Relatório do Segundo Trabalho de Inteligência Artificial

Leandro Furlam Turi

Abstract

A classificação é um exemplo de reconhecimento de padrões. Neste trabalho onze algoritmos capazes de realizar esta tarefa foram aplicados e avaliados a quatro bases de dados bem conhecidas no mundo de *Machine Learning*. Os algoritmos são: *ZeroR*, *Aleatório*, *Aleatório Estratificado*, *OneR Probabilístico*, *Naive Bayes Gaussiano*, *KmeansCentroides*, *KGACentroides*, *Knn*, *DistKnn*, *Árvore de Decisão* e *Florestas de Árvores*. As bases de dados são *Iris*, *Digits*, *Wine* e *Breast Cancer*. Após experimentação, aferiu-se que apenas os métodos *Knn* e *DistKnn* apresentaram-se como melhores métodos.

Keywords: Classificadores, Aprendizagem Supervisionada, Aprendizagem Não Supervisionada.

1. Introdução

Em estatística, *classificação* é o problema de identificar a qual categoria uma determinada observação pertence, com base em um conjunto de dados contendo observações conhecidas. Exemplos são atribuir um e-mail como sendo *spam* ou *não-spam* [1], e atribuir um diagnóstico a um determinado paciente com base nas características observadas (sexo, pressão arterial, presença ou ausência de determinados sintomas, etc.) [1].

Na terminologia de *Machine Learning*, a classificação é considerada um exemplo de aprendizagem supervisionada, ou seja, aprendizagem onde um conjunto de treinamento de observações corretamente identificadas está disponível [1]. Ao caso em que procedimento ocorre de maneira não supervisionada, este é conhecido como *clustering*, e envolve agrupar dados em categorias com base em alguma medida de similaridade ou distância inerentes.

A classificação é um exemplo de reconhecimento de padrões. Na maior parte dos problemas, as observações individuais são analisadas em um conjunto de propriedades quantificáveis, conhecidas como *variáveis*. Essas propriedades podem ser categóricas, ordinais, valor de inteiro ou valor real [1]. Outros classificadores trabalham comparando observações com observações anteriores por meio de uma função de semelhança ou distância.

Um algoritmo que implementa classificação é conhecido como *classificador*. Esse termo às vezes também se refere à função matemática, implementada por um algoritmo de classificação, que mapeia dados de entrada para uma categoria.

Este trabalho consiste em realizar uma comparação experimental entre um conjunto pré-definido de técnicas de aprendizado e classificação automática aplicadas a alguns problemas de classificação.

Neste trabalho onze *classificadores* foram aplicados a quatro bases de dados bem conhecidas no mundo de *Machine Learning*. Os algoritmos são: *ZeroR*, *Aleatório*, *Aleatório Estratificado*, *OneR Probabilístico*, *Naive Bayes Gaussiano*, *KmeansCentroides*, *KGACentroides*, *Knn*, *DistKnn*, *Árvore de Decisão* e *Florestas de Árvores*. As bases de dados são *Iris*, *Digits*, *Wine* e *Breast Cancer*. Dado que cada classificador possui um conjunto específico de regras dinâmicas, a técnica de validação cruzada estratificada foi aplicada buscando tornar os resultados comparáveis.

2. Metodologia

Para cada base de dados, o procedimento experimental foi dividido em duas etapas: A primeira etapa consiste no treino e teste com 3 rodadas de validação cruzada estratificada de 10 *folds* dos classificadores que não possuem hiperparâmetros, isto é, os classificadores *ZeroR*, *Aleatório*, *Aleatório Estratificado*, *OneR Probabilístico*, *Naive Bayes Gaussiano*. A segunda etapa consiste no treino, validação e teste dos classificadores que precisam de ajuste de hiperparâmetros, isto é, os classificadores *KmeansCentroides*, *KGACentroides*, *Knn*, *DistKnn*, *Árvore de Decisão* e *Florestas de Árvores*. Neste caso o procedimento

de treinamento, validação e teste foi realizado através de 3 rodadas de ciclos aninhados de validação e teste, com o ciclo interno de validação contendo 4 *folds* e o externo de teste com 10 *folds*.

Todos os classificadores utilizaram implementam ou são implementados através da biblioteca de classificadores do *scikit-learn*[1], inclusive a própria validação cruzada estratificada, dada pela combinação das classes *cross_validate*[2] e *RepeatedStratifiedKFold*[3], com o parâmetro *random_state* setado em 36851234.

