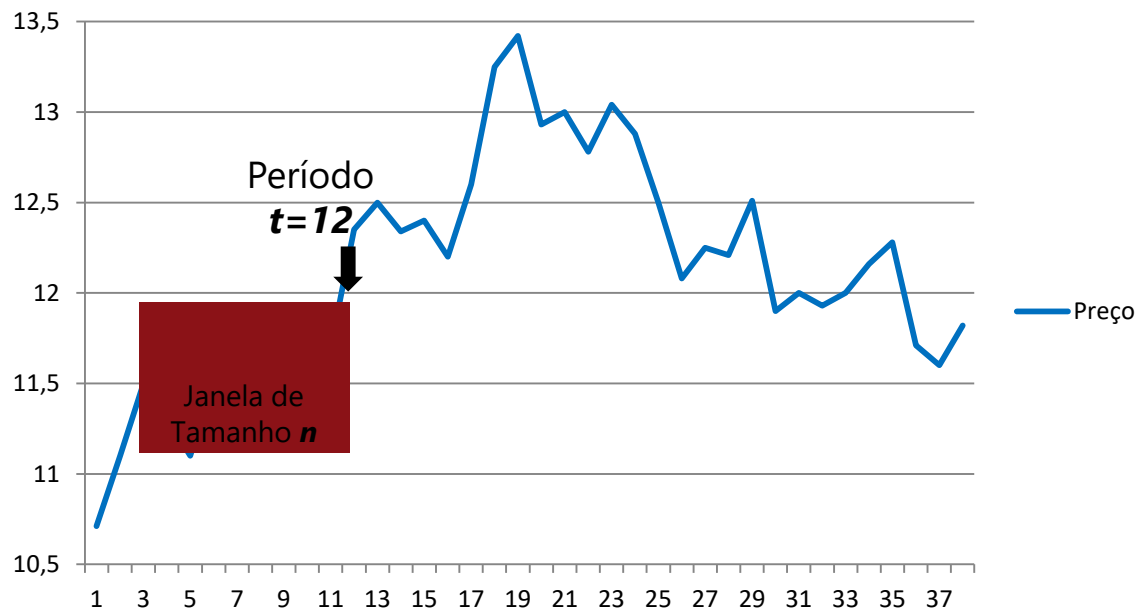
□ Simples (SMA):

$$SMA_t = \frac{\sum_{i=0}^n C_{t-i}}{n}$$

Fundamentação Teórica

Transformação – Extração de Indicadores Técnicos

Médias Móveis

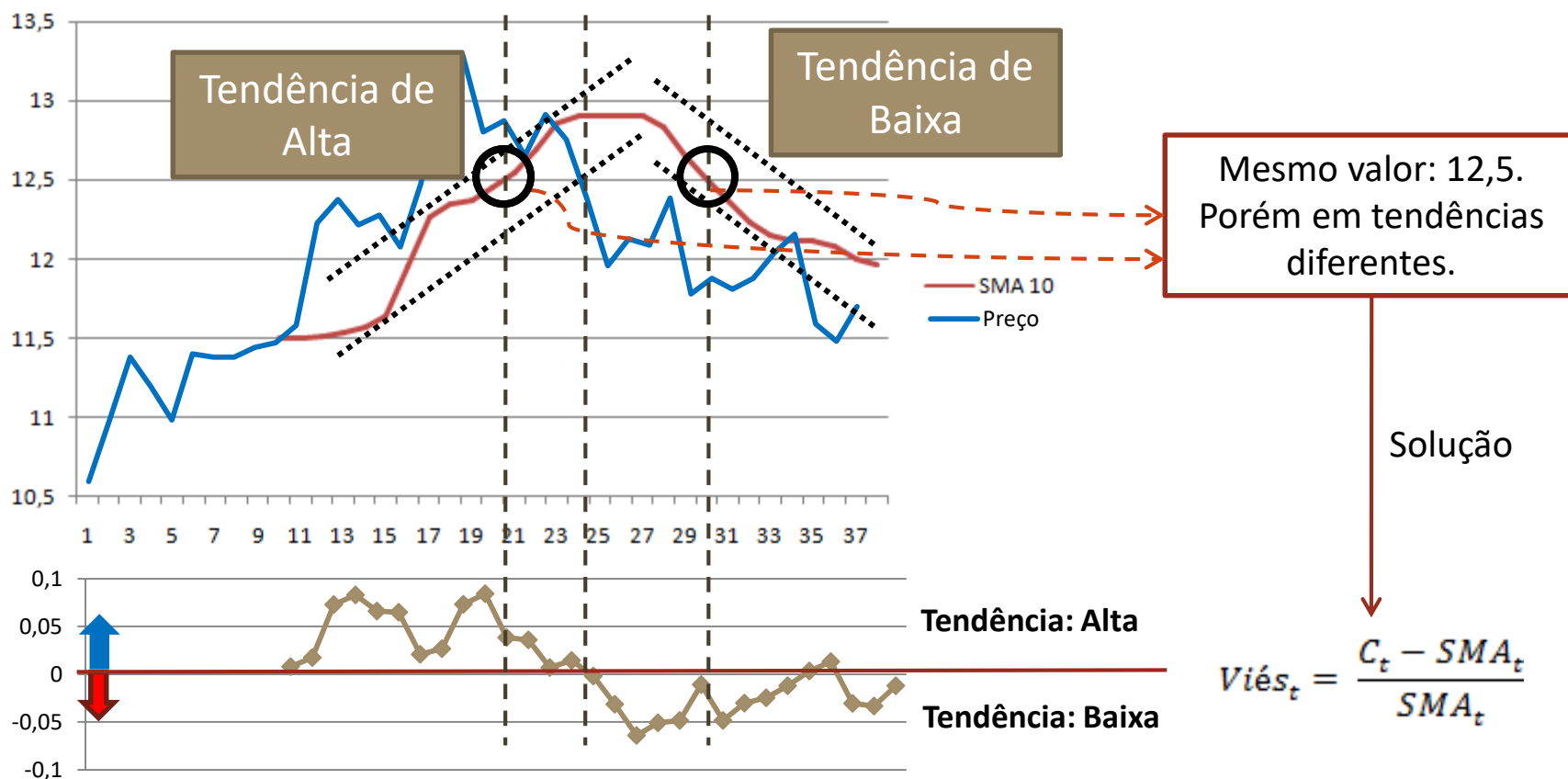


Simple (SMA):

$$SMA_t = \frac{\sum_{i=0}^n C_{t-i}}{n}$$

Fundamentação Teórica

Transformação – Médias Móveis



Fundamentação Teórica

- Transformação – Outros Indicadores
 - Taxas de mudança do preço, podendo indicar um movimento tendencioso
 - *Commodity Channel Index (CCI)*
 - Níveis de sobre-compra e sobre-venda
 - Williams %R, *Relative Strength Index (RSI)*, *Aroon Up* e *Aroon Down*
 - Apontam tendências e reversões
 - Estocásticos %K e %D, Moving Average Convergence/Divergence (MACD), Indicadores Direcionais (ADX, +DI, -DI), Weighted Moving Average (WMA) e Psychological Line (PSY)

