



II CONPESQ Congresso de Pesquisa, Pós-Graduação e Inovação

Os novos rumos da ciência pós-pandemia

12 a 16 de abril de 2021 Universidade Federal do Cariri - UFCA

ANÁLISE DE MÉTODOS DE CLASSIFICAÇÃO: UMA COMPARAÇÃO ENTRE MLP, SVM E K-MEANS

Nicolas Moreira de Carvalho Gomes¹, Matheus Souza Leite², Marcos André Queiroz Machado³, Rosilda Benicio de Souza⁴.

Universidade Federal do Cariri - UFCA- Av. Tenente Raimundo Rocha N° 1639 - Bairro
Cidade Universitária, Juazeiro do Norte – CE, CEP: 63048-080.
nic982010@hotmail.com

1. INTRODUÇÃO

O uso da classificação permeia as diversas áreas do conhecimento. Sua aplicação pode ser vista em situações como diagnóstico de doenças, busca na internet, filtro de e-mails, identificação de pessoas, seleção automática de qualidade, perfil de clientes, análise de sequências de DNA, detecção de fraudes, marketing, entre muitas outras. A solução de um problema de classificação trata do agrupamento de objetos em classes de objetos similares. O grau de dificuldade para classificar um conjunto de dados em algum número de classes depende da variabilidade nos valores característicos.

Recentemente, metodologias de classificação têm sido apresentadas a partir da Inteligência Artificial (IA), um ramo da Ciência da Computação cujo objetivo é desenvolver sistemas que executam funções desempenhadas pelo ser humano usando conhecimento e raciocínio. Estas metodologias são de grande importância e um ramo usual em algumas áreas como mineração de dados, estatística, engenharia, ciências da computação entre outras ciências que trabalham com análises de dados. Alguns métodos são usuais para a tarefa de classificação, por exemplo como o *K-means* (ALAMSYAH et al., 2018; LI, 2020), que faz aglomerações a partir de um centróide definido aleatoriamente, ou uma Rede Neural Artificial (RNA) (TEH et al., 2020; UMAMAHESWARI; PRINCESS; KUMAR, 2020), que aprenderá com erros e acertos imitando o cérebro humano ou o *Support Vector Machine* (SVM) (QI; SILVESTROV; NAZIR, 2017; TEH et al., 2020; WU et al., 2018) que gera hiperplanos a partir de vetores para a classificação.

A solução de um problema de classificação trata do agrupamento de objetos em classes de objetos similares. O grau de dificuldade para classificar um conjunto de dados em algum número de classes depende da variabilidade nos valores característicos. Para um certo objeto, basicamente, dois fatores podem ter efeito sobre essa variabilidade: a complexidade dos dados e o ruído (DUDA et al, 2000). Se o conjunto de dados apresenta um baixo nível

1. Bolsista PIBITI/UFCA no Programa Institucional de Iniciação em Desenvolvimento Tecnológico e Inovação da UFCA.
2. Voluntário no Programa Institucional de Iniciação Científica e Tecnológica da UFCA.
3. Bolsista PIBIC/UFCA no Programa Institucional de Bolsa de Iniciação Científica da UFCA.
4. Orientadora.

de ruído e complexidade com relações linearmente independentes para as características, então a tarefa de classificação é um problema linearmente separável e existe um hiperplano ótimo que conduz à uma classificação com taxa de reconhecimento de 100%. Por outro lado, se o problema de classificação é um problema não linearmente separável, essa taxa de reconhecimento é frequentemente impossível de ser alcançada. Em geral, busca-se determinar a probabilidade de um dado objeto pertencer a cada uma das possíveis classes. Portanto, o problema linearmente separável é relativamente mais simples para resolver do que um problema não linearmente separável (VAPNIK, 1998; ELIZONDO, 2006).