A busca em grade do ciclo interno considerou os seguintes valores de hiperparâmetros de cada técnica de aprendizado:

- KMeansCentroides: $n_clusters = 1, 3, 5, 7$
- KGACentroides: $n_clusters = 1, 3, 5, 7$
- Knn: $n\_neighbors = 1, 3, 5, 7$
- DistKnn: $n\_neighbors = 1, 3, 5, 7$
- Árvore de Decisão: $max\_depth = None, 3, 5, 10$
- Florestas de Árvores: $n\_estimators = 10, 20, 50, 100$

Os dados utilizados no conjunto de treino em cada rodada de teste foram padronizados via normalização *z-score*. Os valores de padronização obtidos nos dados de treino foram os mesmos utilizados para padronizar os dados do respectivo conjunto de teste. Para avaliar cada resultado, a métrica de acurácia foi aplicada, onde devido a padronização, a melhor performance é a unitária.

Todos os classificadores foram limitados a um tempo de execução máximo de 2 horas. Acerca do método *KGACentroides*, os hiperparâmetros genéticos utilizados foram aqueles que melhor foram avaliados no Trabalho 1, ou seja,

[1] `https://scikit-learn.org/stable/index.html`

[2] `https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.cross_validate.html`

[3] `https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.RepeatedStratifiedKFold.html`

tamanho populacional de 50 indivíduos, quantidade máxima de 100 iterações, taxas de cruzamento e mutação de 75% e 20%, além do tempo de execução limitado a 1s.

Para otimizar e simplificar todas as etapas aqui definidas, *Pipelines*[4] foram aplicadas para realizar as rodadas de validação cruzada.

2.1. Métodos Implementados

Dado que nem todos os classificadores estavam disponíveis na biblioteca do *scikit-learn*, estes foram implementados, a saber o classificador *OneR Probabilístico* e o método *KCentróides*, que serivirá como base para o *KMeansCentroides* e *KGACentroides*.

2.1.1. OneR Probabilístico

O método consiste no seguinte algoritmo: Seja o conjunto de preditores P. Tome um preditor $p \in P$. Para cada característica de p, calcule o erro total de uma regra r definida da seguinte maneira: (1) Conte quantas vezes cada valor de destino (classe) aparece. (2) Encontre a classe mais frequente. (3) Faça com que a regra atribua essa classe a esta característica de p. Repita este procedimento em todo P. Escolha o preditor com o menor erro total.

Um requisito para o algoritmo apresentado é que todas as características devem possuir um conjunto moderado de valores discretos, o que não é o caso das características existentes nas bases de dados utilizadas neste trabalho. Portanto, foi realizada uma discretização nas características das bases de dados. Para isto, o método de discretização utilizado foi o *KBinsDiscretizer*[5] da biblioteca *scikit-learn*, onde o parâmetro *strategy* utilizado foi definido como *kmeans*. Para simplificação, em vez de se definir um hiperparâmetro para este método, foi

[4] `https://scikit-learn.org/stable/modules/generated/sklearn.pipeline.Pipeline.html`

[5] `https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.KBinsDiscretizer.html`

adotado que o número de intervalos de cada característica será o dobro do número de classes da base.

Dado que o método é de natureza relativamente simples, este pode ser implementado de maneira quase *pythonica*[6], portanto seu tempo de processamento e quantidade de linhas de código são relativamente pequenos.

2.1.2. KCentróides

O método *KCentroides* utiliza um algoritmo de agrupamento para definir k grupos de exemplos de cada classe na base de treino. Assumindo que uma base de dados possui *ncl* classes, o algoritmo *KCentroides* forma inicialmente $k \times ncl$ grupos, sendo k grupos em cada uma das *ncl* classes. Em seguida, são calculados os centróides de cada um dos grupos e este centróide é associado a classe do grupo a partir do qual foi gerado. O método possui como hiperparâmetro o valor de k.

Para realizar uma classificação, o *KCentroides* verifica qual o centróide mais próximo do elemento a ser classificado e retorna a sua classe.

Para definir o classificador *KmeansCentroides*, o método *Kmeans* é passado para o *Kcentroides* em sua criação. De forma análoga, para o método *KGA-Centroides*, o método de agrupamento Algoritmo Genético, implementado no Trabalho 1, foi passado para o *Kcentroides* em sua criação.

3. Experimentos Realizados

Dado que todos os classificadores foram limitados a um tempo de execução máximo de 2 horas, os tempos de processamento de cada método para cada base de dados são apresentados na Tabela 1, onde destaca-se o alto tempo de processamento dos métodos implementados neste trabalho, sobretudo o *KGA-Centroids*. Note que este foi capaz de executar apenas para a base *Iris*. Isto

[6] `https://pt.stackoverflow.com/questions/192343/o-que-%C3%A9-c%C3%B3digo-pyth%C3%B4nico#:~:text=A%20express%C3%A3o%20pythonico%2C%20originada%20no,com%20algumas%20solu%C3%A7%C3%B5es%20extremamente%20simples.`

decorre principalmente da falta de otimização na maioria das etapas do algoritmo.

