

Introduction to sklearn

Introduction to Machine Learning

Sophie Robert

L3 MIASHS | Semestre 2

2023-2024

Outline

Introduction
to sklearn

Sophie Robert

Introduction

Standard
Workflow: Fit,
Train, Predict

Available
models

1 Introduction

2 Standard Workflow: Fit, Train, Predict

3 Available models

Introduction
to sklearn

Sophie Robert

Introduction

Standard
Workflow: Fit,
Train, Predict

Available
models

Introduction

Scikit-learn

Introduction
to sklearn

Sophie Robert

Introduction

Standard
Workflow: Fit,
Train, Predict

Available
models

Sklearn

Scikit-learn, or sklearn, is a **must-know** machine learning library in Python, which provides a wide range of tools for various machine learning tasks, including classification, regression, clustering, and more.

Sklearn

Scikit-learn, or sklearn, is a **must-know** machine learning library in Python, which provides a wide range of tools for various machine learning tasks, including classification, regression, clustering, and more.

It provides:

- An easy to use standard API.
- Most Machine Learning algorithms.

Scikit-learn

Introduction
to sklearn

Sophie Robert

Introduction

Standard
Workflow: Fit,
Train, Predict

Available
models

- Born out of the Google Summer of Code project in 2007.
- Developed and maintained by a diverse community of contributors worldwide.
- Actively developed on GitHub, with continuous contributions from researchers, developers, and data scientists.

Why use scikit-learn ?

- 1 **Extensive documentation:** Comprehensive guides and examples for effective usage.

Introduction
to sklearn

Sophie Robert

Introduction

Standard
Workflow: Fit,
Train, Predict

Available
models

Why use scikit-learn ?

Introduction
to sklearn

Sophie Robert

Introduction

Standard
Workflow: Fit,
Train, Predict

Available
models

- 1 **Extensive documentation:** Comprehensive guides and examples for effective usage.
- 2 **Wide range of algorithms:** Supports various ML tasks, including classification, regression, clustering, and dimensionality reduction.

Why use scikit-learn ?

Introduction
to sklearn

Sophie Robert

Introduction

Standard
Workflow: Fit,
Train, Predict

Available
models

- 1 **Extensive documentation:** Comprehensive guides and examples for effective usage.
- 2 **Wide range of algorithms:** Supports various ML tasks, including classification, regression, clustering, and dimensionality reduction.
- 3 **Consistent API:** Smooth transition between different algorithms with a consistent workflow.

Why use scikit-learn ?

Introduction
to sklearn

Sophie Robert

Introduction

Standard
Workflow: Fit,
Train, Predict

Available
models

- 1 **Extensive documentation:** Comprehensive guides and examples for effective usage.
- 2 **Wide range of algorithms:** Supports various ML tasks, including classification, regression, clustering, and dimensionality reduction.
- 3 **Consistent API:** Smooth transition between different algorithms with a consistent workflow.
- 4 **Active community:** Strong community of developers, researchers, and data scientists contributing regularly so it stays on the top of the art.

Why use scikit-learn ?

Introduction
to sklearn

Sophie Robert

Introduction

Standard
Workflow: Fit,
Train, Predict

Available
models

- 1 **Extensive documentation:** Comprehensive guides and examples for effective usage.
- 2 **Wide range of algorithms:** Supports various ML tasks, including classification, regression, clustering, and dimensionality reduction.
- 3 **Consistent API:** Smooth transition between different algorithms with a consistent workflow.
- 4 **Active community:** Strong community of developers, researchers, and data scientists contributing regularly so it stays on the top of the art.
- 5 **Integration with other libraries:** Seamless integration with NumPy, SciPy, and matplotlib (pyData ecosystem).

Why use scikit-learn ?

Introduction
to sklearn

Sophie Robert

Introduction

Standard
Workflow: Fit,
Train, Predict

Available
models

- 1 **Extensive documentation:** Comprehensive guides and examples for effective usage.
- 2 **Wide range of algorithms:** Supports various ML tasks, including classification, regression, clustering, and dimensionality reduction.
- 3 **Consistent API:** Smooth transition between different algorithms with a consistent workflow.
- 4 **Active community:** Strong community of developers, researchers, and data scientists contributing regularly so it stays on the top of the art.
- 5 **Integration with other libraries:** Seamless integration with NumPy, SciPy, and matplotlib (pyData ecosystem).
- 6 **Performance optimization:** Optimized C and Cython implementations for efficient computations.