Para o presente trabalho, analisamos oito conjuntos de dados diferentes e empregamos os métodos de classificação do *K-means*, SVM e a RNA do tipo *multilayer perceptron* (MLP). Em seguida comparamos os métodos para avaliarmos o desempenho dos três em diferentes situações. Uma vez que cada conjunto de dados tem características próprias, como dimensionalidade, número de classes, apresenta ou não sobreposição/não linearidade, é esperado atingir altas taxas de acerto na classificação de qualquer conjunto de dados, das diversas áreas do conhecimento. A plataforma computacional escolhida para o desenvolvimento do estudo foi o programa R dada as suas características mencionadas por Ritter et al (2019) como sendo uma plataforma de código aberto, gratuita e com ampla capacidade extensível, isto é, tem o potencial de realizar diversas atividades, desde que sua capacidade a comporte. Nesta perspectiva, este trabalho tem como objetivo o estudo e desenvolvimento de uma arquitetura de algoritmos que gere a eficiência máxima em acertos, e tomando como base, conjuntos de dados recorrentes na literatura.

2. MATERIAL E MÉTODOS

Foram analisados oito conjuntos de dados. A Tabela 1 descreve os dados quanto a suas características como instâncias, atributos, número de classes, nome de cada classe e a instância de cada classe. Os dados descritos de Círculos Concêntricos e Espirais Concêntricas, foram obtidos de forma simulada, enquanto que os dados de Diabetes, Câncer de mama, Coluna vertebral, Sementes, Vinho e *Yeast* foram obtidos a partir da *UCI Machine Learning Repository* (FRANK, ASUNCION, 2013).

Tabela 1 - Descrição dos dados usados para a classificação

| Dados | Instâncias | Atributos | Nº de classes | Classes | Instâncias |
|-----------------------|------------|-----------|---------------|------------------|------------|
| Círculos Concêntricos | 750 | 2 | 2 | Círculo interno | 375 |
| | | | | Círculo externo | 375 |
| Espirais Concêntricas | 750 | 2 | 2 | Espiral interna | 375 |
| | | | | Espiral externa | 375 |
| Diabetes | 768 | 8 | 2 | Positivo | 268 |
| | | | | Negativo | 500 |
| Câncer de mama | 458 | 9 | 2 | Benigno | 458 |
| | | | | Maligno | 241 |
| Coluna Vertebral | 310 | 6 | 3 | Hernia | 60 |
| | | | | Espondilolistese | 150 |
| | | | | Normal | 100 |
| Sementes | 210 | 7 | 3 | Kama | 70 |
| | | | | Rosa | 70 |
| | | | | Canadian | 70 |
| Vinho | 188 | 13 | 3 | Cultivar 1 | 59 |
| | | | | Cultivar 2 | 71 |

| | | | | | |
|--------------|------|---|---|---------------|-----|
| | | | | Cultivar 3 | 58 |
| <i>Yeast</i> | 1484 | 8 | 4 | Citoesqueleto | 463 |
| | | | | Nuclear | 429 |
| | | | | Mitocondrial | 244 |
| | | | | Outros | 348 |

Fonte: Própria dos autores.

Cada conjunto de dados escolhido oferece um desafio único para os algoritmos de classificação. Com isso podemos analisar os métodos que se mostram superiores em desempenho. No conjunto de dados gerado artificialmente dos Círculos Concêntricos, as classes possuem uma boa separação. Por outro lado, o conjunto de Espirais concêntricas possuem muitos objetos sobrepostos, principalmente próximo ao centro, que propositalmente ocasionarão erros. Os outros conjuntos, além de possuírem dados sobrepostos, podem possuir também diferentes quantidades de classes com diferentes instâncias, como o caso do *Yeast* que é considerado o mais difícil de classificar.

Os algoritmos empregados para a classificação foram o *K-means* e a RNA do tipo MLP, onde o número de interações para cada conjunto de dados foi de 50, 100, 500, 1000 e 5000 iterações. Na MLP foi analisado também o emprego de diferentes funções de aprendizagem, onde foram empregadas as funções *Standart Back-Propagation*, *Backprop Momentum*, *Backprop Weight Decay* e *Quickprop*. Dentre as funções possíveis, estas foram escolhidas por permitirem uma comparação mais direta entre elas e entre o número de iterações

Para o SVM, através da função Kernel RBF ou radial, é possível resolver problemas não linearmente separáveis, através do mapeamento para um espaço de maior dimensão. Existem dois parâmetros que podem ser variados em busca de um melhor resultado para o aprendizado do classificador, são eles: γ (gamma) e C (custo) (OLIVEIRA JUNIOR, 2010). Assim, foi realizada a alteração do termo de regularização na formulação de Lagrange, representado no algoritmo computacional como sendo a variável *Cost*, bem como também foi alterada a variável *gamma* as quais foram realizadas mediante a uma sequência de estimativas empíricas até encontrar os melhores valores que gerassem maior eficiência.

