

ESTUDO COMPARATIVO ENTRE REDES NEURAIS E MÁQUINAS DE VETORES DE SUPORTE NA PREVISÃO DE SÉRIES ECONÔMICAS

Mêuser Valença¹

Vítor Rafael Ferraz Ferreira²

Resumo

Apresenta-se, neste texto, uma análise dos usos das redes neurais artificiais, especificamente a rede Multi-Layer Perceptron (MLP), e das máquinas de vetores de suporte, especificamente a Sequential Minimal Optimization (SMO). Serão apresentados os conceitos básicos de redes MLP assim como alguns conceitos das SVMs. Foi definido um modelo de MLP e um modelo de SVM. Será mostrado um pouco sobre o mercado de câmbio no mundo. Usou-se os métodos para avaliar seus desempenhos na previsão da cotação do euro em relação ao dólar usando o período de 01/08/2017 à 31/08/2017 com intervalo entre as cotações de 30 minutos. Tentou-se prever a menor cotação um período a frente. Os resultados mostram que a SVM SMO teve melhor desempenho.

Palavras-chave: SVM; SMO; MLP; Multi-Layer Perceptron; Sequential Minimal Optimization

1. Introdução

Segundo relatório trienal do *Bank for International Settlements* o mercado de câmbio global pode ser considerado o maior mercado em volume negociado e também o de maior liquidez. Operando um volume médio de 5,1 trilhões de dólares em abril de 2016. O dólar americano esteve envolvido em um dos lados da negociação em 88% das vezes, enquanto o euro esteve presente em 31% do montante.

Sendo assim o dólar e o euro individualmente são as duas moedas que foram negociadas em maior volume. Estando em mais da metade das negociações. A venda de euros e compra de dólares ou compra de euros e venda de dólares também chamada de par EUR/USD ou USD/EUR representaram 23% de todo o volume operado no FOREX.

¹ Escola Politécnica de Pernambuco – Universidade de Pernambuco – UPE – Professor do curso de Engenharia da computação da Escola Politécnica de Pernambuco. meuser@ecomp.poli.br

² Escola Politécnica de Pernambuco – Universidade de Pernambuco – UPE – Aluno do curso de Engenharia da computação da Escola Politécnica de Pernambuco. vrff@ecomp.poli.br

Devido então a expressividade deste mercado e a grande importância do câmbio que impacta desde a microeconomia até as políticas macroeconômicas é importante a existências de trabalhos científicos para previsão da cotação das moedas. Trabalhos que podem ajudar dos pequenos investidores particulares até os governos dos países.

Este trabalho aplica dois métodos que são bastantes utilizados na previsão de séries temporais e analisa os seus desempenhos na previsão da cotação do euro frente ao dólar.

O artigo está organizado da seguinte forma. Na seção 2 será exposto a fundamentação teórica dos algoritmos utilizados durante os experimentos. A terceira seção apresenta a metodologia utilizada nas coletas e preparo dos dados para serem usados nos experimentos. Os resultados e suas análises são descritas na seção 4. Na quinta seção, são apresentadas as considerações finais.

2. Fundamentação teórica

2.1. Redes neurais artificiais

As redes neurais artificiais (RNA) são assim chamadas porque foram inspiradas na forma de comunicação do sistema nervoso humano. No sistema nervoso um neurônio transmissor envia um sinal elétrico para um neurônio receptor. Se este sinal for forte o suficiente para alcançar o limite de ativação do neurônio receptor ele passará para outros neurônios um sinal elétrico.

De forma análoga funcionam as RNA. Onde um neurônio recebe um sinal e aplicará uma função, chamada função de ativação, no resultado do somatório de todos os sinais recebidos multiplicados por seus respectivos pesos. São várias as funções matemáticas que podem ser utilizadas como função de ativação. Mas as mais comumente utilizadas são a função linear, função sigmoide logística e a tangente hiperbólica, segundo Valença [1]. O peso X_0 é 1 pois é o limite excitatório do neurônio. Após aplicado a função de ativação é repassado para o neurônio da camada seguinte o resultado da função conforme ilustrado na figura 1.

Em uma das suas formas mais simples, uma RNA teria apenas duas camadas de neurônios. Uma camada de entrada, onde seriam recebidos os valores de entrada do problema em questão e uma camada de saída que seriam os valores de saída do sistema.