	Iris	*Digits*	*Wine*	*Breast Cancer*
ZeroR	0,031	0,047	0,031	0,031
Random	0,031	0,047	0,031	0,031
Stratified Random	0,016	0,063	0,016	0,031
OneR	1,078	155,859	3,047	9,969
Gaussian Naive Bayes	0,047	0,094	0,047	0,047
KmeansCentroids	85,219	743,250	89,297	152,641
KGACentroids	6738,672	†	†	†
Knn	2,469	351,125	2,703	108,344
DistKnn	1,625	142,344	1,734	40,656
Decision Tree	1,063	10,781	1,328	6,047
Random Forest	50,906	115,344	53,188	71,516

Tabela 1: Comparações entre tempos de processamento, em segundos.

3.1. Iris

Inicialmente serão analisados os resultados obtidos através da aplicação dos classificadores na base de dados *Iris*. Estatísticas obtidas do conjunto de pontuação durante as rodadas de teste em cada divisão da validação cruzada são apresentadas na Tabela 2. Note que todas as médias apresentaram-se dentro dos limites do intervalo de confiança (com o método *Stratified Random* possuindo o maior intervalo) e sobretudo que a maioria das métricas estão na mesma ordem de grandeza, a não ser nos classificadores puramente aleatórios (*ZeroR*, *Random* e *Stratified Random*). Estes era de se esperar que performassem mau, dada sua natureza randômica.

Visualmente, podemos observar que os métodos *Naive Bayes Gaussiano*, *KmeansCentroides*, *Knn*, *DistKnn*, *Árvore de Decisão* e *Florestas de Árvores* foram os que melhor performaram (Figura 1). Sobretudo os métodos *Árvore de Decisão* e *Florestas de Árvores*, que nem sequer apresentaram *outliers*.

Mas confirmemos nossa hipótese de que os métodos *Árvore de Decisão* e *Florestas de Árvores* performaram melhor. A Tabela 3 apresenta os $p-$valores de

	Média	*Desvio Padrão*	*Limite Inferior*	*Limite Superior*
ZeroR	3,333e-01	1,110e-16	3,333e-01	3,333e-01
Random	3,133e-01	1,049e-01	2,735e-01	3,532e-01
Stratified Random	3,000e-01	1,177e-01	2,553e-01	3,447e-01
OneR	8,956e-01	6,596e-02	8,705e-01	9,206e-01
Gaussian Naive Bayes	9,511e-01	5,145e-02	9,316e-01	9,707e-01
KmeansCentroids	9,489e-01	6,130e-02	9,256e-01	9,722e-01
Knn	9,422e-01	6,828e-02	9,163e-01	9,682e-01
DistKnn	9,467e-01	6,301e-02	9,227e-01	9,706e-01
Decision Tree	9,467e-01	4,355e-02	9,301e-01	9,632e-01
Random Forest	9,556e-01	4,969e-02	9,367e-01	9,744e-01
KGACentroids	8,577e-01	8,347e-02	8,266e-01	8,266e-01

Tabela 2: Estatísticas das acurácias obtidas em cada *fold* do ciclo externo para a base de dados *Iris*.

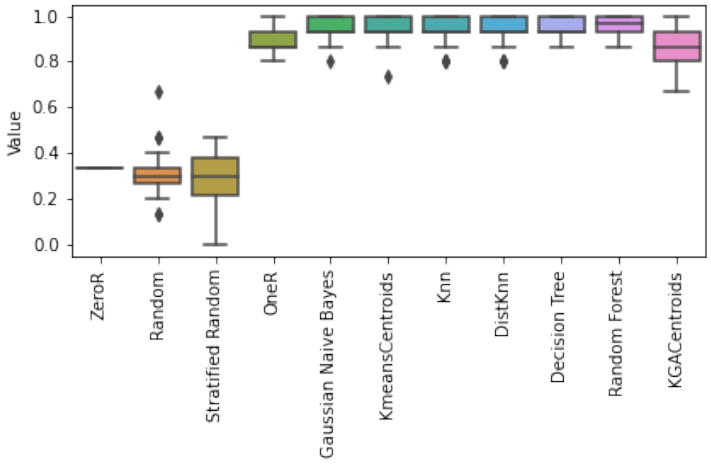

Figura 1: Boxplot dos resultados de cada classificador em cada *fold* para a base de dados *Iris*.

cada par de classificador, para um nível de significância de 95%. Segundo [2], o $p-$valor é a probabilidade de se obter uma estatística de teste igual ou mais extrema que aquela observada em uma amostra, sob a hipótese nula. Por exemplo, em testes de hipótese, pode-se rejeitar a hipótese nula a 5% caso o $p-$valor seja menor que 5%. Quando aceitamos hipótese nula como verdadeira, afirmamos

não haver diferença entre os grupos experimentais no nível da população.

