UNIVERSIDADE ESTADUAL DE PONTA GROSSA
SETOR DE CIÊNCIAS AGRÁRIAS E DE TECNOLOGIA
DEPARTAMENTO DE INFORMÁTICA

ALCEU DE SOUZA BRITTO DOS SANTOS
MATEUS FELIPE DA SILVA JUNGES
RAFAEL DE MATTOS
VICTOR PIOTROVSKI BEGHA

TRABALHO FINAL DE INTELIGÊNCIA COMPUTACIONAL

PONTA GROSSA
2019

ALCEU DE SOUZA BRITTO DOS SANTOS
MATEUS FELIPE DA SILVA JUNGES
RAFAEL DE MATTOS
VICTOR PIOTROVSKI BEGHA

TRABALHO FINAL DE INTELIGÊNCIA COMPUTACIONAL

Trabalho apresentado para a disciplina de Inteligência Computacional, do curso de Engenharia de Computação da Universidade Estadual de Ponta Grossa.
Professor: Prof. Dr. José Carlos Ferreira da Rocha

PONTA GROSSA
2019

SUMÁRIO

1 INTRODUÇÃO

Dentro do estudo da aprendizagem de máquina, o objetivo principal é realizar o treinamento de algoritmos para detectar e aprender padrões, além de auxiliar na tomada de decisão dentro de determinadas situações. Tudo isso é feito através da análise sobre conjuntos extensos de dados, ao invés de uma programação direta para execução de funções - isso permite que os algoritmos se tornem mais eficientes e acurados com base nas informações já conhecidas.

Um classificador pode ser definido como a forma na qual o algoritmo faz a aprendizagem dos dados fornecidos. São muitos os classificadores disponíveis para uso, cada um deles tem um melhor desempenho em certo tipo de dados, ou para certa funcionalidade do código.

Neste trabalho, são apresentados e comparados os resultados referentes à comparação de três classificadores que atuaram sobre três bases de dados diferentes, os classificadores são o *Naive Bayes,* o *Perceptron* (em diferentes exemplos de configuração possíveis, o J48 (ID3) e o SVM, os quais atuaram sobre as bases *Congressional Voting Records, Iris* e *Breast Cancer Wisconsin Diagnostic.*

2 CLASSIFICADORES UTILIZADOS

2.1 NAIVE BAYES

O algoritmo de Naive Bayes é um classificador probabilístico baseado no Teorema de Bayes, inicialmente criado pelo matemático Thomas Bayes no século XVIII e com sua formulação moderna elaborada por Pierre-Simon Laplace em 1812.

Por ser um algoritmo simples e rápido, possui um bom desempenho como classificador, além do fato de precisar de um conjunto de dados pequeno para concluir classificações a um excelente nível de precisão.

A principal característica deste algoritmo é desconsiderar completamente a correlação entre as variáveis, ou seja, se uma fruta é considerada uma laranja se ela possuir cor amarela, for redonda e contar com aproximadamente 10 centímetros de diâmetro, o algoritmo de Naive Bayes não vai considerar a correlação entre esses dados, e tratará cada um de forma independente.

Frequentemente este algoritmo é aplicado em processamento de diagnósticos médicos, para diagnósticos de doenças. Basicamente, consiste em calcular uma probabilidade *a posteriori*, ou seja, o paciente possuir a doença, dado que recebeu um resultado positivo, multiplicando a probabilidade *a priori*, que é possuir a doença, pela probabilidade de receber um resultado positivo, dado que tem a doença. Com esses dados, é calculado a probabilidade a posteriori da negação, ou seja, não possuir a doença dado que o paciente recebeu um resultado positivo. Para concluir, os dados são normalizados, dividindo-se o resultado pela soma das probabilidades.

2.2 PERCEPTRON

O perceptron multicamadas é um tipo de rede neural artificial profunda. Alguns trabalhos recentes com este classificador mostraram que ele é capaz de aproximar um operador XOR, assim como várias outras funções lineares.