Nessa camada de saída todos os neurônios receberiam os sinais (valores) de todos os neurônios da camada de entrada.

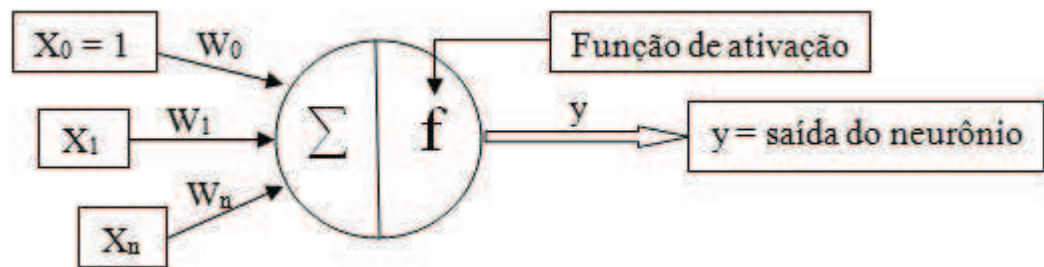


Figura 1: Ilustração do modelo matemático de um neurônio artificial segundo McCulloch e Pitts
[Fonte: Coutinho, Lara, UTILIZANDO RESERVOIR COMPUTING PARA REALIZAR DIAGNÓSTICO PRECOCE DE MAL DE ALZHEIMER, p23.]

Só que as redes com apenas a camada de entrada e a camada de saída se mostraram eficientes apenas para os problemas linearmente separáveis. Exemplos de redes com apenas duas camadas são a Perceptron e Adaline. O que diferencia essas duas redes é a função de ativação.

Devido à necessidade de resolução de problemas que não são linearmente separáveis descobriu-se as redes neurais Multi-Layer Perceptron (MLP). Que além das camadas de entrada e saída tem também uma ou mais camadas intermediárias de neurônio. A presença de camadas intermediárias permite a aproximação de qualquer função contínua e de qualquer função matemática dependendo da quantidade de camadas intermediárias existente na MLP. HAYKIN [2]

As redes MLP são bastante utilizadas para classificação e previsão de dados. As redes MLP devem ter uma boa capacidade de generalização. Deve-se treinar a rede para que ela obtenha os melhores resultados.. O treinamento de uma rede consiste em ajustar os pesos entre as camadas de neurônios.

Um algoritmo que é bastante utilizado para ajustar os pesos é o *backpropagation*. Esse algoritmo é dividido em duas fases. Na primeira fase os sinais vão passando da camada de entrada até chegar na camada de saída. Na segunda fase o resultado da camada de saída é comparado com o valor desejado. E dependendo da diferença entre o valor calculado e valor desejado os pesos entre os neurônios das camadas vão sendo ajustados.