Sob o contexto mencionado e os valores em negrito na Tabela 3, aceitamos a hipótese nula aos pares contendo apenas os métodos puramente aleatórios (*ZeroR*, *Random* e *Stratified Random*) e aos pares contendo os métodos anteriormente mencionados como melhores (*Naive Bayes Gaussiano*, *KmeansCentroides*, *Knn*, *DistKnn*, *Árvore de Decisão* e *Florestas de Árvores*). Ou seja, não podemos afirmar que há diferença entre os resultados experimentais no nível da população, portanto, não podemos afirmar que *Árvore de Decisão* e *Florestas de Árvores* foram os melhores classificadores. Apenas que **o conjunto de métodos *Naive Bayes Gaussiano*, *KmeansCentroides*, *Knn*, *DistKnn*, *Árvore de Decisão* e *Florestas de Árvores* performaram melhor na base de dados *Iris***.

	ZeroR	*Rand*	*Strat*	*OneR*	*GNB*	*KM*	*KNN*	*DKNN*	*DT*	*RFor*	*KGA*
ZeroR		**3,131e-01**	**1,380e-01**	1,246e-28	6,614e-33	1,135e-30	3,413e-29	2,772e-30	6,650e-35	1,976e-33	4,569e-25
Rand	**5,764e-01**		**6,473e-01**	5,771e-21	4,441e-21	1,008e-21	4,768e-20	8,335e-21	4,101e-22	5,554e-22	2,457e-19
Strat	**5,106e-01**	**8,082e-01**		4,726e-20	2,406e-22	2,822e-22	2,824e-21	6,874e-22	6,874e-22	5,865e-22	2,599e-19
OneR	1,469e-06	1,670e-06	1,715e-06		3,156e-04	2,497e-03	4,256e-03	1,158e-03	1,726e-05	5,187e-06	1,122e-03
GNB	1,133e-06	1,645e-06	1,689e-06	9,325e-04		**8,454e-01**	**3,801e-01**	**6,015e-01**	**6,015e-01**	**5,725e-01**	7,356e-07
KM	1,170e-06	1,683e-06	1,714e-06	5,205e-03	**1,000e+00**		**5,219e-01**	**8,012e-01**	**8,315e-01**	**5,725e-01**	4,607e-05
KNN	1,213e-06	1,718e-06	1,740e-06	1,020e-02	**3,785e-01**	**6,506e-01**		**3,256e-01**	**6,253e-01**	**2,061e-01**	2,546e-05
DKNN	1,174e-06	1,711e-06	1,743e-06	3,020e-03	**6,270e-01**	**1,000e+00**	**4,237e-01**		**1,000e+00**	**3,545e-01**	5,902e-06
DT	1,042e-06	1,647e-06	1,731e-06	2,401e-04	**6,179e-01**	**6,444e-01**	**6,444e-01**	**1,000e+00**		**1,608e-01**	2,516e-07
RFor	1,123e-06	1,613e-06	1,743e-06	1,431e-04	**6,078e-01**	**6,583e-01**	**2,234e-01**	**3,778e-01**	**1,817e-01**		8,948e-08
KGA	1,450e-06	2,000e-06	2,000e-06	1,734e-03	3,200e-05	3,403e-04	1,001e-04	6,964e-05	1,112e-05	8,243e-06	

Tabela 3: Resultados pareados ($p-$valores) de testes de hipótese entre os pares de métodos para a base de dados *Iris*. Na matriz triangular superior estão situados os resultados do teste$-t$ pareado (amostras dependentes) e na matriz triangular inferior os resultado do teste não paramétrico de wilcoxon. Os valores que rejeitam a hipótese nula para um nível de significância de 95% estão destacados em negrito.

3.2. Digits

Neste segundo conjunto de experimentos, as estatísticas obtidas do conjunto de pontuação durante as rodadas de teste em cada divisão da validação cruzada são apresentadas na Tabela 4. Novamente, nota-se que todas as médias apresentaram-se dentro dos limites do intervalo de confiança e que as médias estão na mesma ordem de grandeza, com apenas o *Stratified Random* diferenciando-se desta característica no limite inferior do intervalo de confiança (mas ainda sim próximo aos demais). Acerca do desvio padrão, os métodos diferenciam-se entre si, mas permanecem numa ordem de grandeza pequena. Novamente os classificadores puramente aleatórios (*ZeroR*, *Random* e *Stratified Random*) performaram mau e desta vez o classificador *OneR* equiparou-se a estes.

Diferentemente da base de dados *Iris*, podemos destacar a boa acurácia média de menos métodos, a saber, *KmeansCentroids*, *Knn*, *DistKnn* e *Random Forest* (veja a Figura 2). Note que dentre estes, *DistKnn* e *Random Forest* possuem a menor variação (menor desvio-padrão). Vejamos se nossa análise pode ser validada através dos testes envolvendo o cálculo do $p-$valor.

