

Seleção de Características

Abordagem Wrapper

Prof. Dr. Danilo S. Sanches



Introdução

- Seleção de atributos é o processo que remove atributos irrelevantes e redundantes do conjunto de dados;
- O modelo, por sua vez, será de complexidade reduzida e, portanto, mais fácil de interpretar.

All Features



Feature Selection



Final Features

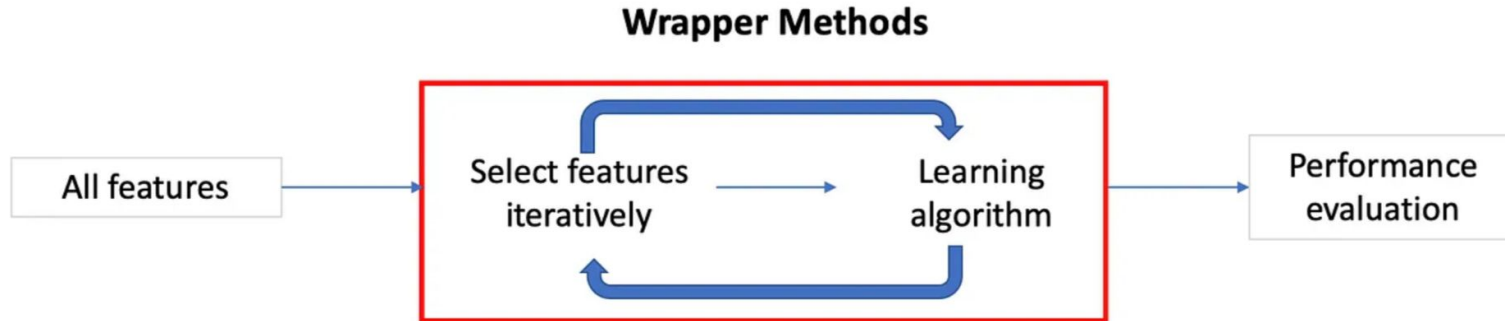


Fonte:

<https://themanoftalent.medium.com/feature-selection-9b1609f1f6b0>

Abordagem Wrapper

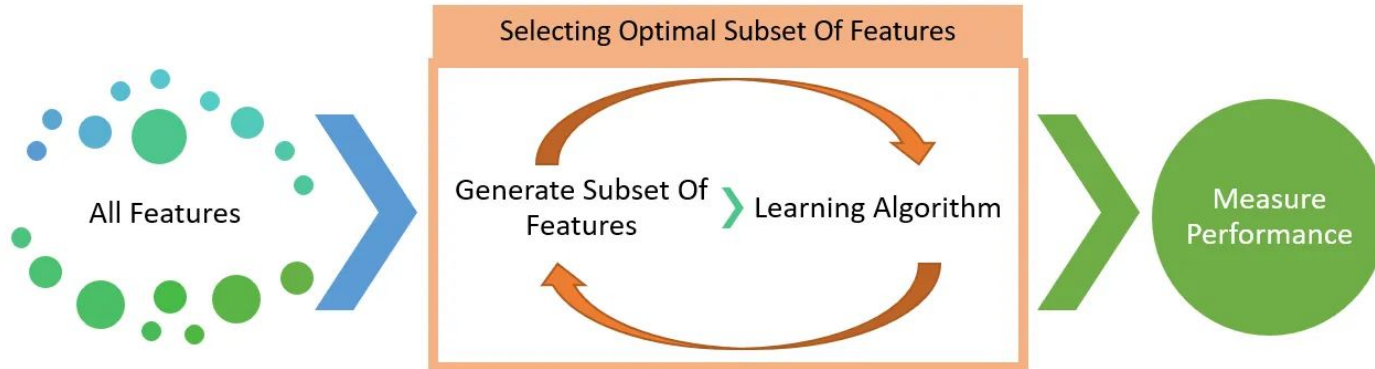
- Encontra o melhor conjunto de variáveis, dentre as disponíveis nos seus dados, para um determinado algoritmo de Machine Learning;
- Depende do algoritmo de Machine Learning que será utilizado.