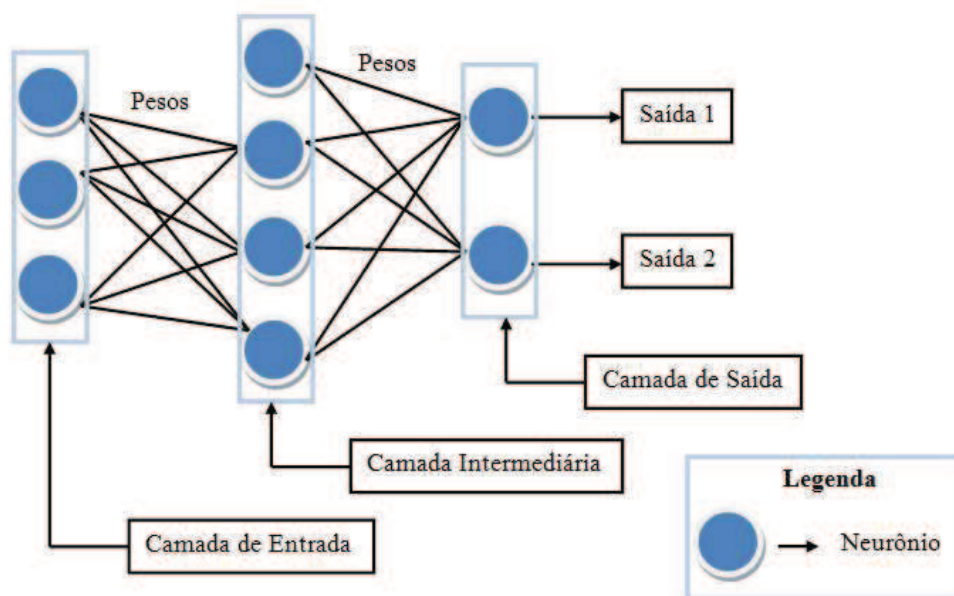


Figura 2 – Representação de uma rede MLP com uma camada intermediária.

[Fonte: Coutinho, Lara, UTILIZANDO RESERVOIR COMPUTING PARA REALIZAR DIAGNÓSTICO PRECOCE DE MAL DE ALZHEIMER, p25.]

2.2. Máquinas de vetores de suporte

As máquinas de vetores de suporte (SVMs, do inglês Support Vector Machines) é um algoritmo de aprendizado de máquina que assim como as redes MLP são utilizadas para a classificação e previsão de dados. As SVM são fundamentadas na teoria de aprendizado estatístico de Vapnik[3]. Essa teoria define parâmetros obrigatórios para que sejam garantidas uma boa capacidade de generalização.

Assim como as redes neurais MLP as SVMs são exemplos de técnicas de aprendizagem de máquinas que utilizam o aprendizado supervisionado. Ou seja, há como mensurar o desempenho até chegar a um modelo que apresente resultados satisfatórios. Nesse tipo de aprendizado o algoritmo sabe o quanto está próximo do resultado esperado. É importante as técnicas apresentarem capacidade de generalização por que tem que ter um bom índice de acerto quando apresentado novos dados e não somente com os dados apresentados durante o treinamento.

O grande desafio do algoritmo é dividir o grupo de dados através de hiperplanos. Para que a partir daí possa tomar as decisões para prever e classificar os dados. As primeiras SVMs lidavam apenas com problemas linearmente separáveis. Por exemplo as

SVMs de margens rígidas não permitiam que houvessem dados entre as margens ou vetores de suporte.

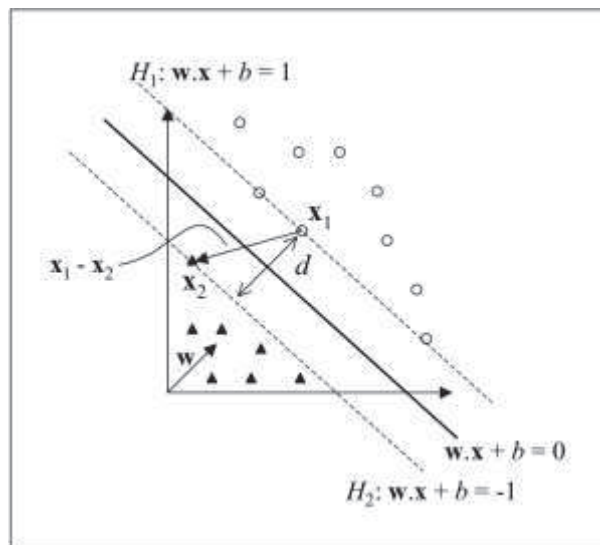


Figura 3. Exemplo de SVM com margem rígida

[Fonte: Lorena, Ana Carolina. Carvalho, André de, Uma introdução às Support Vector Machines, p12.]

A figura 3 é um exemplo de uma SVM com margens rígidas. Os hiperplanos H_1 e H_2 seriam as margens. O objetivo do algoritmo do SVM é justamente encontrar um hiperplano separador entre as margens com a distância máxima possível da reta para as margens. No caso das SVMs com margens suaves há uma maior flexibilidade para que os dados permaneçam entre os hiperplanos H_1 e H_2 no exemplo da figura 3. Ainda na figura 3 X_1 e X_2 seriam as margens ou vetores de suporte.

Como no mundo real os problemas em sua maioria não são linearmente separáveis as SVMs lineares em muitos casos não conseguem encontrar um hiperplano que divida satisfatoriamente os dados. A solução encontrada foi mapear os dados do seu espaço original para um novo espaço de dimensão maior. Segundo Cover [4] quando a transformação é não linear e a dimensão do espaço de características seja suficientemente alta é grande a probabilidade encontrar um hiperplano que divida os dados.