	Média	*Desvio Padrão*	*Limite Inferior*	*Limite Superior*
ZeroR	1,013e-01	2,436e-03	1,004e-01	1,022e-01
Random	1,005e-01	2,643e-02	9,050e-02	1,106e-01
Stratified Random	1,015e-01	1,776e-02	9,472e-02	1,082e-01
OneR	1,758e-01	2,493e-02	1,664e-01	1,853e-01
Gaussian Naive Bayes	7,843e-01	3,024e-02	7,728e-01	7,958e-01
KmeansCentroids	9,523e-01	1,644e-02	9,461e-01	9,586e-01
Knn	9,755e-01	1,117e-02	9,713e-01	9,798e-01
DistKnn	9,761e-01	9,951e-03	9,723e-01	9,799e-01
Decision Tree	8,579e-01	1,977e-02	8,504e-01	8,654e-01
Random Forest	9,777e-01	9,864e-03	9,740e-01	9,815e-01

Tabela 4: Estatísticas das acurácias obtidas em cada *fold* do ciclo externo para a base de dados *Digits*.

A Tabela 5 apresenta os resultados ($p-$valores) dos testes de hipótese entre os pares de métodos. Curiosamente, não há evidência estatística de que existam

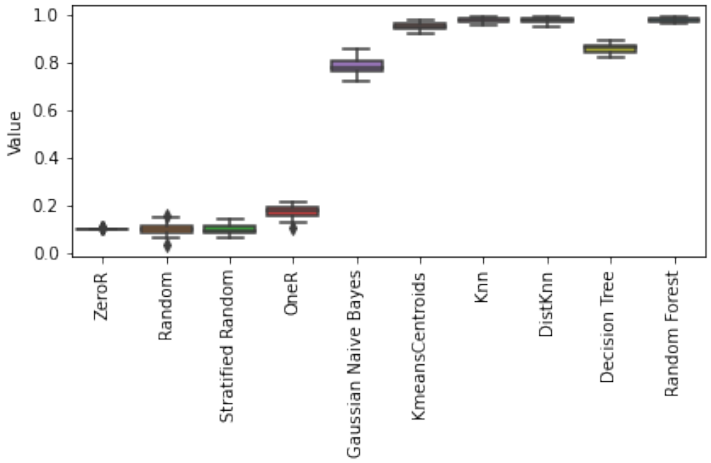

Figura 2: Boxplot dos resultados de cada classificador em cada *fold* para a base de dados *Digits*.

diferenças entre os resultados envolvendo os classificadores *DistKnn* e *Knn* (coloquemos então o *Knn* no grupo de melhores) e do *Random Forest* a estes dois. A partir do exposto, podemos apenas afirmar que os **métodos *Knn*, *DistKnn* e *Random Forest* performaram melhor na base de dados *Digits***.

	ZeroR	*Rand*	*Strat*	*OneR*	*GNB*	*KM*	*KNN*	*DKNN*	*DT*	*RFor*
ZeroR		**8,812e-01**	**9,578e-01**	7,750e-16	4,231e-41	1,340e-51	6,060e-57	2,733e-58	3,371e-47	2,115e-57
Rand	**7,672e-01**		**8,717e-01**	8,443e-11	1,983e-36	2,074e-42	3,094e-44	2,200e-44	1,568e-41	2,246e-44
Strat	**5,505e-01**	**9,886e-01**		5,385e-15	3,445e-38	9,969e-47	1,275e-48	3,980e-50	3,337e-45	1,764e-48
OneR	1,799e-06	3,327e-06	1,811e-06		2,276e-37	3,798e-43	2,529e-46	9,971e-47	2,770e-40	2,850e-45
GNB	1,798e-06	1,817e-06	1,820e-06	1,819e-06		2,578e-23	7,556e-26	9,354e-26	1,025e-12	1,003e-24
KM	1,792e-06	1,816e-06	1,796e-06	1,819e-06	1,819e-06		4,072e-10	4,015e-11	6,858e-19	1,740e-09
KNN	1,742e-06	1,809e-06	1,807e-06	1,815e-06	1,798e-06	2,654e-06		**5,576e-01**	1,333e-22	**2,665e-01**
DKNN	1,735e-06	1,812e-06	1,800e-06	1,816e-06	1,812e-06	2,636e-06	**8,933e-01**		1,910e-22	**4,096e-01**
DT	1,804e-06	1,813e-06	1,813e-06	1,811e-06	1,822e-06	1,816e-06	1,806e-06	1,805e-06		1,773e-22
RFor	1,751e-06	1,813e-06	1,809e-06	1,808e-06	1,814e-06	3,294e-06	**3,555e-01**	**3,803e-01**	1,796e-06	

Tabela 5: Resultados pareados ($p-$valores) de testes de hipótese entre os pares de métodos para a base de dados *Digits*. Na matriz triangular superior estão situados os resultados do teste$-t$ pareado (amostras dependentes) e na matriz triangular inferior os resultado do teste não paramétrico de wilcoxon. Os valores que rejeitam a hipótese nula para um nível de significância de 95% estão destacados em negrito.