Fundamentação Teórica

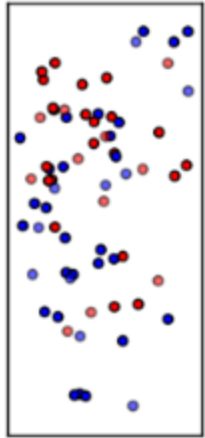
- Transformação
 - ▣ Z-normalization / Standardization

$$x' = \frac{(x - \mu)}{\sigma}$$

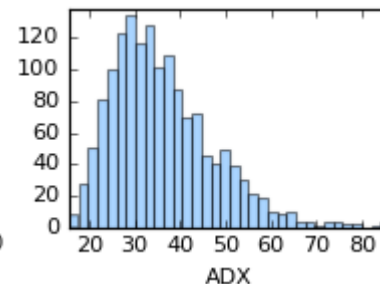
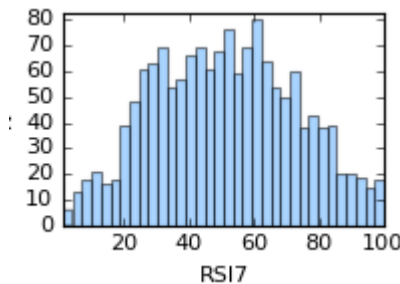
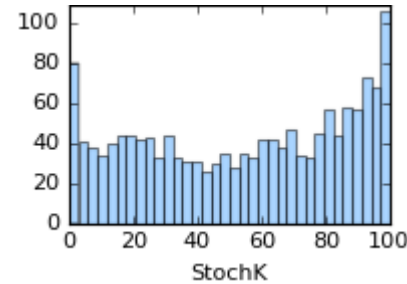
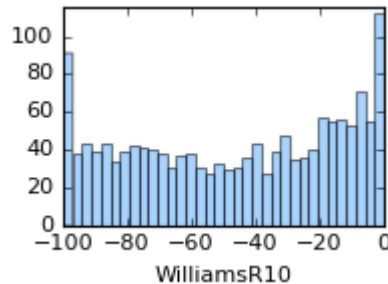
Fundamentação Teórica

Característica dos Dados

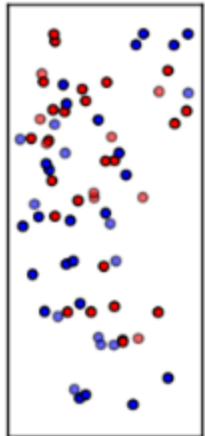
ADX - StochK



- Não linear;
- Complexa;
- Dados não paramétricos.

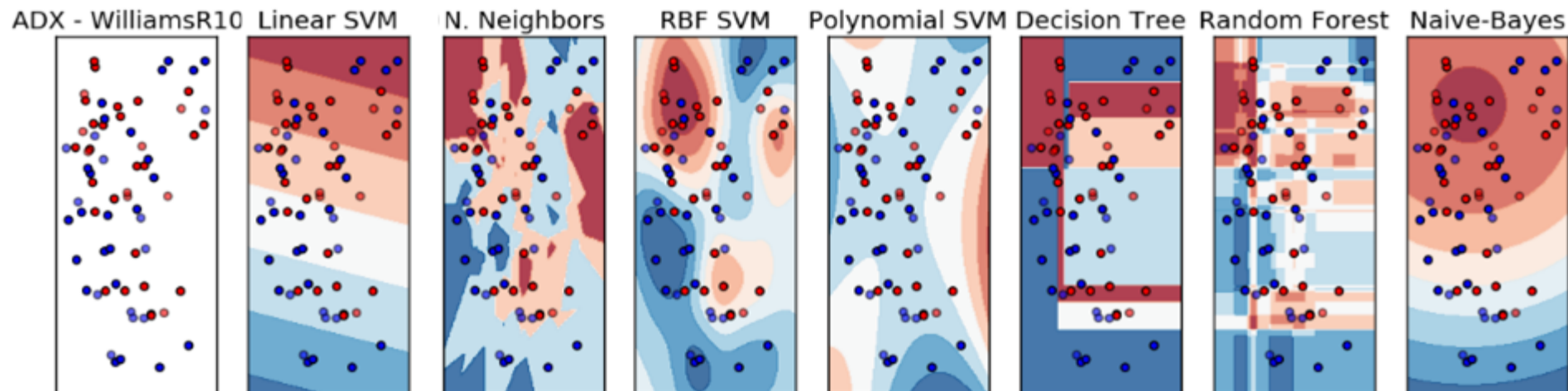
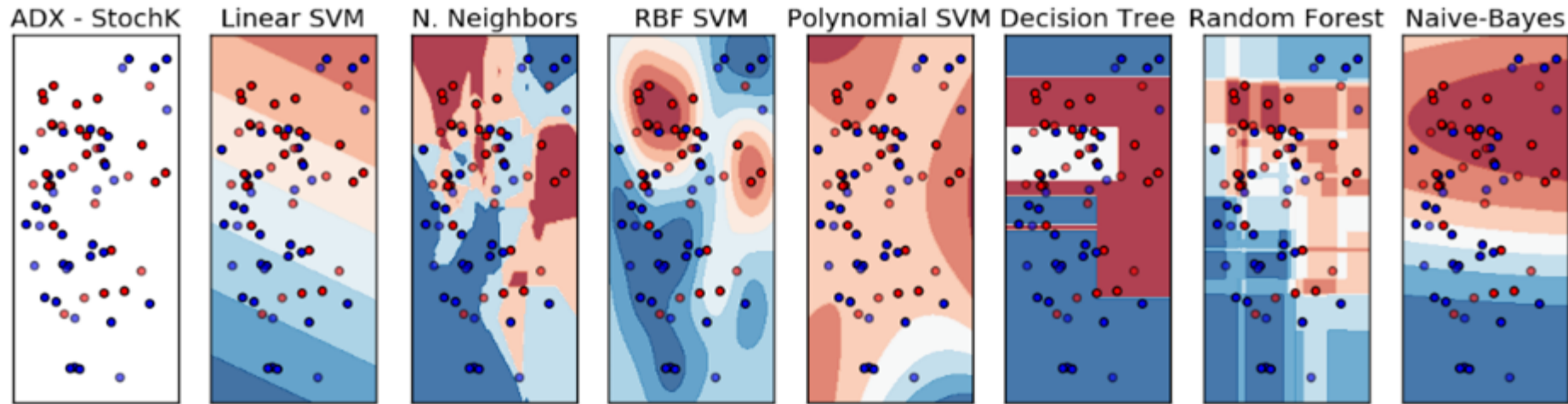