Um perceptron multicamadas é composto por mais de um perceptron. Estes, são compostos por uma camada de entrada para receber o sinal, uma camada de saída que toma uma decisão ou faz uma predição sobre uma dada entrada e, entre essas duas, um número arbitrário de camadas ocultas. Esses algoritmos são normalmente aplicados a problemas de machine learning supervisionados, o que significa dizer que treinam sob um conjunto de pares de entrada e saída e aprendem a modelar a correlação entre os pares.

Podemos fazer uma analogia do perceptron multi camadas com o jogo de ping pong. São basicamente dois movimentos, sendo um para frente e um para trás. No passo para frente, o sinal se move da camada de entrada através das camadas ocultas até a camada de saída, e a decisão na camada de saída é medida contra os rótulos de verdade do solo. No passo para trás, usa-se *backpropagation* e a regra da cadeia de cálculo, derivadas parciais da função de erro respeitando-se os vários pesos e assim, propagados de volta através do perceptron multi camadas. A rede continua assim até que o erro diminua. Isso é chamado de convergência.

2.3 J48 (ID3)

O algoritmo ID3 foi um dos primeiros algoritmos de árvore de decisão, baseado em sistemas de inferência e em conceitos de sistemas de aprendizagem. Ele funciona basicamente construindo árvores de decisão a partir de um dado conjunto de entrada, sendo que a árvore resultante é usada para classificar as amostras futuras.

Este algoritmo separa um conjunto de treinamento em vários subconjuntos, de forma que estes contenham exemplos de uma única classe. Essa divisão ocorre baseado em um único atributo, que, por sua vez, é selecionado com base em uma propriedade estatística chamada de ganho de informação. O ganho de informação consiste em uma medida estatística usada para construção de árvores de decisão a fim de escolher o melhor atributo de teste entre todos os envolvidos no nó em questão. O atributo com maior ganho de informação é o que melhor classifica o conjunto de amostras de treinamento (MOTTA, 2004). Assim, a profundidade final da árvore de decisão é minimizado.

Após a formação dessa árvore de decisão, ela é avaliada, utilizando dados que não tenham sido usados no treinamento, permitindo que possamos estimar como a árvore generaliza os dados e se adapta a novas situações, além de estimar a proporção de erros e acertos ocorridos na construção da árvore.

2.4 SVM COM KERNEL POLINOMIAL

O SVM com Kernel Polinomial representa a similaridade de vetores em um espaço de características sobre polinômios das variáveis originais, permitindo a aprendizagem de modelos não lineares.

Muito popular no Processamento de Linguagem Natural, o Kernel Polinomial não utiliza somente as características fornecidas pelos dados que foram usados como entrada, mas também leva em conta as combinações dos dados.

O espaço de características de um núcleo polinomial é o mesmo de uma regressão polinomial, porém não utiliza a ampliação combinatória no número dos parâmetros que devem ser aprendidos.

3 BASES DE DADOS ANALISADAS

3.1 CONGRESSIONAL VOTING RECORDS

Este conjunto de dados se refere aos resultados de votações no Congresso dos Estados Unidos no ano de 1984. Os atributos (colunas) são dispostos de forma que a primeira coluna é o partido (nome da classe), que pode receber os valores ‘democrat’ ou ‘republican’, e cada uma das demais colunas é o resultado de uma votação, com valores possíveis ‘y’ ou ‘n’; cada linha representa, então, um congressista, com seu partido e o seu voto em cada uma das pautas (‘y’ para sim, ‘n’ para não e ‘?’ para desconhecido ou inexistente), com 435 instâncias. A classificação foi baseada na classe principal que é o partido.

3.2 IRIS

Base que se refere às características de três espécies de plantas: *Iris Setosa*, *Iris Versicolour* e *Iris Virginica*, sendo que para cada uma dessas classes possui 50 instâncias para um total de 150. Os atributos são as características da flor da planta, no caso, comprimento e largura das pétalas e sépalas, em centímetros. A implementação com os algoritmos testados buscou classificar corretamente a classe da espécie.