3.3. Wine

De forma análoga e para a base de dados *Wine*, as estatísticas obtidas do conjunto de pontuação durante as rodadas de teste em cada divisão da validação cruzada são apresentadas numericamente na Tabela 6 e visualmente na Figura 3. Neste caso todos os métodos que já foram considerados melhores anteriomente apresentaram alguns outliers e dentre estes, *Decision Tree* é o que apresenta maiores variações. Apenas a título de menção, novamente os métodos aleatórios (*ZeroR*, *Random* e *Stratified Random*) performaram mal e que o método *OneR* ficou na média entre estes dois conjuntos, com um grande intervalo inter-quartis.

Portanto, procuraremos confirmar que os classificadores *Gaussian Naive Bayes*, *KmeansCentroids*, *Knn*, *DistKnn*, *Decision Tree* e *Random Forest* foram os melhores.

	Média	*Desvio Padrão*	*Limite Inferior*	*Limite Superior*
ZeroR	3,993e-01	2,471e-02	3,900e-01	4,087e-01
Random	3,157e-01	1,150e-01	2,720e-01	3,594e-01
Stratified Random	3,318e-01	1,070e-01	2,912e-01	3,725e-01
OneR	7,136e-01	9,791e-02	6,764e-01	7,508e-01
Gaussian Naive Bayes	9,734e-01	4,820e-02	9,551e-01	9,917e-01
KmeansCentroids	9,491e-01	6,533e-02	9,243e-01	9,739e-01
Knn	9,550e-01	5,334e-02	9,348e-01	9,753e-01
DistKnn	9,549e-01	5,549e-02	9,338e-01	9,760e-01
Decision Tree	9,011e-01	7,886e-02	8,711e-01	9,310e-01
Random Forest	9,813e-01	3,338e-02	9,686e-01	9,939e-01

Tabela 6: Estatísticas das acurácias obtidas em cada *fold* do ciclo externo para a base de dados *Wine*.

A Tabela 7 apresenta os resultados ($p-$valores) dos testes de hipótese entre os pares de métodos. Nesta base de dados, alguns fatos importantes ocorreram: (1) existem evidências de diferenças entre os resultados obtidos pelo *Decision Tree* em relação a todos os demais, e de forma análoga, (2) que há diferenças entre o *Random Forest* a todos os demais, a não ser ao *Gaussian Naive Bayes* segundo o teste$-t$ para as amostras dependentes.

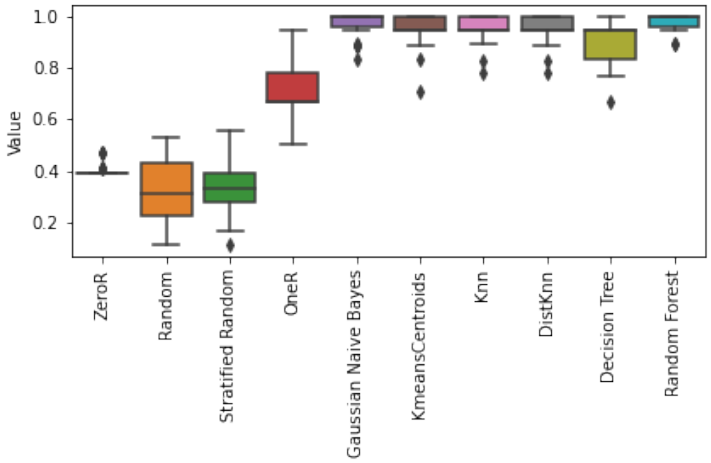

Figura 3: Boxplot dos resultados de cada classificador em cada *fold* para a base de dados *Wine*.

Voltemos à Tabela 6. Segundo as médias, o método *Random Forest* performou melhor, seguido pelo *Gaussian Naive Bayes*. Mas de (2), não há evidência estatística de que há diferença entre estes dois classificadores. No mesmo sentido desta análise, entram neste conjunto de melhores resultados, segundo o teste$-t$, os métodos *Knn*, *DistKnn* e *KmeansCentroids* (veja que estão em negrito na Tabela 7). Ou seja, **seguindo a perspectiva do teste$-t$, o conjunto de classificadores *Gaussian Naive Bayes*, *KmeansCentroids*, *Knn*, *DistKnn*, *Decision Tree* e *Random Forest* performaram melhor. Mas seguindo a perspectiva do teste de wilcoxon e os resultados de médias, o método *Random Forest* performou melhor**.