ADX - WilliamsR10



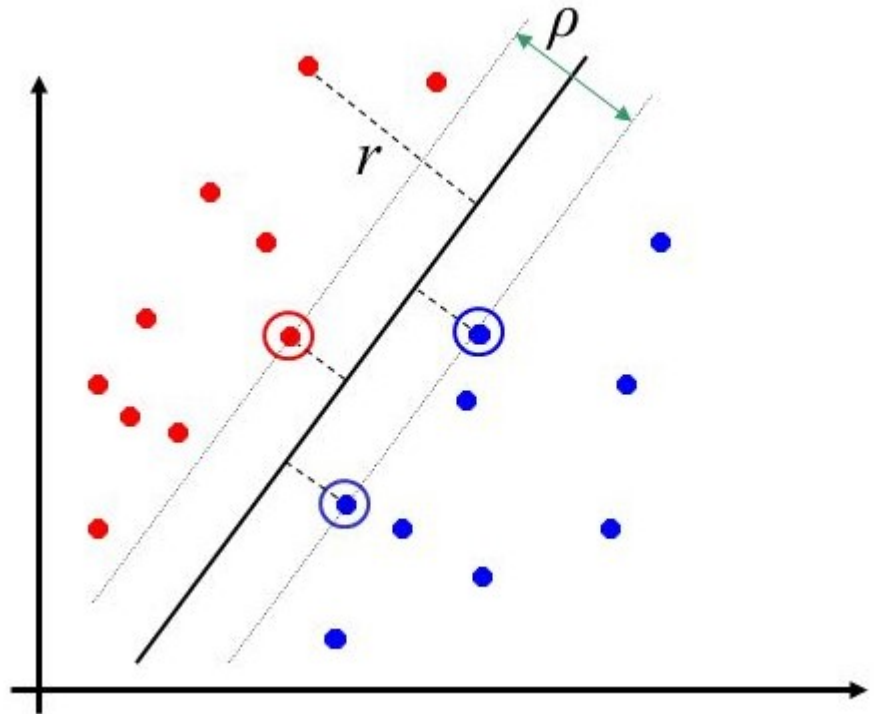
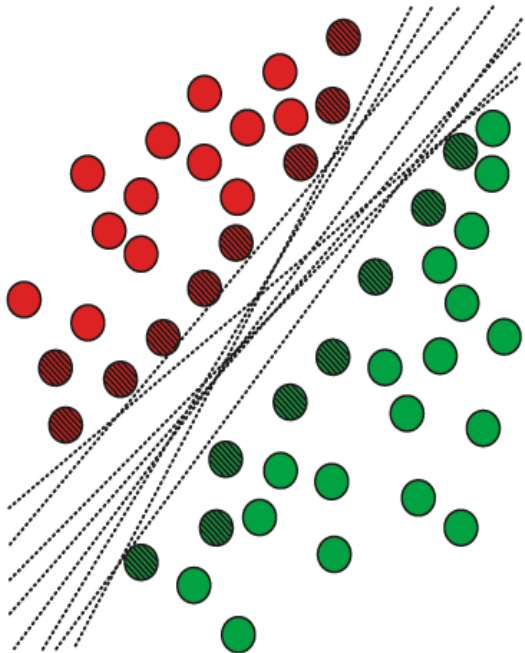
Fundamentação Teórica

Característica dos Dados



SVM

- Structural Risk Minimization (SRM)
 - ▣ Statistical learning theory for controlling generalization to determine the ideal tradeoff between structural complexity and empirical risk

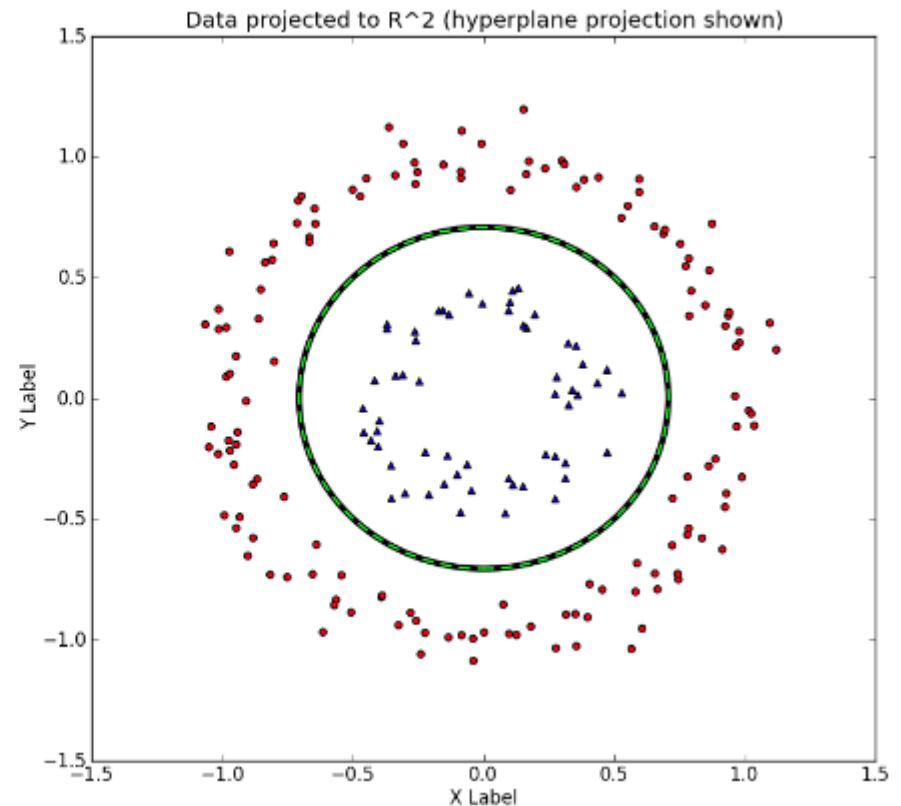
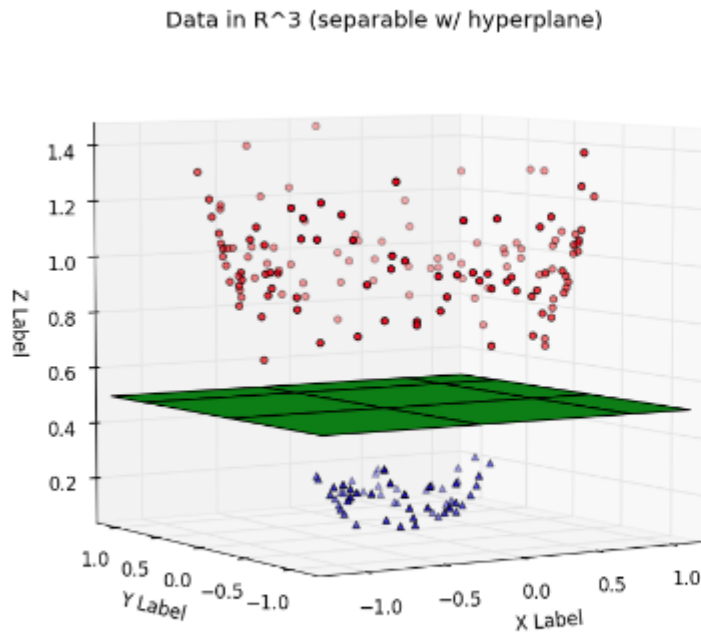