Foi empregado o método de *cross validation*, de modo que 80% dos dados são usados para treino e 20% para teste, até que todos os objetos do conjunto de dados tenham sido testados e treinados. Todas as ferramentas desenvolvidas foram implementadas no programa estatístico, *R-Program* (R Core Team, 2015).

3. REFERENCIAL TEÓRICO

K-means

K-means é um método de classificação tradicional que particiona um conjunto de n dimensões em uma quantidade k de aglomerados. Este método leva em consideração apenas a distância euclidiana simples para classificar os dados (MACQUEEN, 1967; COVER, HART, 1967; ZHANG, FANG; 2013). Sendo k um valor definido pelo operador, este valor será ideal quando for igual a quantidade de classes presente no conjunto de dados, e este se tornará a quantidade de centróides gerados aleatoriamente no conjunto de dados e que irão gerar o aglomerado a partir dos objetos mais próximos. Os centróides podem ser ajustados a partir de um número de iterações que possam satisfazer a um critério de parada.

Redes Neurais Artificiais

As Redes Neurais Artificiais (RNA) são um sistema de processamento que calculam funções matemáticas, normalmente não lineares. As RNAs são inspiradas no cérebro humano, onde o sistema vai simular um aprendizado baseado nos dados que processa. Existem uma ou mais camadas nesse processamento, que são interligadas por conexões, que constituem os nodos, e que geralmente estão associadas a pesos. Os pesos passam a ponderar a entrada recebida por cada neurônio e armazenam o “conhecimento” do modelo. A aprendizagem ocorre através de um ajuste nos pesos, para que se adequem a resposta esperada (HAYKIN, 2001). O *perceptron* multicamadas, ou *multilayer perceptron* (MLP), que são RNAs que apresentam pelo menos uma camada intermediária ou escondida, são mais adequados para resolver problemas não linearmente separáveis. Em uma rede multicamadas, o processamento realizado por cada nodo é definido pela combinação dos processamentos realizados pelos nodos da camada anterior que estão conectados a ele. O algoritmo de aprendizado mais conhecido para treinamento destas redes é o algoritmo *back-propagation*. O algoritmo *back-propagation* é um algoritmo supervisionado que utiliza pares (entrada, saída desejada) para, por meio de um mecanismo de correção de erros, ajustar os pesos da rede. O treinamento ocorre em duas fases, em que cada fase percorre a rede em um sentido. Estas duas fases são chamadas de fase *forward* e fase *backward*. A fase *forward* é utilizada para definir a saída da rede para um dado padrão de entrada. A fase *backward* utiliza a saída desejada e a saída fornecida pela rede para atualizar os pesos de suas conexões (RUMELHART *et al*, 1986).

Support Vector Machines

Para Mitchell (1997) as Máquinas de Vetores de Suporte (SVMs, do Inglês *Support Vector Machines*) se baseiam no aprendizado de máquinas e estão recebendo cada vez mais notoriedade por parte da comunidade científica. Moreira et al (2014) afirma que este modelo de aprendizagem por análise estatística possibilita a diminuição de erros empíricos durante o processo de classificação, além de que também pode separar as classes com uma maior eficiência através da maximização a margem de separação entre elas. Lorena et al. (2007) discorre que é de suma importância o estudo dos parâmetros do algoritmo SVM para o desenvolvimento de suas respectivas soluções para os problemas impostos à ciência. Uma vez que a escolha de um determinado parâmetro pode ser interessante para uma ocasião, e não ser interessante a outra.