3.3 BREAST CANCER WISCONSIN DIAGNOSTIC

Essa base foi construída a partir de imagens de células para verificar a presença de câncer de mama em 569 instâncias. Os seus atributos incluem uma série de características dos núcleos células nas imagens - raio, textura, perímetro, área, etc. A classificação busca verificar corretamente o diagnóstico das células (‘M’ = maligno, ‘B’ = benigno).

4 RESULTADOS E DISCUSSÃO

4.1 ACURÁCIA E PRECISÃO DOS CLASSIFICADORES

Para avaliar o desempenho dos classificadores, foram verificadas para cada classificador respectivamente em cada base de dados, a porcentagem de acurácia (isto é, valores médios próximos do valor esperado) e a porcentagem de precisão (valores com pouca dispersão).

Cada classificador foi testado para cada uma das bases de dados. Em todos os casos, o teste foi feito através do software Weka (“Waikato Environment for Knowledge Analysis”) versão 3, desenvolvido na Universidade de Waikato, Nova Zelândia. O arquivo da base de dados era lido em formato .csv, em seguida escolhendo a classe que será analisada e então realizando a classificação com os parâmetros escolhidos.

No caso do Perceptron multicamadas, foram testadas oito configurações, alterando número de camadas intermediárias, número de épocas e taxas de aprendizagem da seguinte forma:

Configuração	Nº Camadas	Nº Épocas	Taxa aprendizagem
Perceptron(3, 500, 0.3) - P1	3	500	0.3
Perceptron(3, 500, 0.5) - P2	3	500	0.5
Perceptron(3, 750, 0.3) - P3	3	750	0.3
Perceptron(3, 750, 0.5) - P4	3	750	0.5
Perceptron(4, 500, 0.3) - P5	4	500	0.3
Perceptron(4, 500, 0.5) - P6	4	500	0.5
Perceptron(4, 750, 0.3) - P7	4	750	0.3
Perceptron(4, 750, 0.5) - P8	4	750	0.5

Quadro 1 - Configurações do Perceptron

4.1.1 Base Congressional Voting Records

Usando a base “Congressional Voting Records”, verificou-se a capacidade dos classificadores em obter corretamente o partido dos congressistas. Cada classificador obteve acurácia e precisão conforme os seguintes dados:

Congressional Voting Records		
Classificador	**Acurácia**	**Precisão**
Naive Bayes	0,901	0,918
Perceptron(3, 500, 0.3) - P1	0,952	0,948
Perceptron(3, 500, 0.5) - P2	0,947	0,940
Perceptron(3, 750, 0.3) - P3	0,956	0,953
Perceptron(3, 750, 0.5) - P4	0,947	0,940
Perceptron(4, 500, 0.3) - P5	0,947	0,939
Perceptron(4, 500, 0.5) - P6	0,936	0,929
Perceptron(4, 750, 0.3) - P7	0,947	0,939
Perceptron(4, 750, 0.5) - P8	0,940	0,934
J48 (ID3)	0,961	0,960
SVM	0,961	0,958

Quadro 2 - Resultados “Congressional Voting Records”

4.1.2 Base Iris

Usando a base “Iris”, verificou-se a capacidade dos classificadores em escolher corretamente a espécie da planta. Os resultados são mostrados a seguir:

Iris		
Classificador	**Acurácia**	**Precisão**
Naive Bayes	0,953	0,976
Perceptron(3, 500, 0.3) - P1	0,973	0,987
Perceptron(3, 500, 0.5) - P2	0,960	0,980
Perceptron(3, 750, 0.3) - P3	0,967	0,983
Perceptron(3, 750, 0.5) - P4	0,960	0,980
Perceptron(4, 500, 0.3) - P5	0,980	0,990
Perceptron(4, 500, 0.5) - P6	0,967	0,983
Perceptron(4, 750, 0.3) - P7	0,973	0,987
Perceptron(4, 750, 0.5) - P8	0,967	0,983
J48 (ID3)	0,960	0,980
SVM	0,968	0,983

Quadro 3 - Resultados “Iris”

4.1.3 Base Breast Cancer Wisconsin Diagnostic

Usando a base “Breast Cancer Wisconsin Diagnostic”, os capacitores deveriam calcular corretamente o valor da classe de diagnóstico (‘M’ ou ‘B’, maligno ou benigno). Foram obtidos os seguintes resultados:

Breast Cancer Wisconsin Diagnostic		
Classificador	**Acurácia**	**Precisão**
Naive Bayes	0,931	0,918
Perceptron(3, 500, 0.3) - P1	0,965	0,949
Perceptron(3, 500, 0.5) - P2	0,963	0,955
Perceptron(3, 750, 0.3) - P3	0,963	0,952
Perceptron(3, 750, 0.5) - P4	0,970	0,965
Perceptron(4, 500, 0.3) - P5	0,968	0,957
Perceptron(4, 500, 0.5) - P6	0,965	0,956
Perceptron(4, 750, 0.3) - P7	0,968	0,960
Perceptron(4, 750, 0.5) - P8	0,967	0,957
J48 (ID3)	0,940	0,933
SVM	0,977	0,965

Quadro 4 - Resultados “Breast Cancer Wisconsin Diagnostic”

4.2 ANÁLISE ESTATÍSTICA

Para realização dos testes, foi usado a linguagem R, utilizando a IDE RStudio e funções da biblioteca PMCMR (“Pairwise Multiple Comparisons of Mean Rank Sums”).

Com base nas acurácias e precisões obtidas de cada algoritmo para cada dataset podemos fazer a análise sobre eles usando o Teste de Friedman. O teste de Friedman é um teste estatístico não-paramétrico desenvolvido por Milton Friedman. Semelhante ao ANOVA, é utilizado para detectar diferenças nos tratamentos em várias experimentos de teste. O procedimento envolve a classificação de cada linha (ou *bloco*), então considerando os valores dos postos de colunas. O teste de

Friedman é usado para medidas repetidas de análise unidirecional de variância dos postos.

Os dados tabelados de acurácia e precisão estão listados abaixo.

Acurácia												
	Naive Bayes	P1	P2	P3	P4	P5	P6	P7	P8	P9	J48	SVM
base 1	0,901	0,952	0,947	0,956	0,947	0,947	0,936	0,947	0,940	0,961	0,961	0,901
base 2	0,953	0,973	0,960	0,967	0,960	0,980	0,967	0,973	0,967	0,960	0,968	0,953
base 3	0,931	0,965	0,963	0,963	0,970	0,968	0,965	0,968	0,967	0,940	0,977	0,977

Quadro 5 - Resultados de acurácia de cada classificador

Precisão												
	Naive Bayes	P1	P2	P3	P4	P5	P6	P7	P8	P9	J48	SVM
base1	0,918	0,948	0,940	0,953	0,940	0,939	0,929	0,939	0,934	0,960	0,958	0,958
base2	0,976	0,987	0,980	0,983	0,980	0,990	0,983	0,987	0,983	0,980	0,983	0,983
base3	0,918	0,949	0,955	0,952	0,965	0,957	0,956	0,960	0,957	0,933	0,965	0,965

Quadro 6 - Resultados de precisão de cada classificador

Considera-se que a análise será feita utilizando um nível de significância de 5%. Ao realizar o teste de Friedman para blocos não replicados e tratamentos em ambos os dados analisados (acurácia e precisão), e em seguida aplicando o teste de Post-hoc para classificar hierarquicamente essa diferença dos grupos (pois apenas o teste de Friedman revela apenas se existe pelo menos um grupo que difere dos demais - o teste Friedman com Post-hoc é necessário, portanto, para saber exatamente quais grupos diferem estatisticamente). No caso, cada par de classificadores terá um *p-value* que indica qual hipótese deve ser considerada. Se o p-value do par for maior que o nível de significância (5%), considera-se que a hipótese nula é verdadeira, ou seja: esses dois classificadores não diferem

estatisticamente entre si (estarão no mesmo agrupamento na tabela de classificação, que será mostrada a seguir). Se o p-value da comparação for menor que o nível de significância, eles diferem estatisticamente e estarão agrupados separadamente.