SVM

□ High dimensional feature mapping

□ $f(x) = \text{sgn} \left(\sum_{i=1}^N y_i \alpha_i \cdot k(x_j, x_i) + b \right)$

□ Radial Basis Function: $k(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|)^d$



Previsão em Ações - Objetivos

- ❑ Criar um método de classificação do dia seguinte em ações da bolsa de valores em “alta” ou “baixa”.
- ❑ Obter o melhor conjunto de atributos (*Feature Engineering*);
- ❑ Simulação de compra e venda de acordo com os sinais gerados pelo modelo.
- ❑ Gerar um modelo generalista e único que possa classificar qualquer ação, não apenas a ação que foi treinado.

Método

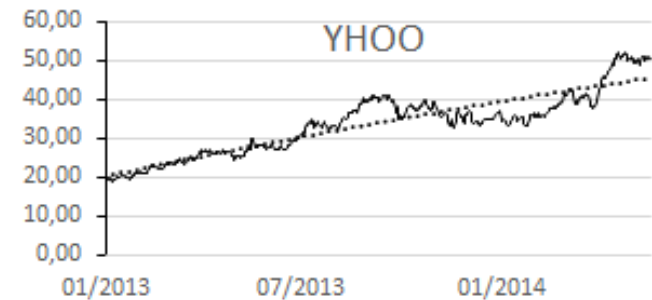
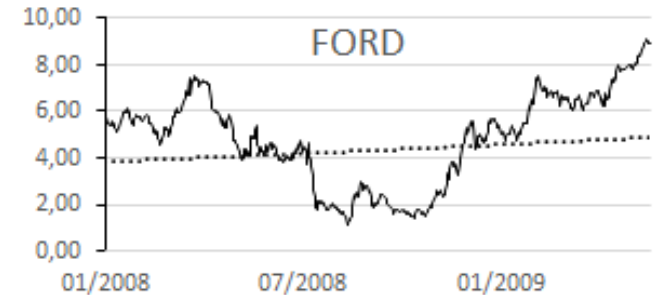
- Feature Engineering
 - ▣ Genetic Algorithm for Time Window Length Optimization – (GATWLO)
 - Encontra o melhor conjunto de tamanhos de janelas de tempo para os atributos além de selecionar apenas os atributos mais relevantes
- Conjuntos de Dados de Treinamento
 - ▣ Ações das bolsas Down Jones:
 - Microsoft (código: MSFT);
 - Nike (código: NKE);
 - Intel (código: INTC);
 - Goldman Sachs (código: GS).
 - ▣ Janeiro 2000 - Dezembro 2014 – Amostragem de 50% com balanceamento por classe.
 - Desta amostragem, 70% foi utilizado para treinamento e 30% para testes do classificador.

Método

□ Conjuntos de Dados Simulação de Negociações

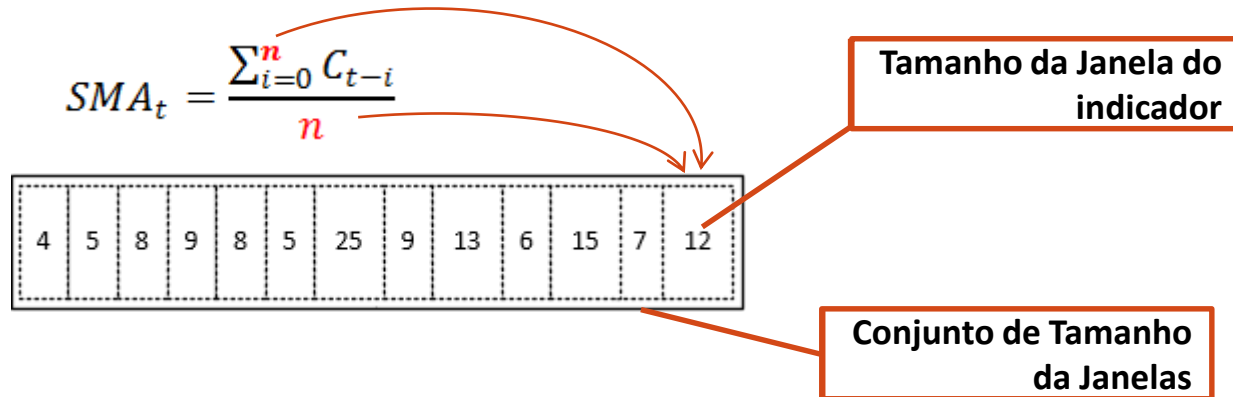
■ Ações das bolsas Down Jones:

- YHOO: 01/01/2013 – 12/31/2014.
Trend: Up-trend;
- FORD: 01/01/2008 – 12/31/2009.
Trend: Side-trend;
- JPM: 11/01/2006 – 10/31/2008.
Trend: Down-trend.