4. RESULTADOS E DISCUSSÃO

Os resultados obtidos experimentalmente demonstraram que nem sempre o resultado esperado ocorreu. No caso da quantidade maior de iterações, nem sempre apresentou uma melhora e, em alguns casos, ocorreu exatamente o contrário. A Tabela 4 ilustra os melhores resultados obtidos para os métodos empregados. Na Tabela 2, é apresentado apenas o melhor resultado obtido para cada função de aprendizagem empregada, podendo demonstrar, em alguns casos, que o crescimento no número de iterações pode não ajudar muito na taxa média de acertos.

Tabela 2 - Parâmetros da MLP e taxa de acertos – Média (Desvio padrão)

| Dados | Função de aprendizagem | Iterações | Média (Desv.Padrão) |
|-----------------------|-------------------------|-----------|---------------------|
| Círculos Concêntricos | <i>Back Propagation</i> | 500 | 1,00 (0,00) |

| | | | |
|-----------------------|------------------------------|------|-------------|
| | <i>Backprop Momentum</i> | 500 | 1,00 (0,00) |
| | <i>Backprop Weight Decay</i> | 500 | 1,00 (0,00) |
| | <i>Quickprop</i> | 5000 | 0,80 (0,27) |
| Espirais Concêntricas | <i>Back Propagation</i> | 5000 | 0,66 (0,03) |
| | <i>Backprop Momentum</i> | 5000 | 0,60 (0,06) |
| | <i>Backprop Weight Decay</i> | 5000 | 0,59 (0,07) |
| | <i>Quickprop</i> | 5000 | 0,53 (0,05) |
| Diabetes | <i>Back Propagation</i> | 50 | 0,77 (0,03) |
| | <i>Backprop Momentum</i> | 50 | 0,77 (0,03) |
| | <i>Backprop Weight Decay</i> | 100 | 0,77 (0,03) |
| | <i>Quickprop</i> | 5000 | 0,69 (0,08) |
| Câncer de mama | <i>Back Propagation</i> | 50 | 0,97 (0,03) |
| | <i>Backprop Momentum</i> | 50 | 0,97 (0,03) |
| | <i>Backprop Weight Decay</i> | 50 | 0,97 (0,03) |
| | <i>Quickprop</i> | 50 | 0,97 (0,02) |
| Coluna Vertebral | <i>Back Propagation</i> | 500 | 0,85 (0,08) |
| | <i>Backprop Momentum</i> | 500 | 0,85 (0,09) |
| | <i>Backprop Weight Decay</i> | 5000 | 0,86 (0,08) |
| | <i>Quickprop</i> | 1000 | 0,84 (0,06) |
| Sementes | <i>Back Propagation</i> | 1000 | 0,96 (0,03) |
| | <i>Backprop Momentum</i> | 1000 | 0,96 (0,03) |
| | <i>Backprop Weight Decay</i> | 5000 | 0,97 (0,03) |
| | <i>Quickprop</i> | 500 | 0,97 (0,03) |
| Vinho | <i>Back Propagation</i> | 100 | 0,99 (0,02) |
| | <i>Backprop Momentum</i> | 50 | 0,99 (0,02) |
| | <i>Backprop Weight Decay</i> | 50 | 0,99 (0,02) |
| | <i>Quickprop</i> | 100 | 0,99 (0,02) |
| Yeast | <i>Back Propagation</i> | 1000 | 0,65 (0,06) |
| | <i>Backprop Momentum</i> | 500 | 0,63 (0,04) |
| | <i>Backprop Weight Decay</i> | 500 | 0,63 (0,05) |
| | <i>Quickprop</i> | 1000 | 0,41 (0,09) |

Fonte: Própria dos autores.

No caso da RNA do tipo MLP, podemos perceber tanto a influência do número de iterações quanto da função de aprendizagem (Tabela 2), embora que em alguns casos não apresentem tanta influência assim. Os melhores resultados ocorreram nos conjuntos de Círculo Concêntrico, Vinho, Sementes e Câncer, com a média de acerto entre 97% a 100%. Devido a falta de objetos sobrepostos no conjunto do Círculos Concêntricos, foi possível atingir uma média de 100% de acerto e um desvio padrão 0%, em 500 iterações na maioria das funções de aprendizagem.