2.3. FOREX

FOREign EXchange (FOREX) é como é chamado o mercado global de câmbio de moedas. É o mercado que movimenta o maior volume de dinheiro do mundo. Nesse mercado estão os bancos centrais dos países, os bancos comerciais de todo o mundo, além

das gigantescas empresas multinacionais e os investidores individuais. O FOREX já rompeu a barreira de negociar 5.1 trilhões de dólares em um único dia.

O FOREX envolve as transações entre todas as moedas do mundo. Dependendo apenas de ter alguma corretora que aceite intermediar as negociações. As corretoras cobram um pequeno valor que já é adicionado sobre a cotação do valor de venda. No nosso trabalho escolhemos tratar apenas do par euro e dólar. Ou seja, tentaremos prever a cotação do euro frente ao dólar.

3. Metodologia

3.1. Base de dados e pré-processamento dos dados

Foi escolhido para esse trabalho a cotação com intervalo de meia hora do dia 01/08/2017 às 00:00 até o dia 31/08/2017 15:00. Obteve-se os dados da corretora FxPRO. Para cada período coletou-se a cotação de abertura, a cotação de encerramento, a maior cotação e a menor cotação além da quantidade de transações realizadas.

Criou-se uma matriz das cotações e volume de operações com uma defasagem de 12 períodos (ciclos de meia hora) para cada tipo de dado (cotação de abertura, menor cotação, maior cotação, cotação de encerramento e o volume de transações). Através dessa matriz foram obtidas as variáveis que mais influenciaram nas previsões.

Após a defasagem foi feita a correlação entre as 60 variáveis (12 de cada um dos cinco tipos) que foram escolhidas inicialmente para serem os valores de entrada das redes neurais e a variável de saída. Nesse caso é a menor cotação um período a frente. A correlação serve para determinar quais das variáveis de entrada tem maior influência na variável a ser prevista.

Escolheu-se trabalhar apenas com as variáveis que apresentaram correlação acima de 0,9. Atenderam a esse pré-requisito apenas as 10 cotações mais recente de cada. E nenhum dos 12 valores referentes ao volume de transações apresentou a correlação acima de 0,9. Ficando então 40 variáveis (10 cotações de abertura, 10 de menor cotação, 10 de maior cotação e 10 cotações de encerramento) de entrada para prever a menor cotação para um período de meia hora a frente.

Foi aplicado a normalização nos dados restantes. Processo esse que consiste em manter os dados em um intervalo numérico que seja melhor aceito pelas redes neurais pois estarão todos em uma mesma escala. A normalização impede que as redes neurais

deem mais importância a um dado que ao outro indevidamente, assim como pode impedir que as redes tenham problemas com os mínimos locais e que perca a capacidade de generalização.

3.2. Modelo da MLP

Usando o software WEKA fez-se vários testes variando a taxa de aprendizagem, a quantidade de épocas e o momentum. Em todos os testes foi aceita a sugestão do WEKA que utiliza como quantidade de neurônios na camada escondida a metade da soma da quantidade de neurônios na camada de entrada, no nosso caso 41, com a quantidade de neurônios na camada de saída, no caso 1 neurônio. Totalizando então 21 neurônios na camada intermediária.

Entre os testes tentando achar o que desse os melhores resultados chegou-se ao seguinte modelo de rede neural MLP:

Quantidade de neurônios na camada de entrada: 40

Quantidade de neurônios na camada intermediária: 21

Quantidade de neurônios na camada de saída: 1

Quantidade de épocas: 1000

Taxa de aprendizagem: 0,01

Momentum: 0,01

Definido o modelo, foram realizados trinta experimentos e foram calculados os valores do Erro Médio Quadrático (EMQ) e Erro Médio Absoluto (EMA).

3.3. Modelo da SVM

Ao utilizar o WEKA foi escolhido utilizar a função SMOReg. A função SMOReg é uma otimização das SVMs para problemas de regressão. Fez-se vários testes alterando apenas o parâmetro c (é um parâmetro que pune mais os erros cometidos pelo algoritmo) e o kernel. Após os testes ficou estabelecido o seguinte modelo de SVM:

c : 100

Kernel: RBF Kernel

Assim como no caso das redes neurais foram feitos trinta experimentos com o modelo de SVM definido e calculados os valores do erro médio quadrático e do erro médio absoluto.