Why use scikit-learn ?

Introduction
to sklearn

Sophie Robert

Introduction

Standard
Workflow: Fit,
Train, Predict

Available
models

- 1 All usual data science steps are included:** Built-in support for cross-validation and model evaluation, feature selection and data preprocessing, ...

Why use scikit-learn ?

Introduction
to sklearn

Sophie Robert

Introduction

Standard
Workflow: Fit,
Train, Predict

Available
models

- 1 All usual data science steps are included:** Built-in support for cross-validation and model evaluation, feature selection and data preprocessing, ...
- 2 Open Source:** Freely available, modifiable, and encourages collaboration and innovation.

Introduction
to sklearn

Sophie Robert

Introduction

Standard
Workflow: Fit,
Train, Predict

Available
models

Standard Workflow: Fit, Train, Predict

Standard Workflow: Create, Train, Evaluate, Predict

Introduction
to sklearn

Sophie Robert

Introduction

Standard
Workflow: Fit,
Train, Predict

Available
models

The typical workflow in scikit-learn involves three steps:

- **Constructor:** Initialize the model with hyperparameters.
- **Fit:** Training the model on the training data.
 - Involves adjusting the model parameters using the provided training data.
- **Evaluate:** Evaluating the model's performance on the training set.
 - Assess how well the model has learned from the training data.
- **Predict:** Applying the trained model to make predictions on new data.
 - Use the learned patterns to predict outcomes for unseen instances.

Example workflow

Introduction
to sklearn

Sophie Robert

Introduction

Standard
Workflow: Fit,
Train, Predict

Available
models

```
# Create the model
from sklearn.neighbors import KNeighborsClassifier
model = KNeighborsClassifier(n_neighbors=3)

# Fit the model
model.fit(X_train, y_train)

# Train: Evaluate on the training set
training_accuracy = model.score(X_train, y_train)

# Predict on new data
predictions = model.predict(X_test)
```

Introduction
to sklearn

Sophie Robert

Introduction

Standard
Workflow: Fit,
Train, Predict

Available
models

Available models

Scikit-Learn Models - Linear Models

Introduction
to sklearn

Sophie Robert

Introduction

Standard
Workflow: Fit,
Train, Predict

Available
models

Scikit-learn provides a comprehensive collection of machine learning models :

Linear Models

- `LinearRegression`
- `Ridge`
- `Lasso`
- `ElasticNet`
- `LogisticRegression`
- `SGDClassifier`

Scikit-Learn Models - KNN

Introduction
to sklearn

Sophie Robert

Introduction

Standard
Workflow: Fit,
Train, Predict

Available
models

k-Nearest Neighbors

- `KNeighborsClassifier`
- `KNeighborsRegressor`

Scikit-Learn Models - SVM

Introduction
to sklearn

Sophie Robert

Introduction

Standard
Workflow: Fit,
Train, Predict

Available
models

Support Vector Machines

- SVC (Support Vector Classification)
- NuSVC
- SVR (Support Vector Regression)
- NuSVR

Scikit-Learn Models - Tree-based Models

Introduction
to sklearn

Sophie Robert

Introduction

Standard
Workflow: Fit,
Train, Predict

Available
models

Tree-based Models

- `DecisionTreeClassifier`
- `DecisionTreeRegressor`
- `RandomForestClassifier`
- `RandomForestRegressor`

Scikit-Learn Models - Clustering

Introduction
to sklearn

Sophie Robert

Introduction

Standard
Workflow: Fit,
Train, Predict

Available
models

Clustering

- KMeans
- MiniBatchKMeans
- DBSCAN
- AgglomerativeClustering

Scikit-Learn Models - Dimensionality Reduction

Introduction
to sklearn

Sophie Robert

Introduction

Standard
Workflow: Fit,
Train, Predict

Available
models

Dimensionality Reduction

- PCA (Principal Component Analysis)
- TruncatedSVD

Scikit-Learn Models - Ensemble Methods

Introduction
to sklearn

Sophie Robert

Introduction

Standard
Workflow: Fit,
Train, Predict

Available
models

Ensemble Methods

- VotingClassifier
- VotingRegressor
- BaggingClassifier
- BaggingRegressor