<https://towardsdatascience.com/feature-selection-for-machine-learning-in-python-wrapper-methods-2b5e27d2db31>

Principais técnicas

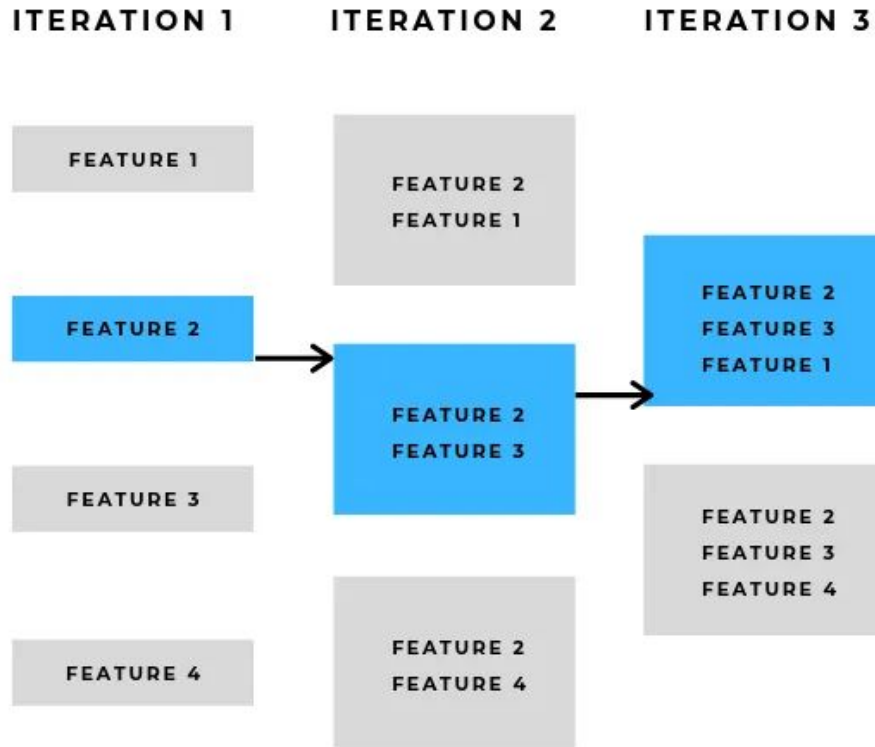
- Sequential Feature Selector (SFS)
 - Forward selection
 - Backward elimination
- Recursive Feature Elimination (RFE)



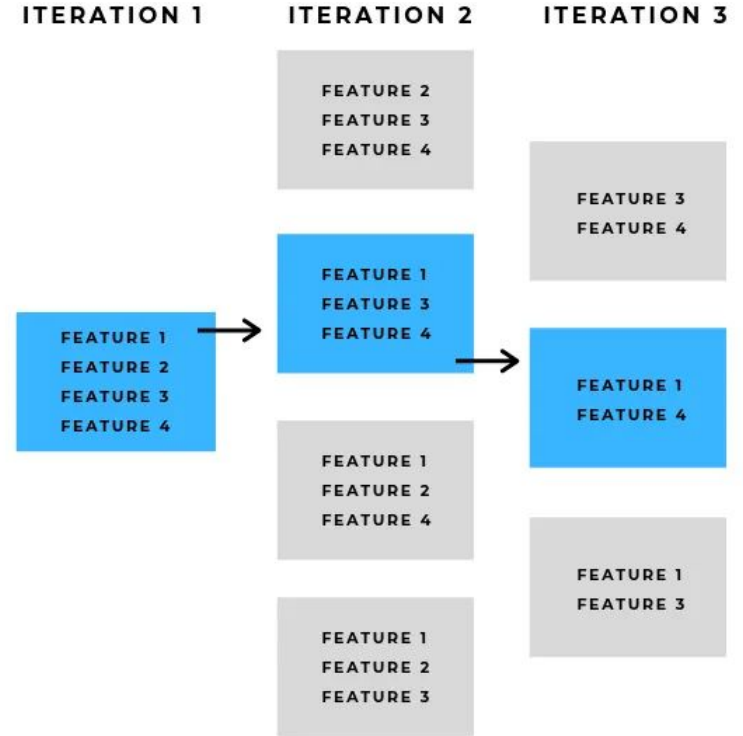
Sequential Feature Selection (SFS)

- A seleção sequencial de atributos (SFS) é um algoritmo greedy que adiciona ou remove iterativamente atributos de um conjunto de dados para melhorar o desempenho de um modelo preditivo;
- SFS pode ser seleção para frente (**forward**) ou para trás (**backward**);
- Classe **Sequential Feature Selector**: oferece suporte à seleção para frente e para trás;
- Funciona adicionando ou removendo iterativamente atributos de um conjunto de dados para melhorar o desempenho de um modelo preditivo.

