

## Segundo Trabalho de Inteligência Artificial 2019/2

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. As técnicas escolhidas são: ZeroR, OneR, OneR Probabilístico, Centróide, Centróide OneR, Naive Bayes Gaussiano, Knn, Árvore de Decisão, Rede Neural e Florestas de Árvores. As bases de dados a serem utilizadas são iris, digits, wine e breast cancer.

Para cada base, o procedimento experimental será dividido em duas etapas.

A primeira etapa consiste no treino e teste com validação cruzada de 10 folds dos classificadores que não possuem hiperparâmetros, isto é, os classificadores ZeroR, OneR, OneR Probabilístico, Centróide, Centróide OneR e Naive Bayes Gaussiano. Os resultados de cada classificador devem ser apresentados numa tabela contendo a média das acurácias e o desvio padrão dos resultados obtidos em cada fold, e também através do boxplot dos resultados de cada classificador em cada fold.

A segunda etapa consiste no treino, validação e teste dos classificadores que precisam de ajuste de hiperparâmetros, isto é, os classificadores Knn, Árvore de Decisão, Redes Neurais e Florestas de Árvores. Neste caso o procedimento de treinamento, validação e teste será realizado através de ciclos aninhados de validação e teste, com o ciclo interno de validação contendo 4 folds e o externo de teste com 10 folds. A busca em grade (grid search) do ciclo interno deve considerar os seguintes valores de hiperparâmetros de cada técnica de aprendizado:

Knn: [n\_neighbors = 1, 3, 5, 7, 10]

Árvore de Decisão: [max\_depth = None, 3, 5, 10]

Rede Neural: { [max\_iter = 50, 100, 200], [hidden\_layer\_sizes=(15,)]}

Florestas de Árvores: [n\_estimators = 10, 20, 50, 100]

Os resultados de cada classificador devem ser apresentados numa tabela contendo a média das acurácias e o desvio padrão dos resultados obtidos em cada fold do ciclo externo, e também através do boxplot dos resultados de cada classificador em cada fold.

Os métodos ZeroR, OneR, OneR Probabilístico, Centróide, Centróide OneR devem ser implementados. Os métodos Naive Bayes Gaussiano, Knn, Árvores de Decisão, Rede Neural e Florestas de Árvores estão disponíveis no scikit-learn. Suas descrições podem ser acessadas respectivamente em:

<https://scikit-learn.org/stable/modules/generated/>

[sklearn.naive\\_bayes.GaussianNB.html#sklearn.naive\\_bayes.GaussianNB](https://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.GaussianNB.html#sklearn.naive_bayes.GaussianNB)

<https://scikit-learn.org/stable/modules/generated/>

[sklearn.tree.DecisionTreeClassifier.html#sklearn.tree.DecisionTreeClassifier](https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html#sklearn.tree.DecisionTreeClassifier)

<https://scikit-learn.org/stable/modules/generated/>

[sklearn.neural\\_network.MLPClassifier.html#sklearn.neural\\_network.MLPClassifier](https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html#sklearn.neural_network.MLPClassifier)

<https://scikit-learn.org/stable/modules/generated/>

[sklearn.ensemble.RandomForestClassifier.html#sklearn.ensemble.RandomForestClassifier](https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html#sklearn.ensemble.RandomForestClassifier)

## **1. Artigo**

Após a realização dos experimentos, um artigo descrevendo todo o processo experimental realizado deverá ser escrito em latex usando o software overleaf. O artigo deve ser estruturado contendo os seguintes componentes:

1. Título
2. Resumo
3. Seção 1. Introdução
4. Seção 2. Descrição Resumida dos Datasets
5. Seção 3. Descrição Resumidas dos Métodos Implementados ( ZeroR, OneR, OneR Probabilístico, Centróide, Centróide OneR)
6. Seção 4. Descrição dos Experimentos Realizados
  - a. Seção 4.1 Iris
  - b. Seção 4.2 Digits
  - c. Seção 4.3 Wine
  - d. Seção 4.4 Breast Cancer
7. Seção 5. Conclusões
  - a. Análise geral dos resultados
  - b. Contribuições do Trabalho
  - c. Melhorias e trabalhos futuros
8. Referências Bibliográficas

## **2. Informações Complementares**

Os gráficos bloxplot requeridos no treino e no teste devem ser gerados usando função específica do pacote seaborn (ver instruções de instalação e uso no apêndice A deste enunciado).

O apêndice B deste enunciado apresenta instruções de instalação e uso do overleaf para a escrita do artigo.