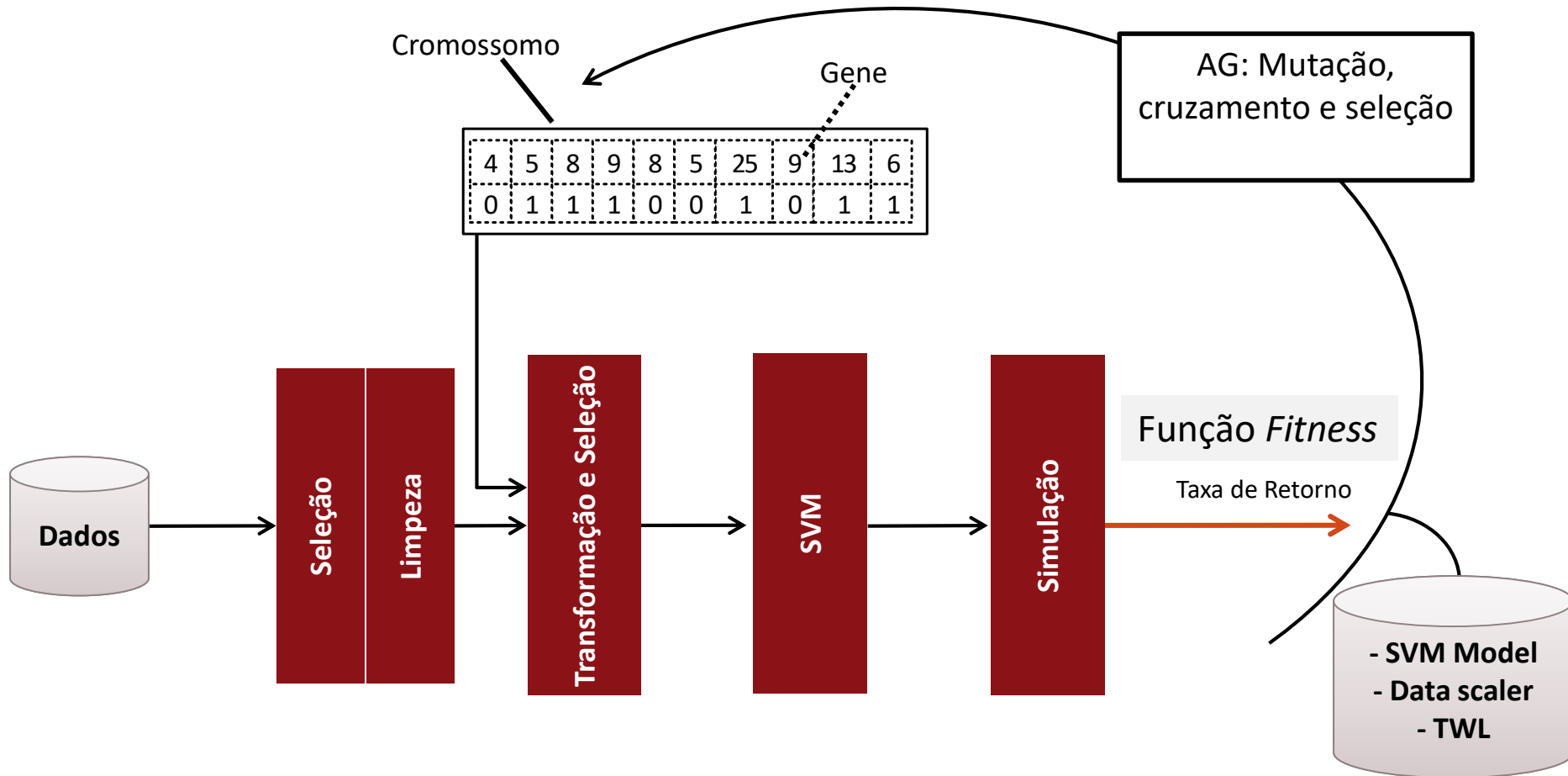
Método

□ GA for Feature Optimization



Método

GA for Feature Optimization



Método

□ Avaliação – Algoritmo de Negociação

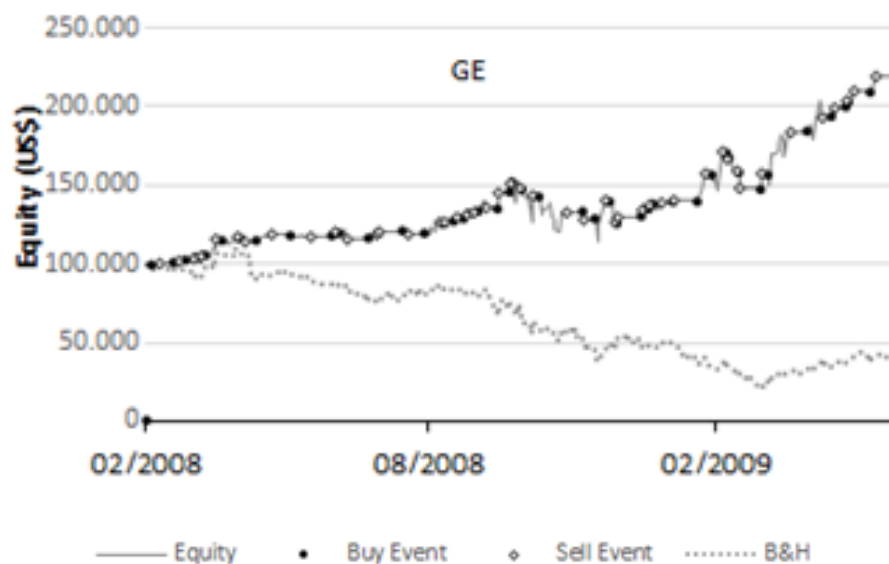
```
If  $f(x')$  = yes
    If amount stocks = 0
        Then buy
    else
        Then hold
If  $f(x')$  = no
    if amount stocks > 0
        Then sell all
```

$$RR(x) = \frac{TP(x)}{Inv} * 100$$

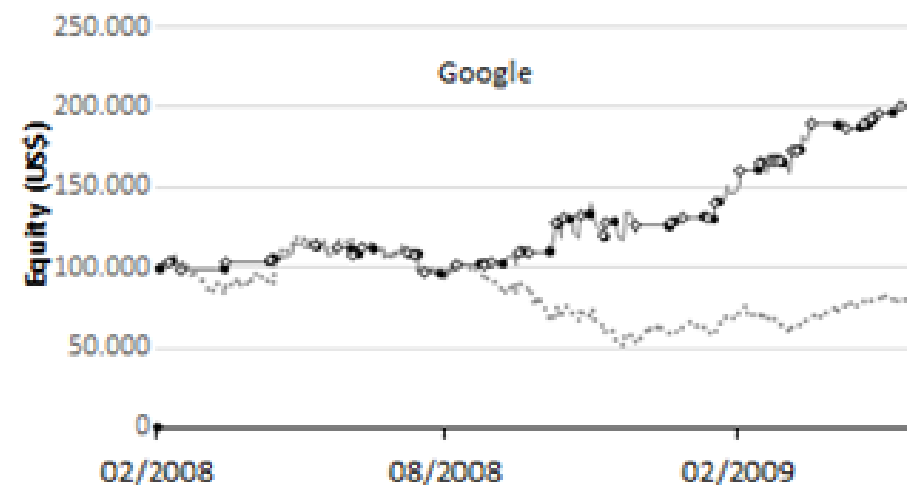
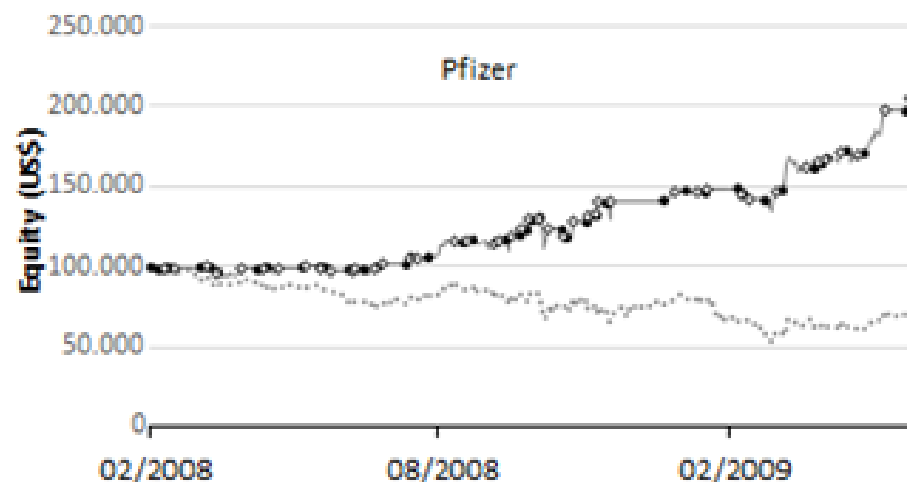
Experimentos Realizados