A Tabela 3 descreve os parâmetros que apresentaram os melhores resultados no uso do algoritmo do SVM. Com a pretensão de observar o comportamento do número de vetores suporte e a taxa de acerto, os cálculos foram realizados uniformizando e não uniformizando os valores de custo, nem da variável *gamma*.

Tabela 3 - Parâmetros do SVM – Média (desvio padrão) do número de vetores de suporte, custo e *gamma*

| Dados | Vetores de suporte | Custo | <i>gamma</i> |
|-----------------------|--------------------|-------|--------------|
| Círculos Concêntricos | 108,80 (2,14) | 0,1 | 0,5 |

| | | | |
|-----------------------|----------------|-----------------------|-------------|
| Espirais Concêntricas | 201,80 (24,12) | 1000, 10000 ou 100000 | 2 |
| Diabetes | 450,40 (5,31) | 1 | 0,5 |
| Câncer de mama | 208,20 (18,43) | 1 ou 10 | 0,5 ou 1 |
| Coluna Vertebral | 168,80 (6,21) | 1 ou 10 | 0,5 |
| Sementes | 99,00 (24,63) | 1 ou 10 | 0,5, 1 ou 2 |
| Vinho | 135,40 (1,85) | 10 | 0,5 |
| <i>Yeast</i> | 961 (42,85) | 1 | 0,5 |

Fonte: Própria dos autores.

Para o *K-means*, o número de iterações não influenciou nos resultados de forma que pudesse ter sido notada. Para a maioria dos conjuntos de dados, 50 iterações já foi o suficiente para ter o mesmo resultado de 5000 iterações e com o desvio padrão praticamente nulo. No entanto, para o conjunto de dados *Yeast*, foi notado um decréscimo na média de acerto assim como um decréscimo no desvio padrão, demonstrando uma leve influência do número de iterações. Os resultados obtidos pelos métodos *K-means* e SVM estão dispostos na Tabela 4, juntamente com melhores resultados obtidos pelo MLP.

Tabela 4 – Taxa de acertos em média (desvio padrão) por classes e em geral, para os métodos *Kmeans*, RNA e SVM

| Dados | Classes | <i>K-means</i> | RNA | SVM |
|-----------------------|------------------|----------------|-------------------|-------------------|
| Círculos Concêntricos | Círculo interno | 0,82(0,00) | 1,00(0,00) | 1,00(0,00) |
| | Círculo externo | 0,45(0,00) | 1,00(0,00) | 1,00(0,00) |
| | Geral | 0,64(0,00) | 1,00(0,00) | 1,00(0,00) |
| Espirais Concêntricas | Espiral interna | 0,40(0,00) | 0,65(0,43) | 0,99(0,02) |
| | Espiral externa | 0,65(0,00) | 0,54(0,32) | 1,00(0,01) |
| | Geral | 0,53(0,00) | 0,60(0,38) | 1,00(0,02) |
| Diabetes | Positivo | 0,50(0,00) | 0,59(0,07) | 0,49(0,04) |
| | Negativo | 0,76(0,00) | 0,86(0,05) | 0,86(0,03) |
| | Geral | 0,63(0,00) | 0,73(0,06) | 0,68(0,04) |
| Câncer de mama | Benigno | 0,98(0,00) | 0,96(0,03) | 0,95(0,03) |
| | Maligno | 0,92(0,00) | 0,99(0,02) | 0,98(0,03) |
| | Geral | 0,95(0,00) | 0,98(0,03) | 0,97(0,03) |
| Coluna Vertebral | Hernia | 0,02(0,00) | 0,78(0,26) | 0,72(0,11) |
| | Espondilolistese | 0,51(0,00) | 0,95(0,07) | 0,95(0,06) |
| | Normal | 0,83(0,00) | 0,78(0,09) | 0,75(0,10) |
| | Geral | 0,45(0,00) | 0,84(0,14) | 0,81(0,09) |
| Sementes | Kama | 0,83(0,00) | 0,96(0,06) | 0,89(0,07) |
| | Rosa | 0,89(0,00) | 0,96(0,04) | 0,96(0,03) |
| | Canadian | 0,96(0,00) | 0,99(0,03) | 0,96(0,06) |
| | Geral | 0,89(0,00) | 0,97(0,04) | 0,94(0,05) |
| Vinho | Cultivar 1 | 1,00(0,00) | 1,00(0,00) | 0,88(0,04) |
| | Cultivar 2 | 0,89(0,00) | 0,97(0,04) | 1,00(0,00) |
| | Cultivar 3 | 1,00(0,00) | 1,00(0,00) | 0,79(0,11) |
| | Geral | 0,96(0,00) | 0,99(0,01) | 0,89(0,05) |
| <i>Yeast</i> | Citoesqueleto | 0,68(0,12) | 0,63(0,03) | 0,62(0,04) |
| | Nuclear | 0,19(0,26) | 0,55(0,08) | 0,55(0,05) |
| | Mitochondrial | 0,59(0,01) | 0,56(0,08) | 0,49(0,05) |
| | Outros | 0,43(0,03) | 0,80(0,12) | 0,80(0,10) |
| | Geral | 0,47(0,11) | 0,64(0,08) | 0,62(0,06) |