## **3. Condições de Entrega**

O trabalho deve ser feito individualmente e submetido por e-mail até as 23:59 horas do dia 25/11/2019 para o endereço [fvarejao@gmail.com](mailto:fvarejao@gmail.com) com o subject IA\_TRABALHO\_2\_NomedoAluno\_SobrenomedoAluno. O e-mail deve conter um arquivo .pdf com o artigo produzido no trabalho e .zip contendo todos os arquivos com código fonte em python utilizados no trabalho. Tanto arquivo pdf quanto zip devem possuir o mesmo nome do subject do e-mail enviado. Note que a data limite já leva em conta um dia adicional de tolerância para o caso de problemas de submissão via rede. Isso significa que o aluno deve submeter seu trabalho até no máximo um dia antes da data limite. Se o aluno resolver submeter o trabalho na data limite, estará fazendo isso assumindo o risco do trabalho ser cadastrado no sistema após o prazo. Em caso de recebimento do trabalho após a data limite, o

trabalho não será avaliado e a nota será ZERO. Aluno que receber zero por este motivo e vier pedir para o professor considerar o trabalho estará cometendo um ato de DESRESPEITO ao professor e estará sujeito a perda adicional de pontos na média.

### **Observação importante**

**Caso haja algum erro neste documento, serão publicadas novas versões e divulgadas erratas em sala de aula. É responsabilidade do aluno manter-se informado, freqüentando as aulas ou acompanhando as novidades na página da disciplina na Internet.**

### **Apêndice A. . Boxplots usando seaborn**

Instruções de instalação para python3:

```
sudo apt install python3 python3-pip
pip3 install seaborn
```

```
# arquivo bloxplot.py
```

```
# library & dataset
```

```
import seaborn as sns
```

```
Instruções de instalaçãoimport matplotlib.pyplot as plt
```

```
def example1():
```

```
    mydata=[1,2,3,4,5,6,12]
```

```
    sns.boxplot(y=mydata) # Also accepts numpy arrays
```

```
    plt.show()
```

```
def example2():
```

```
    df = sns.load_dataset('iris')
```

```
    #returns a DataFrame object. This dataset has 150 examples.
```

```
    #print(df)
```

```
    # Make boxplot for each group
```

```
    sns.boxplot( data=df.loc[:,:] )
```

```
    # loc[:,:] means all lines and all columns
```

```
    plt.show()
```

```
example1()
```

```
example2()
```

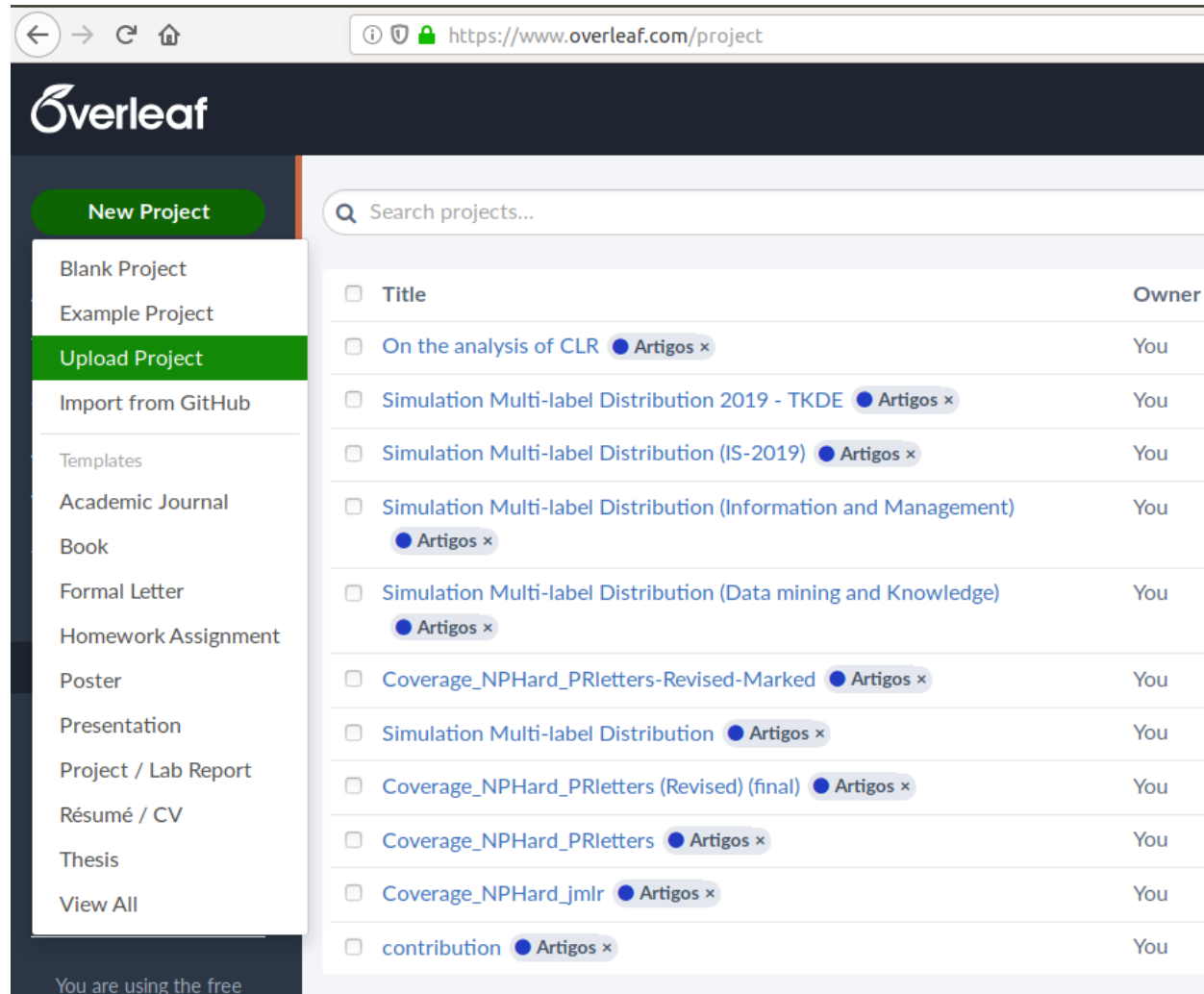
Uso em python3:

```
python3 boxplot.py
```