Esse processo é realizado para todos os pares de classificadores. Assim, aplicando o teste de Friedman com Post-hoc para os blocos (conjuntos de dados para cada classificador) de acurácia usando um nível de significância de 5% podemos afirmar que temos a seguinte diferença entre os classificadores:

Classificação por acurácia	
Classificador	**Agrupamento**
SVM	a
Perceptron(4, 500, 0.3) - P5	ab
Perceptron(3, 500, 0.3) - P1	ab
Perceptron(4, 750, 0.3) - P7	ab
Perceptron(3, 750, 0.3) - P3	ab
Perceptron(3, 750, 0.5) - P4	ab
Perceptron(4, 750, 0.5) - P8	ab
Perceptron(3, 500, 0.5) - P2	ab
Perceptron(4, 500, 0.5) - P6	ab
J48	ab
Naive Bayes	b

Quadro 7 - Divisão dos classificadores por desempenho em acurácia

Os classificadores estão ordenados com base na acurácia média, da maior acurácia (SVM) até a menor (Naive Bayes).

O significado de cada grupo (ex. ‘a’) é que classificadores no mesmo grupo não diferem estatisticamente entre si: no caso, classificadores marcados com ‘a’ não

diferem entre si dentro deste grupo, e classificadores marcados com ‘b’ não diferem entre si dentro deste outro grupo. Nota-se que conjuntos de dados que não diferem estatisticamente não são exatamente iguais: isso apenas significa que, em uma população, análises estatísticas são aproximadamente as mesmas. Portanto, a partir desses agrupamentos, demonstra-se que:

- O algoritmo SVM é significativamente melhor que o Naive Bayes, e tem desempenho semelhante aos classificadores do tipo Perceptron e J48 (ID3).
- O algoritmo Naive Bayes não é significativamente pior que os classificadores Perceptron e J48 (ID3), porém se comparado ao SVM, tem um desempenho estatisticamente mais baixo.

Quanto à precisão, se aplicado o teste de Friedman, determina-se que a precisão de cada classificador não difere estatisticamente dos demais. Ou seja, diferente da acurácia, todos os classificadores apresentam precisão semelhante, sem diferenças estatísticas significativas.

5 CONCLUSÃO

Com base nos testes realizados, podem ser feitas algumas conclusões. Primeiramente, nota-se que todos os classificadores sempre obtiveram acurácia e precisão superior a, pelo menos, 90%, ou seja, todos possuem um bom desempenho ao classificar as bases de dados testadas.

Através do conjunto de dados e da análise estatística realizada sobre ele, demonstra-se que o algoritmo SVM obteve a maior acurácia média, porém sem diferir estatisticamente do Perceptron (em qualquer uma de suas configurações) ou do J48 (ID3): isso significa que, dependendo da aplicação, esses três tipos de classificadores conseguirão alcançar um alto desempenho, com o Naive Bayes conseguindo um desempenho marginalmente pior. Notou-se também que todos os classificadores tiveram uma precisão média estatisticamente semelhante, ou seja, sem dispersão significativa nos dados.

REFERÊNCIAS

Congressional Voting Records Data Set. UCI Machine Learning Repository. Disponível em <https://archive.ics.uci.edu/ml/datasets/congressional+voting+records>. Acesso em 28 de julho de 2019.

Iris Data Set. UCI Machine Learning Repository. Disponível em <https://archive.ics.uci.edu/ml/datasets/Iris>. Acesso em 28 de julho de 2019.

Breast Cancer Wisconsin Diagnostic Data Set. UCI Machine Learning Repository. Disponível em <https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)>. Acesso em 28 de julho de 2019.

PARSANIA, V.; JANI, N.; BHALODIYA, N. **Applying Naive bayes, BayesNet, PART, JRip and OneR Algorithms on Hypothyroid Database for Comparative Analysis**. International Journal of Darshan Institute on Engineering Research and Emerging Technologies, Rajkot - Índia, v. 3, n. 1, jun. 2014.

AMARAL, F. C. N. **Data Mining**. São Paulo: Berkeley, 2001.

MANGIAFICO, S. S. **Friedman Test**. Summary and Analysis of Extension Program Evaluation in R. Disponível em <https://rcompanion.org/handbook/F_10.html>. Acesso em 30 de julho de 2019.