Fonte: Própria dos autores.

Os melhores resultados ocorreram nos dados de Câncer, Sementes e Vinho onde a média de acerto ficou entre 89% a 96%, para todos os métodos de classificação. Demonstrando que, apesar de ser um método mais simples, o *K-means* pode apresentar bons resultados a depender do conjunto de dados trabalhados. E quanto aos erros, podem ser explicados devido a sobreposição de dados ou então pela confusão entre as diferentes quantidades de objetos dentro do conjunto, já que é um método que vai agrupar os vizinhos mais próximos.

O SVM ganha um destaque importante, onde seus resultados se mostraram substancialmente melhores que os do *K-means*, e muito próximos dos da MLP. Em especial na classificação do conjunto de dados de Espirais concêntricas, gerados artificialmente, onde conseguiu atingir uma taxa de 100% de acerto, superando os resultados da MLP. Os dados de Espirais se tornam mais complexos para a classificação, uma vez que seus objetos mais próximos do centro tendem a apresentar uma certa sobreposição. Entretanto, para conjuntos de dados como o Vinho, o SVM foi superado pelo *K-means*.

Em um comparativo com os resultados obtidos utilizando o *K-means*, as RNAs do tipo MLP se sobrepõem ao atingir 97% em comparação com 89% para os dados de Sementes e 99% em comparação com 96% para os de Vinho. Além disso, para os piores resultados da MLP houve um melhor desempenho em comparação com os resultados do *K-means*: 66% contra 52% nos dados de Espirais Concêntricas; 77% contra 67% nos de Diabetes e 65% contra 46% nos de *Yeast*. No entanto, para os dados de Coluna Vertebral os dois métodos apresentaram um resultado semelhante, alcançando 97% e 96% de média de acertos para a MLP e o *K-means* respectivamente.

5. CONCLUSÕES

Se fossemos avaliar os métodos utilizados quanto a sua complexidade e sofisticação, a RNA do tipo MLP seria o método mais sofisticado e complexo, o *K-means* o mais simples e o SVM o intermediário entre os dois. Mediante aos resultados expostos, o desenvolvimento das arquiteturas algorítmicas geradas no programa R utilizando uma aprendizagem supervisionada, se mostraram eficientes para a classificação da maioria dos conjuntos de dados. Vale ainda ressaltar que a eficiência atingida para estes conjuntos de dados não garante o mesmo percentual para outros, uma vez que se trata de um campo de conhecimento com muitos detalhes e extrema variabilidade de ocasiões. Evidentemente as RNAs tiveram um melhor desempenho na classificação da maioria dos conjuntos de dados analisados. No entanto, mesmo com uma metodologia mais avançada, por simular o cérebro humano, as redes neurais podem não superar tanto os métodos mais tradicionais e simples como o *K-means* em alguns poucos casos. Em casos onde há uma semelhança de resultados, é preferível o método mais simples por apresentar um custo e uma complexidade menores.

