

Machine Learning Implementation

How to implement machine learning algorithms?

Tools for Machine Learning



Weka

- <https://www.cs.waikato.ac.nz/ml/weka/>



Python

- <https://www.python.org/>

Data Format

Data Format

- CSV
- ARFF
- LIBSVM

Data Format

- CSV

- ARFF

- LIBSVM

Label	D_1	D_2	D_3	...	D_n
1					
0					
0					
1					
0					
...					
1					

Data Format

- CSV
- ARFF
- LIBSVM

```
@RELATION iris

@ATTRIBUTE sepallength  REAL
@ATTRIBUTE sepalwidth   REAL
@ATTRIBUTE petallength  REAL
@ATTRIBUTE petalwidth   REAL
@ATTRIBUTE class        {Iris-setosa,Iris-versicolor,Iris-virginica}

@DATA
5.1,3.5,1.4,0.2,Iris-setosa
4.9,3.0,1.4,0.2,Iris-setosa
4.7,3.2,1.3,0.2,Iris-setosa
4.6,3.1,1.5,0.2,Iris-setosa
5.0,3.6,1.4,0.2,Iris-setosa
5.4,3.9,1.7,0.4,Iris-setosa
4.6,3.4,1.4,0.3,Iris-setosa
5.0,3.4,1.5,0.2,Iris-setosa
4.4,2.9,1.4,0.2,Iris-setosa
4.9,3.1,1.5,0.1,Iris-setosa
5.4,3.7,1.5,0.2,Iris-setosa
4.8,3.4,1.6,0.2,Iris-setosa
4.8,3.0,1.4,0.1,Iris-setosa
4.3,3.0,1.1,0.1,Iris-setosa
5.8,4.0,1.2,0.2,Iris-setosa
```

SCIKIT-LEARN

- Scikit-learn provides a range of supervised and unsupervised learning algorithms via a consistent interface in Python.
- The library is focused on modeling data.

SCIKIT-LEARN Library

2007

Developed by
David Cournapeau


2010

- The first public release (v0.1 beta) was published in late January 2010.

Now

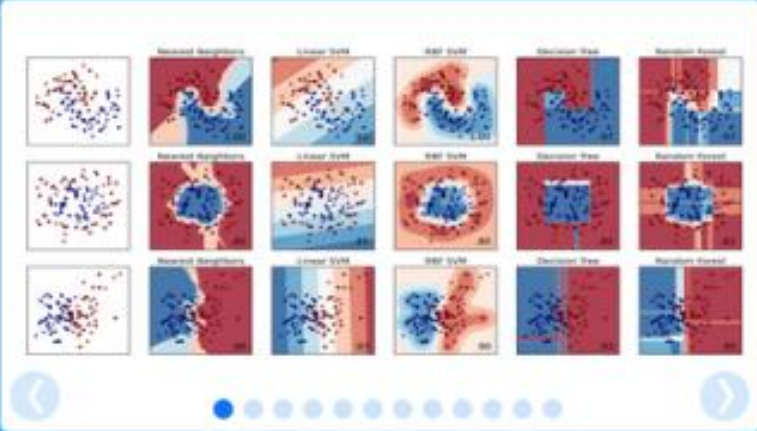
- > 30 active contributors
- Paid sponsorship from INRIA, Google, Tinyclues and the Python Software Foundation.

SCIKIT-LEARN Homepage

[Home](#) [Installation](#) [Documentation](#) [Examples](#)

Google™ Custom Search x

Fork me on GitHub



scikit-learn

Machine Learning in Python

- Simple and efficient tools for data mining and data analysis
- Accessible to everybody, and reusable in various contexts
- Built on NumPy, SciPy, and matplotlib
- Open source, commercially usable - BSD license

Classification

Identifying to which set of categories a new observation belong to.

Applications: Spam detection, Image recognition.

Algorithms: *SVM, nearest neighbors, random forest, ...* — Examples

Regression

Predicting a continuous value for a new example.

Applications: Drug response, Stock prices.

Algorithms: *SVR, ridge regression, Lasso, ...* — Examples

Clustering

Automatic grouping of similar objects into sets.

Applications: Customer segmentation, Grouping experiment outcomes

Algorithms: *k-Means, spectral clustering, mean-shift, ...* — Examples

Dimensionality reduction

Reducing the number of random variables to consider.

Applications: Visualization, Increased efficiency

Algorithms: *PCA, Isomap, non-negative matrix factorization.* — Examples

Model selection

Comparing, validating and choosing parameters and models.

Goal: Improved accuracy via parameter tuning

Modules: *grid search, cross validation, metrics.* — Examples

Preprocessing

Feature extraction and normalization.

Application: Transforming input data such as text for use with machine learning algorithms.

Modules: *preprocessing, feature extraction.* — Examples

Install and Use Scikit-learn

- Install
 - `pip install scikit-learn`
- Check the installation
 - `python -m pip show scikit-learn` # to see which version and where scikit-learn is installed
 - `python -m pip freeze` # to see all packages installed
 - `python -c "import sklearn; sklearn.show_versions()"`