3.4. Análise estatística

Após a conclusão dos experimentos e de ter os valores de duas séries de medidas. Foi feito um teste estatístico para comprovar se um método é melhor que o outro. Nesse trabalho escolheu-se o teste dos postos de Wilcoxon. Para realizar o teste foi utilizada a implementação já existente na linguagem R. Na biblioteca utilizada o nível de significância (α) é definido como 0,05.

No caso do *p-value* ser menor que o nível de significância escolhido se tem a hipótese alternativa. No problema abordado a hipótese alternativa indica que um algoritmo apresentou melhores resultado que o outro algoritmo. E no caso do *p-value* ser maior ou igual a 0,05 se tem a hipótese nula. Que equivaleria a dizer que a SVM e a MLP apresentaram resultados estatisticamente iguais.

4. Resultados

Os resultados da MLP após os experimentos foram os seguintes.

	Multi-Layer Perceptron (MLP)	
	Erro médio absoluto	Erro médio quadrático
1	0,0173	0,0297
2	0,0164	0,0289
3	0,0168	0,029
4	0,0161	0,0284
5	0,0164	0,029
6	0,017	0,0293
7	0,0161	0,0286
8	0,0158	0,0284
9	0,016	0,0292
10	0,0168	0,0294
11	0,0163	0,0289
12	0,0172	0,0297
13	0,0167	0,0291
14	0,0161	0,0285
15	0,016	0,0287

	Multi-Layer Perceptron (MLP)	
	Erro médio absoluto	Erro médio quadrático
16	0,0162	0,0289
17	0,0163	0,0288
18	0,0173	0,0294
19	0,0166	0,0288
20	0,0163	0,0288
21	0,0162	0,0287
22	0,0179	0,0301
23	0,0183	0,0301
24	0,0161	0,029
25	0,0166	0,029
26	0,0158	0,0287
27	0,0168	0,0291
28	0,0165	0,0294
29	0,0168	0,0287
30	0,0165	0,0288

Os resultados da SVM após os experimentos foram os seguintes.

	Support Vector Machine (SVM)	
	Erro médio absoluto	Erro médio quadrático
1	0,0128	0,026
2	0,0129	0,0261
3	0,0127	0,026
4	0,0128	0,026
5	0,0128	0,026
6	0,0128	0,0259
7	0,0127	0,026
8	0,0127	0,026
9	0,0128	0,026
10	0,0127	0,0259
11	0,0128	0,026
12	0,0129	0,026
13	0,0129	0,0261
14	0,0128	0,026
15	0,0128	0,0261

	Support Vector Machine (SVM)	
	Erro médio absoluto	Erro médio quadrático
16	0,0129	0,0261
17	0,0128	0,026
18	0,0127	0,0259
19	0,0128	0,026
20	0,0128	0,026
21	0,0129	0,026
22	0,0128	0,026
23	0,0129	0,0261
24	0,0129	0,026
25	0,0128	0,0259
26	0,0128	0,026
27	0,0127	0,0261
28	0,0127	0,026
29	0,0128	0,026
30	0,0129	0,0261

Foi feito o teste dos postos de Wilcoxon para os erros médios quadráticos da MLP e da SVM. E chegou-se ao seguinte resultado.

```
data: MLP_EMQ and SMO_EMQ
w = 900, p-value = 1.343e-11
alternative hypothesis: true location shift is not equal to 0
```

Figura 4 - Resultado do cálculo do p-value

Foi feito o teste dos postos de Wilcoxon para os erros médios absolutos da MLP e da SVM. E chegou-se ao seguinte resultado.

```
data: MLP_EMA and SMO_EMA
w = 900, p-value = 1.892e-11
alternative hypothesis: true location shift is not equal to 0
```

Figura 5 - Resultado do cálculo do p-value

Tanto nos erros médios quadráticos quanto nos erros médios absolutos o *p-value* é muito menor que o nível de significância estabelecido (0,05). Então para apontar estatisticamente a melhor solução devemos comparar a média dos erros.