6. AGRADECIMENTOS

Agradecemos a Universidade Federal do Cariri (UFCA), que atuou como agência de fomento à pesquisa.

REFERÊNCIAS

ALAMSYAH, M. et al. The Classification of Diabetes Mellitus Using Kernel k-means. **Journal of Physics: Conference Series**, v. 947, n. 1, 2018.

COVER, T.; HART, P. Nearest neighbor pattern classification. *IEEE Transactions on*

Information Theory, v. 13, n. 1, p. 21–27, 1967. ISSN 0018-9448

DAVYDOV, Vasily Vasilyevich. Problemas do ensino desenvolvimental: a experiência da pesquisa teórica e experimental na psicologia. Tradução de José Carlos Libâneo e Raquel AM da Madeira Freitas, 1988.

DUDA, R. O.; HART, P. E.; STORK, D. G. Pattern Classification. Second. New York: John Wiley & Sons, 2000.

ELIZONDO, D. The linear separability problem: Some testing methods. IEEE Transactions on Neural Networks, 2006

FRANK, A.; ASUNCION, A. UCI Machine Learning Repository. abril 2013. Irvine, CA: University of California, School of Information and Computer Science. Disponível em: <<http://archive.ics.uci.edu/ml>>

HAYKIN, S. Redes Neurais Princípios e Prática. Second. Porto Alegre: Bookman, 2001.

LI, M. The bi-criteria seeding algorithms for two variants of k-means problem. **Journal of Combinatorial Optimization**, 2020.

MACQUEEN, J. Some methods for classification and analysis of multivariate observations. In: 5th Berkeley Symposium on Mathematical Statistics and Probability. Berkeley: University of California Press, 1967. p. 281–297.

MOREIRA, A. A., Nery, C. V. M., & Fernandes, F. H. S. (2014). APLICAÇÃO DO ALGORITMO SUPPORT VECTOR MACHINE NA ANÁLISE ESPAÇO-TEMPORAL DO USO E OCUPAÇÃO DO SOLO NA BACIA DO RIO VIEIRA. *Caminhos de Geografia*, 15(50).

OLIVEIRA JUNIOR, G. M. Máquinas de Vetores Suporte: Estudo e Análise de Parâmetros para Otimização de Resultado. 2010. 41f. Dissertação (Graduação) – Centro de Informática, Universidade Federal de Pernambuco, Pernambuco, 2010.

QI, X.; SILVESTROV, S.; NAZIR, T. Data classification with support vector machine and generalized support vector machine. **AIP Conference Proceedings**, v. 1798, 2017.

R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria, 2015. Disponível em: <<https://www.R-project.org>>.

RITTER, Matias do Nascimento; THEY, Ng Haig; KONZEN, Enéas Ricardo. Introdução ao software estatístico R. 2019.v

RUMELHART, D. E.; HINTON, G.; WILLIAMS, R. Learning representations by back-propagating errors. *Nature*, v. 323, n. 11, p. 533–536, 1986.

T. Mitchell. Machine Learning. McGraw Hill, 1997.

TEH, K. et al. Imbalanced learning: Improving classification of diabetic neuropathy from magnetic resonance imaging. **PLoS ONE**, v. 15, n. 12 December, p. 1–15, 2020.

UMAMAHESWARI, R.; PRINCESS, M. R.; KUMAR, P. N. Improved Whale Optimized MLP Neural Network-Based Learning Mechanism for Multiuser Detection in MIMO Communication System. **Journal of Circuits, Systems and Computers**, v. 29, n. 15, 2020.

VAPNIK, V. N. Statistical Learning Theory. New York: John Wiley & Sons, 1998

WU, A. et al. Classification of corn kernels grades using image analysis and support vector machine. **Advances in Mechanical Engineering**, v. 10, n. 12, p. 1–9, 2018.

ZHANG, C.; FANG, Z. An improved k-means clustering algorithm. *Journal of Information & Computational Science*, v. 10, p. 193–199, 2013