	ZeroR	*Rand*	*Strat*	*OneR*	*GNB*	*KM*	*KNN*	*DKNN*	*DT*	*RFor*
ZeroR		3,025e-04	2,049e-03	9,849e-17	1,143e-29	3,263e-26	3,463e-29	1,084e-28	1,728e-24	5,392e-34
Rand	5,903e-04		**6,086e-01**	8,081e-14	1,936e-21	2,825e-20	2,739e-21	5,715e-21	1,156e-19	1,752e-23
Strat	3,903e-03	**6,961e-01**		2,376e-14	1,327e-24	1,149e-22	1,101e-23	2,675e-23	3,757e-23	2,405e-25
OneR	1,566e-06	1,781e-06	2,654e-06		1,124e-13	4,052e-12	1,368e-11	1,293e-11	5,355e-08	3,528e-14
GNB	8,507e-07	1,761e-06	1,766e-06	1,630e-06		2,660e-02	1,758e-02	1,049e-02	3,926e-07	2,010e-01
KM	1,493e-06	1,792e-06	1,767e-06	1,750e-06	3,284e-02		**6,102e-01**	**6,266e-01**	4,144e-03	3,670e-03
KNN	1,424e-06	1,786e-06	1,709e-06	3,699e-06	**1,015e-01**	**1,000e+00**		**9,686e-01**	3,369e-06	1,546e-03
DKNN	1,424e-06	1,782e-06	1,712e-06	3,746e-06	3,900e-02	**9,150e-01**	**1,000e+00**		6,753e-06	2,707e-03
DT	1,595e-06	1,795e-06	1,782e-06	2,177e-05	6,974e-05	8,549e-03	1,226e-04	2,976e-04		4,657e-08
RFor	9,525e-07	1,762e-06	1,738e-06	1,628e-06	**2,008e-01**	2,317e-03	3,686e-03	5,800e-03	1,682e-05	

Tabela 7: Resultados pareados ($p-$valores) de testes de hipótese entre os pares de métodos para a base de dados *Wine*. Na matriz triangular superior estão situados os resultados do teste$-t$ pareado (amostras dependentes) e na matriz triangular inferior os resultado do teste não paramétrico de wilcoxon. Os valores que rejeitam a hipótese nula para um nível de significância de 95% estão destacados em negrito.

3.4. Breast Cancer

Neste último conjunto de testes cujos resultados correspondentes estão apresentados nas Tabelas 8, 9 e Figura 4, novamente os classificadores *Gaussian Naive Bayes*, *KmeansCentroids*, *Knn*, *DistKnn*, *Decision Tree* e *Random Forest* foram os que melhor performaram, com o *OneR* na média e os aleatórios (*ZeroR*, *Random* e *Stratified Random*) de maneira pior. Façamos a análise do $p-$valor.

	Média	*Desvio Padrão*	*Limite Inferior*	*Limite Superior*
ZeroR	6,274e-01	6,966e-03	6,248e-01	6,301e-01
Random	5,033e-01	5,512e-02	4,823e-01	5,242e-01
Stratified Random	5,266e-01	6,996e-02	5,000e-01	5,532e-01
OneR	8,348e-01	3,669e-02	8,209e-01	8,487e-01
Gaussian Naive Bayes	9,338e-01	2,635e-02	9,238e-01	9,438e-01
KmeansCentroids	9,555e-01	3,196e-02	9,434e-01	9,676e-01
Knn	9,654e-01	2,377e-02	9,564e-01	9,745e-01
DistKnn	9,654e-01	2,377e-02	9,564e-01	9,745e-01
Decision Tree	9,268e-01	2,936e-02	9,156e-01	9,379e-01
Random Forest	9,602e-01	2,395e-02	9,511e-01	9,693e-01

Tabela 8: Estatísticas das acurácias obtidas em cada *fold* do ciclo externo para a base de dados *Breast Cancer*.

Inicialmente, destaca-se que os métodos *Knn*, *DistKnn* sequer apresentaram alguma diferença não nula para que fosse computado a estatística de interesse, ou seja, performaram de forma idêntica (ou ao menos muito aproximada).

Ainda, que há dois grupos entre os melhores de correlatados através de ambos os testes teste$-t$ pareado e wilcoxon. A saber, (1) *Gaussian Naive Bayes* e *Decision Tree*; (2) *KmeansCentroids*, *Knn*, *DistKnn* e *Random Forest*. Observando a Tabela 4 nos resultados de médias, vemos que os classificadores pertencentes ao grupo (2) possuem melhores acurácias. Portanto, podemos afirmar que **para a base de dados *Breast Cancer* os métodos *KmeansCentroids*, *Knn*, *DistKnn* e *Random Forest* performaram melhor**.

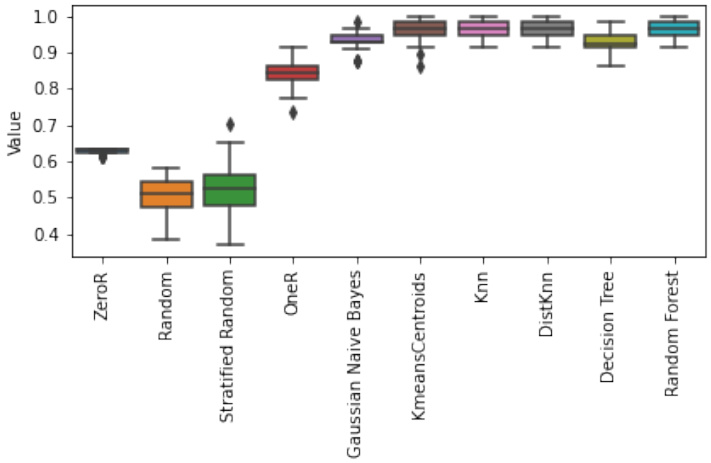

Figura 4: Boxplot dos resultados de cada classificador em cada *fold* para a base de dados *Breast Cancer*.