### **Apêndice B. Artigo em Latex usando Overleaf**

Juntamente com este enunciado foi disponibilizado um arquivo zip com o template de latex para confecção do artigo. O primeiro passo a ser feito é criar uma conta pessoal no Overleaf (<https://www.overleaf.com/register>). Uma vez criada sua conta, deve-se entrar nela. Para incluir

o template no overleaf, basta apenas selecionar "New Project>Upload Project" e selecionar o arquivo zip, como mostrado na figura abaixo. Não é necessário descompactar, faça o upload do zip direto. Lembrar de renomear o artigo após o upload do arquivo.



The screenshot shows the Overleaf website interface. The browser address bar displays <https://www.overleaf.com/project>. The Overleaf logo is in the top left. A sidebar on the left contains a 'New Project' button, which is open, showing a list of options: 'Blank Project', 'Example Project', 'Upload Project' (highlighted in green), 'Import from GitHub', and a 'Templates' section with various document types like 'Academic Journal', 'Book', 'Formal Letter', 'Homework Assignment', 'Poster', 'Presentation', 'Project / Lab Report', 'Résumé / CV', 'Thesis', and 'View All'. At the bottom of the sidebar, it says 'You are using the free'. The main content area has a search bar 'Search projects...' and a table of existing projects.

<input type="checkbox"/>	Title	Owner
<input type="checkbox"/>	On the analysis of CLR <span>● Artigos x</span>	You
<input type="checkbox"/>	Simulation Multi-label Distribution 2019 - TKDE <span>● Artigos x</span>	You
<input type="checkbox"/>	Simulation Multi-label Distribution (IS-2019) <span>● Artigos x</span>	You
<input type="checkbox"/>	Simulation Multi-label Distribution (Information and Management) <span>● Artigos x</span>	You
<input type="checkbox"/>	Simulation Multi-label Distribution (Data mining and Knowledge) <span>● Artigos x</span>	You
<input type="checkbox"/>	Coverage_NPHard_PRletters-Revised-Marked <span>● Artigos x</span>	You
<input type="checkbox"/>	Simulation Multi-label Distribution <span>● Artigos x</span>	You
<input type="checkbox"/>	Coverage_NPHard_PRletters (Revised) (final) <span>● Artigos x</span>	You
<input type="checkbox"/>	Coverage_NPHard_PRletters <span>● Artigos x</span>	You
<input type="checkbox"/>	Coverage_NPHard_jmlr <span>● Artigos x</span>	You
<input type="checkbox"/>	contribution <span>● Artigos x</span>	You