□ Simulação

□ Stress test



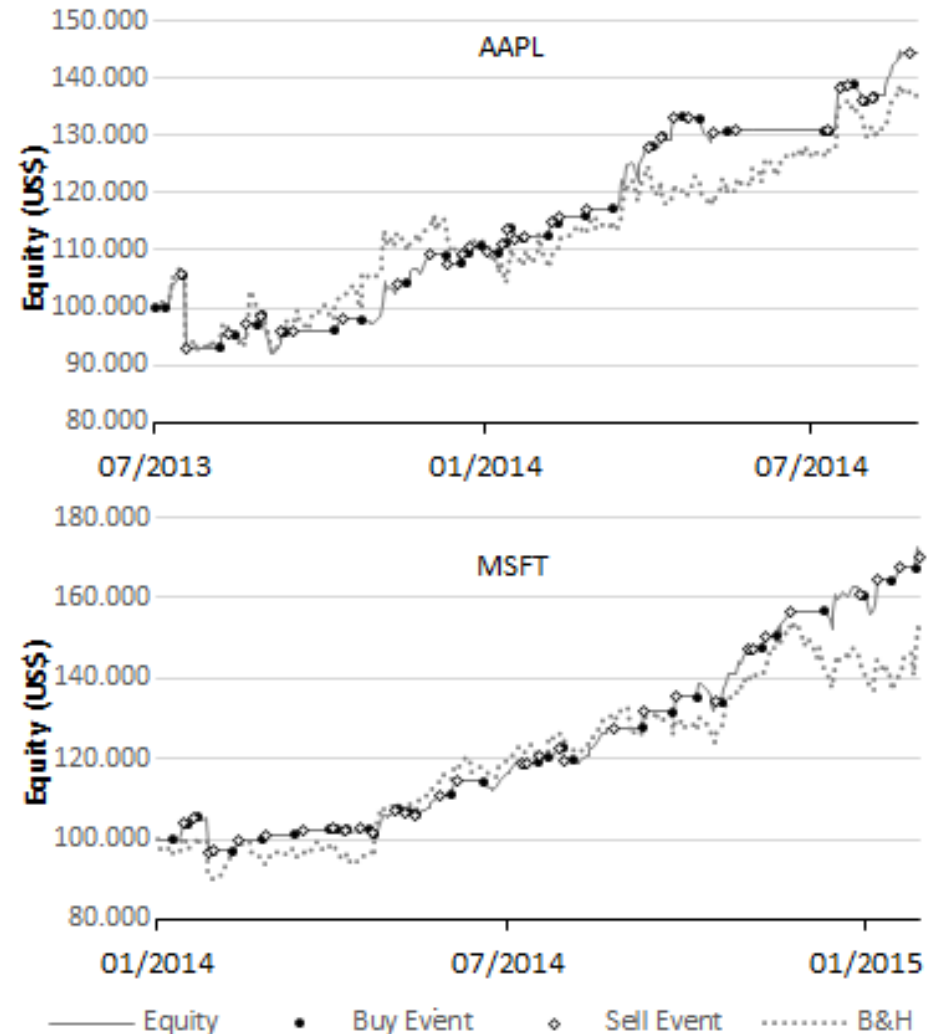
	RR [%]	MDD [%]	RR B&H [%]	MDD B&H [%]
GE	118.9	-25.68	-58.65	-81.37
Pfizer	104.1	-14.49	-29.22	-46.28
Google	106	-18.57	-15.78	-56.73



Experimentos Realizados

- Simulação
 - *Up-trend rally*

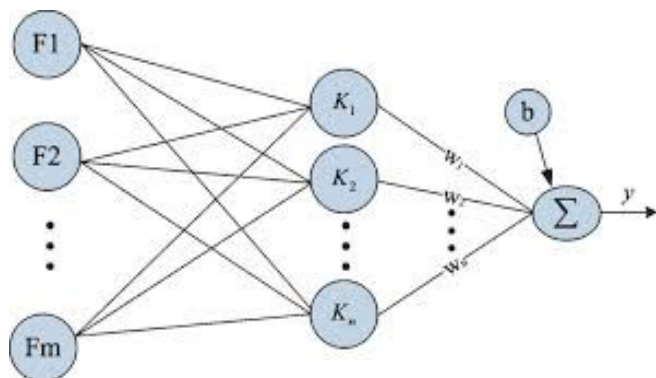
	RR[%]	MDD[%]	RR B&H [%]	MDD B&H [%]
AAPL	69.8	-8.93	51.4	-10.93
MSFT	43.9	-13	38.7	-13.51



Redes Neurais Recursivas

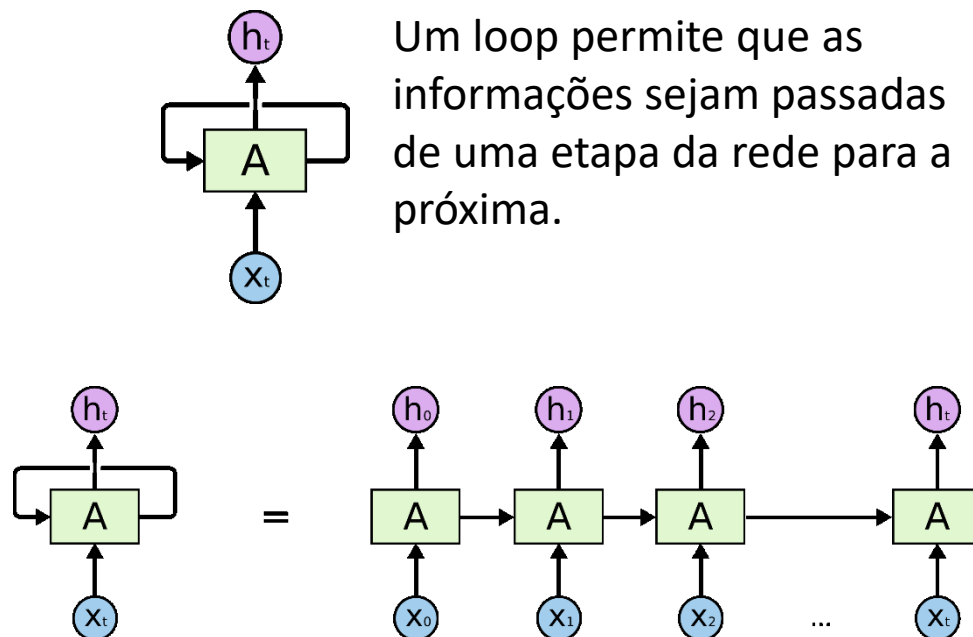
Redes Neurais Recorrentes (Recurrent Neural Network - RNN) levam em consideração a sequência de eventos para classificação.

