Considerando os conjuntos de dados

1. digits – [https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load\_digits.html#sklearn.datasets.load\_digits](https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_digits.html" \l "sklearn.datasets.load_digits)
2. breast\_cancer – [https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load\_breast\_cancer.html#sklearn.datasets.load\_breast\_cancer](https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_breast_cancer.html" \l "sklearn.datasets.load_breast_cancer)
3. wine – [https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load\_wine.html#sklearn.datasets.load\_wine](https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_wine.html" \l "sklearn.datasets.load_wine)

execute os algoritmos a seguir, com os respectivos hiperparâmetros, na busca pelos melhores hiperparâmetros, e avalie o melhor modelo em um conjunto de teste a parte (20% teste, 80% treinamento). Use 3-fold cv para seleção dos hiperparâmetros. Todo o processo será avaliado com a medida f1\_weighted.

|  |  |  |
| --- | --- | --- |
| DecisionTreeClassifier | 'criterion':['gini', 'entropy'], 'max\_depth':[3, 7, 9] | random\_state=0 |
| KNeighborsClassifier | 'n\_neighbors':[1, 3, 5, 7], 'p':[1, 2] |  |
| GaussianNB | Default |  |
| SVC | 'kernel':['linear', 'poly', 'rbf', 'sigmoid'], 'C':[1, 3, 7] |  |
| MLPClassifier | 'hidden\_layer\_sizes':[50, 30, 10] | max\_iter=5000, random\_state=0 |

Ao final, para cada conjunto, exiba a melhor configuração e o valor final da medida no conjunto de teste, como nas tabelas abaixo:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Conjunto | Instâncias | Atributos | Classes | Melhor Configuração | Desempenho no Teste |
| digits | 1797 | 64 | ? | svm com param = 'kernel':['linear', 'poly', 'rbf', 'sigmoid'], 'C':[1, 3, 7] | **0.9750182155092152** |
| breast\_cancer | 569 | 30 | ? | svm com param = 'kernel':['linear', 'poly', 'rbf', 'sigmoid'], 'C':[1, 3, 7] | **0.9576217776258756** |
| wine | 178 | 30 | ? | nb com param = Default | **0.9604749386595953** |

|  |  |  |
| --- | --- | --- |
| digits  tree 0.7572418632025579  knn 0.9638337100004549  nb 0.8158505897021469  **svm 0.9750182155092152**  mlp 0.9083718417581751 | breast\_cancer  tree 0.9100473740898419  knn 0.9235349676545388  nb 0.9362862496669241  **svm 0.9576217776258756**  mlp 0.9291410841746736 | wine  tree 0.863731867759593  knn 0.6998436347104627  **nb 0.9604749386595953**  svm 0.9262906273502898  mlp 0.36288570186875274 |

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

Faça o mesmo procedimento para o conjunto

1. diabetes – [https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load\_diabetes.html#sklearn.datasets.load\_diabetes](https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_diabetes.html" \l "sklearn.datasets.load_diabetes)

Contudo, o mesmo é um conjunto de regressão. Então pesquisem os métodos correspondentes e considere a seguinte tabela:

|  |  |  |
| --- | --- | --- |
| DecisionTreeClassifier | 'criterion':['mse', 'mae'], 'max\_depth':[3, 7, 9] | random\_state=0 |
| KNeighborsClassifier | 'n\_neighbors':[1, 3, 5, 7], 'p':[1, 2] |  |
| SVC | 'kernel':['linear', 'poly', 'rbf', 'sigmoid'], 'C':[1, 3, 7], 'epsilon':[0.1, 0.2] |  |
| MLPClassifier | 'hidden\_layer\_sizes':[50, 30, 10] | max\_iter=9000, random\_state=0 |

Neste caso, a medida de scoring=('neg\_mean\_squared\_error') (veja a semântica da medida (neg)) e a de avaliação mean\_squared\_error e r2\_score. Ao final, faça novamente a tabela para este conjunto de dados.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Conjunto | Instâncias | Atributos | Melhor Configuração | Desempenho no Teste |
| diabetes | 442 | 10 | knn com param = {'n\_neighbors': 7, 'p': 2} | **0.48** |

|  |
| --- |
| diabetes  tree 0.35880719943380274  knn 0.40681573456538517  nb 0.009054973340687627  **svm 0.48378501475332863**  mlp 0.010772175722745359 |