Forward



Backward



Recursive Feature Elimination (RFE)

- Classe: **`sklearn.feature_selection.RFE`**
- RFE é computacionalmente menos complexo e utiliza coeficientes de peso dos atributos ou importância dos atributos para eliminar recursos recursivamente;
- Difere da abordagem SFS, que elimina (ou adiciona) recursos com base em uma métrica de desempenho de classificador definida pelo usuário.

Métodos Wrapper em Python



1.13. Feature selection

The classes in the `sklearn.feature_selection` module can be used for feature selection/dimensionality reduction on sample sets, either to improve estimators' accuracy scores or to boost their performance on very high-dimensional datasets.

Scikit-learn - Algoritmos

RFE

```
class sklearn.feature_selection.RFE(estimator, *,  
n_features_to_select=None, step=1, verbose=0,  
importance_getter='auto')
```

RFECV

```
class sklearn.feature_selection.RFECV(estimator, *, step=1,  
min_features_to_select=1, cv=None, scoring=None, verbose=0,  
n_jobs=None, importance_getter='auto')
```

SequentialFeatureSelector

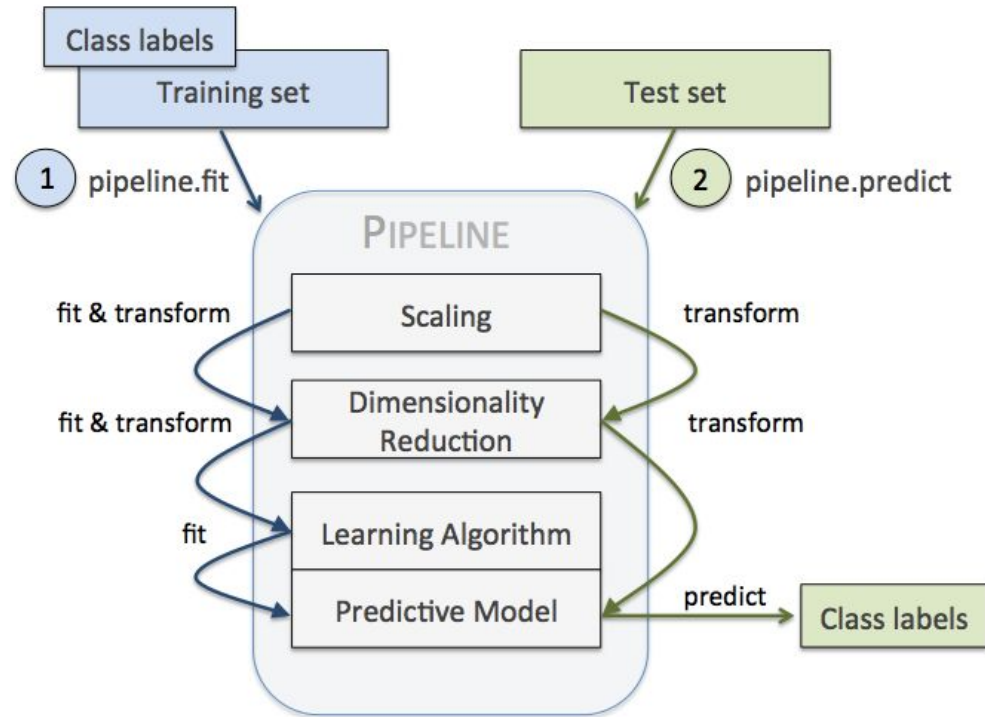
```
class  
sklearn.feature_selection.SequentialFeatureSelector(estimator,  
*, n_features_to_select='auto', tol=None, direction='forward',  
scoring=None, cv=5, n_jobs=None)
```

[\[source\]](#)

Workflow com Pipeline

- Pipeline é um conjunto de tarefas que acontecem em sequência;
- A saída de uma tarefa se torna a entrada da próxima até que no final desta sequência seja gerado o resultado;
- Pipelines tornam as etapas de pré-processamento de dados mais rápida;
- Pode ser facilmente utilizado para o conjunto de teste, tornando ainda mais fácil lidar com conjuntos de teste que ainda não foram transformados ou com novos dados.

Workflow com Pipeline



(Image from Raschka, Sebastian. Python machine learning. Birmingham, UK: Packt Publishing, 2015. Print