Média dos erros médios absolutos	
MLP	SVM
0,016573333	0,012803333

Figura 6 - média aritmética dos 30 erros médios absolutos

Média dos erros médios quadráticos	
MLP	SVM
0,029036667	0,02601

Figura 7 - média aritmética dos 30 erros médios quadráticos

Conforme figura 6 e na figura 7 acima a SVM apresentou menor média tanto no erro médio quadrático quanto no erro médio absoluto. Podendo-se concluir que a SVM apresenta melhores resultados na previsão da cotação do euro frente ao dólar. A SVM pode ter apresentado melhores resultado que a MLP pois devido à convexidade do problema de otimização não apresenta o mesmo problema que as MLPs têm com os mínimos locais.

5. Conclusão

Neste trabalho foram feitas previsões das cotações do euro em relação ao dólar 30 minutos a frente utilizando as máquinas de vetores de suporte (SVM) e as redes neurais Multi-Layer Perceptron (MLP). As previsões feitas foram avaliadas utilizando o erro médio quadrático e o erro médio absoluto.

Foi observado que a eficiência nas previsões está diretamente ligada aos parâmetros escolhidos em cada algoritmo testado. No algoritmo SVM observou-se que com o aumento do valor do parâmetro ‘c’, que penaliza os erros cometidos na previsão, chegou-se a previsões com erros menores. No algoritmo MLP observou-se que aumentando o número de épocas e diminuindo o momento e a taxa de aprendizagem chegou-se a previsões com erros menores.

Quando comparado os erros apresentados pelos dois algoritmos constatou-se que as máquinas de vetores de suporte apresentaram previsões com erros menores que as previsões das redes neurais Multi-Layer Perceptron.

6. Referências bibliográficas

- [1] VALENÇA, M. **Fundamentos das Redes Neurais: exemplos em JAVA**. 2ª ed. Livro Rápido, 2009.
- [2] HAYKIN, S. **Redes Neurais: Princípios e Práticas**. [S.l.]: Bookman, 2007. JAEGER, H.
- [3] B. E. Boser, I. L. Guyon, and V. N. Vapnik. **A training algorithm for optimal margin classifiers**. In Proceedings of the 5th Annual Workshop on Computational Learning Theory, pages 144–152, Pittsburg, Pennsylvania, US, 1992.
- [4] Coutinho, Lara, **UTILIZANDO RESERVOIR COMPUTING PARA REALIZAR DIAGNÓSTICO PRECOCE DE MAL DE ALZHEIMER**
- [5] Lorena, Ana Carolina. Carvalho, André de, **Uma introdução às Support Vector Machines**
- [6] <https://www.bis.org/publ/rpfx16.htm>, acessado em 20/12/2017
- [7] Matos, Paulo; Beviláqua, Giovanni; Filho, Jaime, **Previsão do câmbio real-dólar sob um arcabouço de apreçamento de ativos**

MONOGRAFIA DE FINAL DE CURSO

Avaliação Final (para o presidente da banca)*

No dia 19 de dezembro de 2017, às 11:00 horas, reuniu-se para deliberar a defesa da monografia de conclusão de curso do discente **VITOR RAFAEL FERRAZ FERREIRA**, orientado pelo professor **Mêuser Jorge Silva Valença**, sob título **Redes neurais aplicadas na previsão de séries econômicas**, a banca composta pelos professores:

Sérgio Mário Lins Galdino

Mêuser Jorge Silva Valença

Após a apresentação da monografia e discussão entre os membros da Banca, a mesma foi considerada:

☒ Aprovada

☐ Aprovada com Restrições*

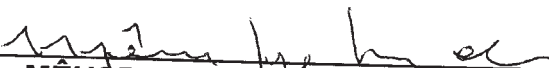
☐ Reprovada

e foi-lhe atribuída nota: 9,0 (nove)

*(Obrigatório o preenchimento do campo abaixo com comentários para o autor)

O discente terá 07 dias para entrega da versão final da monografia a contar da data deste documento.


SÉRGIO MÁRIO LINS GALDINO


MÊUSER JORGE SILVA VALENÇA

* Este documento deverá ser encadernado juntamente com a monografia em versão final.