NN Tradicionais



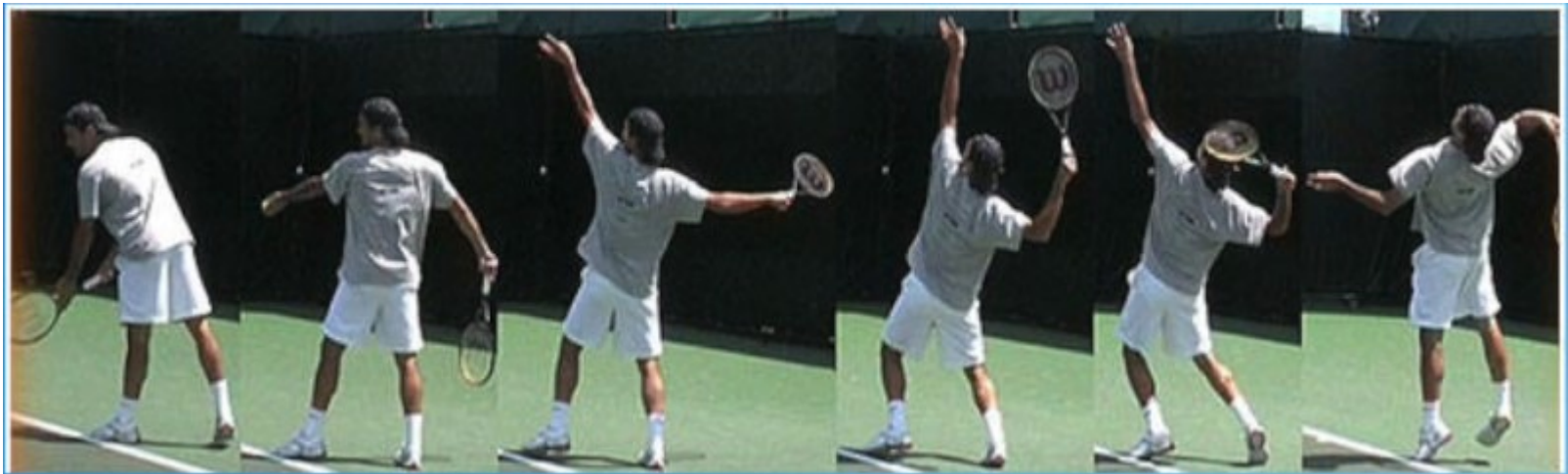
RNN

Um loop permite que as informações sejam passadas de uma etapa da rede para a próxima.



Redes Neurais Recursivas

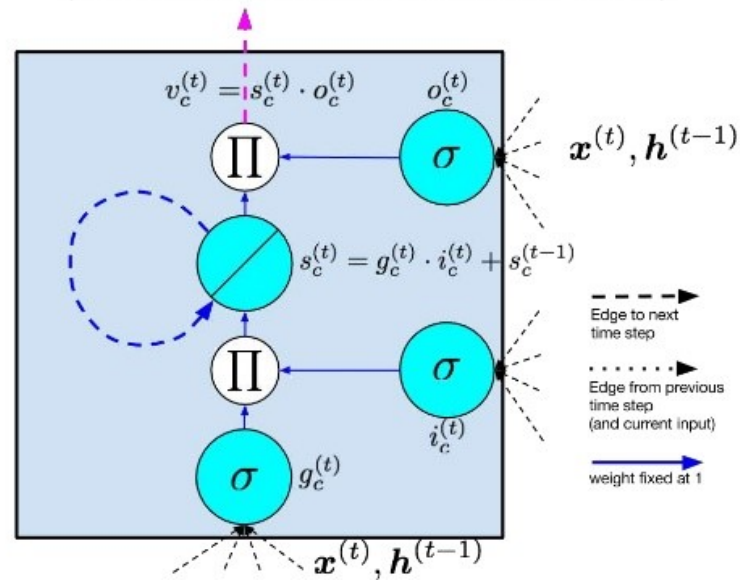
RNN levam em consideração a sequencia de eventos para classificação.



Redes Long Short-Term Memory

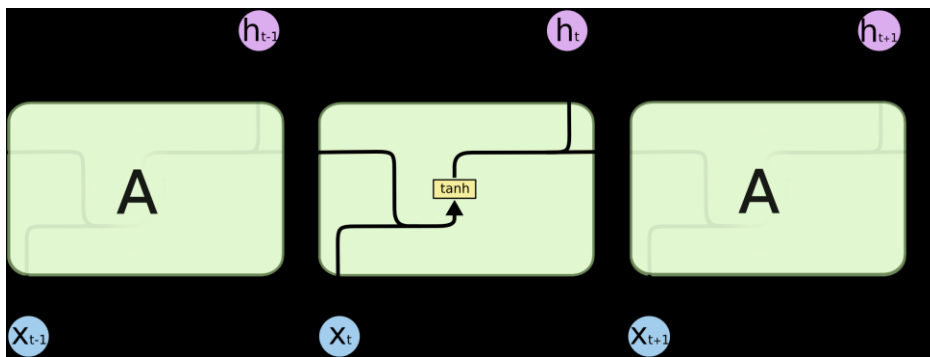
LSTMs foram desenhadas para evitar o problema de dependências de longos períodos. (*Long-term dependency*)

LSTM Memory Cell (Hochreiter & Schmidhuber, 1997)

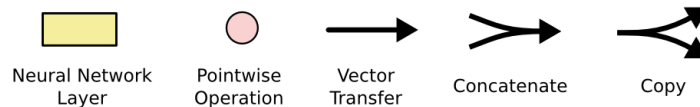
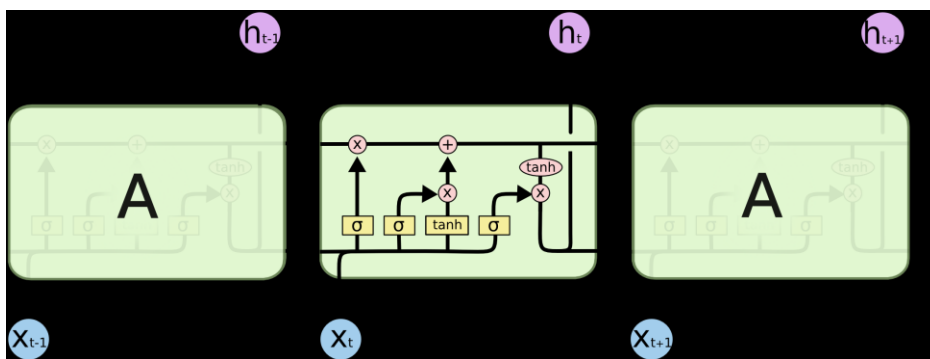


