Grupo: Lucas Bastelli Spagnol e Pedro Gabriel Gonçalves

Atividade 2

Considerando os conjuntos de dados

* 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)
* 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)
* 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 | 10 | SVC com param = {'C': 3, 'kernel': 'rbf'} | **0.9916681909639657** |
| breast\_cancer | 569 | 30 | 2 | SVC com param = {'C': 3, 'kernel': 'linear'} | **0.9650119617224879** |
| wine | 178 | 30 | 3 | SVC com param = {'C': 1, 'kernel': 'linear'} | **0.9728425696167632** |

|  |  |  |
| --- | --- | --- |
| digits  tree 0.850427189256138  KNN 0.9833405264212088  NB 0.8252819198016837  **SVC 0.9916681909639657**  mlp 0.9611593612622309 | breast\_cancer  tree 0.93884293641162  KNN 0.9560669894569158  NB 0.9298245614035088  **SVC 0.9650119617224879**  mlp 0.9125299043062202 | wine  tree 0.9449984172206395  KNN 0.8364350703060381  NB 0.9152765497593083  **SVC 0.9728425696167632**  mlp 0.9449984172206395 |

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

Faça o mesmo procedimento para o conjunto

* 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 | SVR com param = {'C': 7, 'epsilon': 0.1, 'kernel': 'sigmoid'} | **0.35** |

|  |
| --- |
| diabetes  R2:  treeR2 0.06821179414519896  Knn\_R2 0.22012789155372325  **SVR\_R2 0.3592936331106341**  MLP\_R2 0.3256745589785328  Scores:  treeMSE 4778.141926493388  Knn\_MSE 3999.1272643889015  **SVR\_MSE 3285.49549669057**  MLP\_MSE 3457.8916556364948 |

O melhor resultado foi o SVR por possuir o r2\_score mais próximo de 1.