	ZeroR	*Rand*	*Strat*	*OneR*	*GNB*	*KM*	*KNN*	*DKNN*	*DT*	*RFor*
ZeroR		1,540e-12	1,418e-08	3,213e-23	1,354e-31	2,479e-30	6,950e-34	6,950e-34	4,006e-30	6,412e-34
Rand	1,781e-06		**1,710e-01**	6,921e-21	3,813e-26	8,611e-27	1,051e-28	1,051e-28	9,224e-25	2,633e-27
Strat	1,113e-05	**1,645e-01**		2,453e-19	9,913e-23	7,354e-23	6,472e-24	6,472e-24	3,519e-22	1,458e-23
OneR	1,778e-06	1,798e-06	1,811e-06		1,815e-13	1,502e-15	3,214e-15	3,214e-15	1,845e-11	6,259e-16
GNB	1,680e-06	1,790e-06	1,804e-06	1,787e-06		1,603e-03	2,464e-07	2,464e-07	**1,978e-01**	3,252e-06
KM	1,754e-06	1,805e-06	1,812e-06	1,959e-06	2,577e-03		**1,041e-01**	**1,041e-01**	2,684e-04	**3,498e-01**
KNN	1,666e-06	1,794e-06	1,811e-06	2,634e-06	2,060e-05	**1,435e-01**		*	1,744e-08	**2,212e-01**
DKNN	1,666e-06	1,794e-06	1,811e-06	2,634e-06	2,060e-05	**1,435e-01**	*		1,744e-08	**2,212e-01**
DT	1,709e-06	1,805e-06	1,795e-06	2,411e-06	**5,016e-01**	9,293e-04	1,203e-05	1,203e-05		3,376e-07
RFor	1,742e-06	1,807e-06	1,800e-06	1,775e-06	7,001e-05	**2,924e-01**	**2,324e-01**	**2,324e-01**	1,696e-05	

Tabela 9: Resultados pareados ($p-$valores) de testes de hipótese entre os pares de métodos para a base de dados *Breast Cancer*. Na matriz triangular superior estão situados os resultados do teste$-t$ pareado (amostras dependentes) e na matriz triangular inferior os resultado do teste não paramétrico de wilcoxon. Os valores que rejeitam a hipótese nula para um nível de significância de 95% estão destacados em negrito.

4. Conclusões e trabalhos futuros

Realizando uma suma dos melhores resultados em cada base de dados, temos:

- *Iris*: *Naive Bayes Gaussiano*, *KmeansCentroides*, *Knn*, *DistKnn*, *Árvore de Decisão* e *Florestas de Árvores*
- *Digits*: *Knn*, *DistKnn* e *Random Forest*
- *Wine*: (teste$-t$) *Gaussian Naive Bayes*, *KmeansCentroids*, *Knn*, *DistKnn*, *Decision Tree* e *Random Forest*; (wilcoxon) *Random Forest*.
- *Breast Cancer*: *KmeansCentroids*, *Knn*, *DistKnn* e *Random Forest*

Diante desses resultados, pontuamos que, a não ser ao caso da base de dados *Wine* sob a perspectiva do teste de wilcoxon, **os métodos *Knn* e *DistKnn* foram os únicos a apresentarem os melhores resultados em todos os experimentos realizados**.

Acerca de trabalhos futuros, ressalta-se a possibilidade da aplicação de outras métricas estatísticas de forma a desempatar o melhor desempenho na base de dados *Wine*, bem como, se possível, que defina-se um melhor método dentre o conjunto destacado. Também encoraja-se que os classificadores aqui utilizados sejam aplicados a outras bases de dados. Relativo ao método *KGACentroids*, que este seja otimizado de forma que seu tempo de execução não seja tão distoante dos demais, e que assim este também possa ser comparável de maneira eficaz.

Referências

[1] Wikipedia contributors, Statistical classification — Wikipedia, the free encyclopedia, [Online; accessed in 18 Apr. 2021] (2021).
URL `https://en.wikipedia.org/w/index.php?title=Statistical_classification&oldid=1013158778`

[2] Wikipédia, Valor-p — wikipédia, a enciclopédia livre, [Online; accessed in 18 Apr. 2021] (2020).
URL https://pt.wikipedia.org/w/index.php?title=Valor-p&oldid=59724859