Redes Long Short-Term Memory

RNN



LSTM



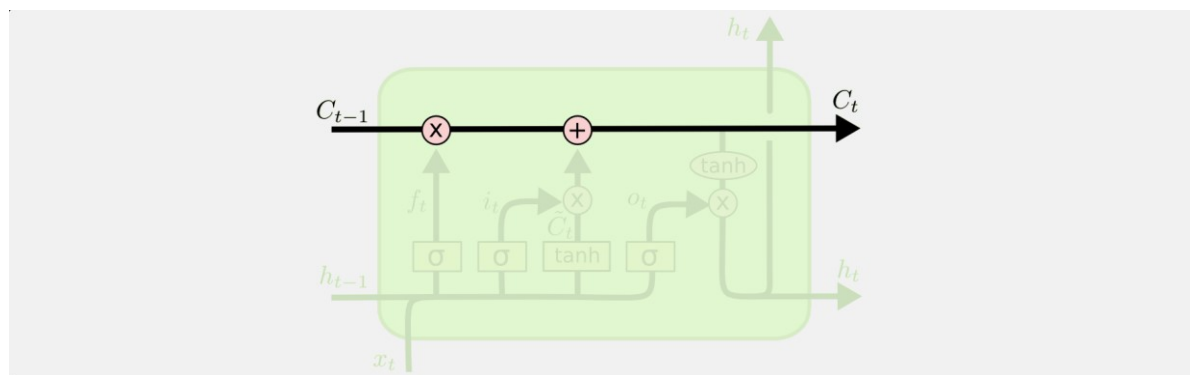
LSTMs também se organizam em cadeia. Porém com 4 camadas de redes neurais internas.

Redes Long Short-Term Memory

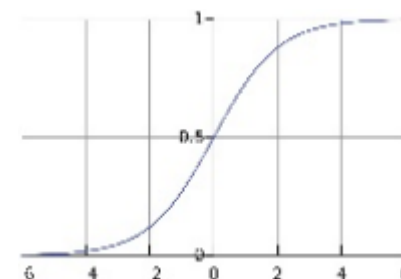
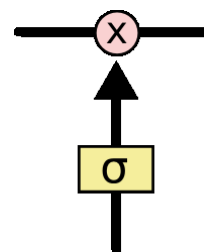
Os estados das células fluem horizontalmente entre os blocos

Os *gates* controlam a alteração de estado de C

C_{t-1} - Cell state



Gates



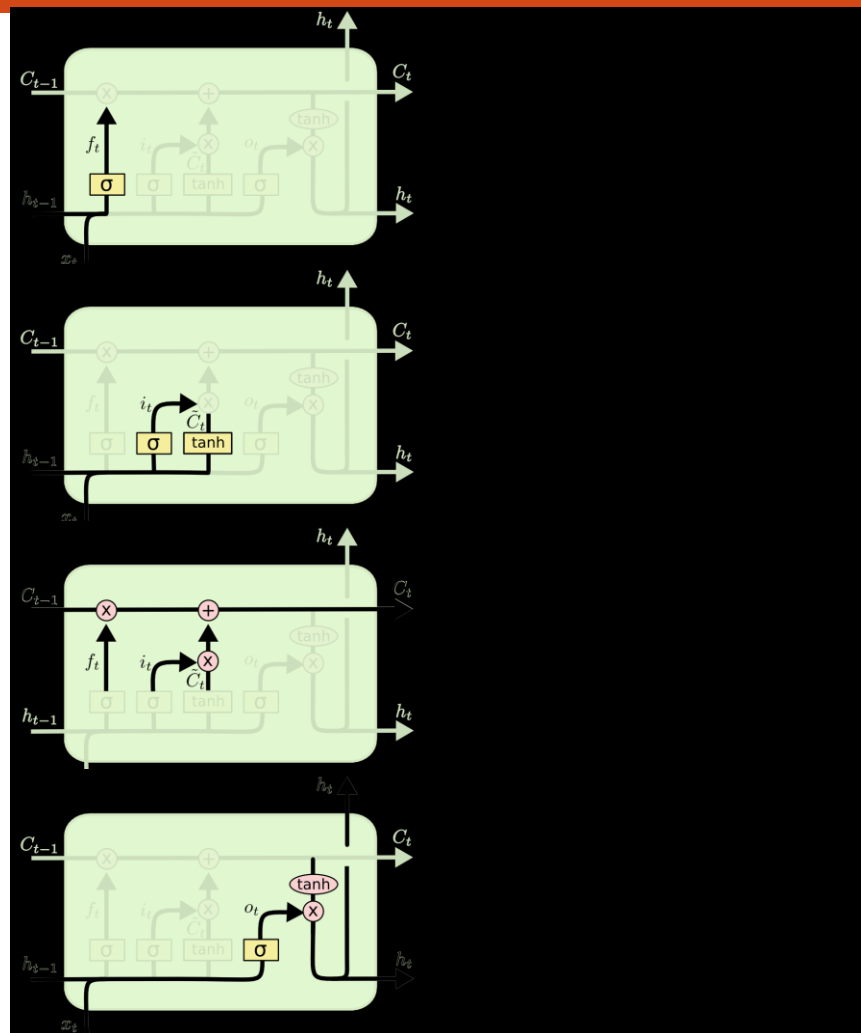
Redes Long Short-Term Memory

Forget gate: Controla o que será esquecido do estado

As próximas duas camadas controlam o que será armazenado no estado da célula

Atualiza o estado de C_{t-1} para o C_t

Decide qual a saída baseado um sigmoid da saída anterior com o novo valor armazenado na célula



Redes Long Short-Term Memory

Experimento:

```
model = Sequential()  
model.add(LSTM(input_dim=5, output_dim=100,  
    return_sequences=True))  
model.add(Dropout(0.25))  
  
model.add(LSTM(  
    output_dim=150, return_sequences=False))  
model.add(Dropout(0.4))  
  
model.add(Dense(  
    output_dim=1))  
model.add(Activation("sigmoid"))
```


Experimentos Realizados

- Resultados Finais
 - Resultados das Negociações

SVM

	RR[%]	RR B&H [%]
AAPL	69,8	51.4
MSFT	43,9	38.7

	RR [%]	RR B&H [%]
GE	118.9	-58.65
Pfizer	104.1	-29.22
Google	106	-15.78

LSTM

	RR[%]	RR B&H [%]
AAPL	120,6	51.4
MSFT	40,7	38.7

	RR [%]	RR B&H [%]
GE	96,1	-58.65
Pfizer	117,5	-29.22
Google	52,4	-15.78

Considerações Finais

□ Conclusões

- Tipo de informação complexa e de difícil previsão
- A otimização de janelas obteve bons resultados
 - A maioria dos atributos reduziram o tamanho da janela em relação aos padrões deste mercado
- LSTM tem um bom potencial, ainda é necessário otimização na arquitetura e alguns parâmetros