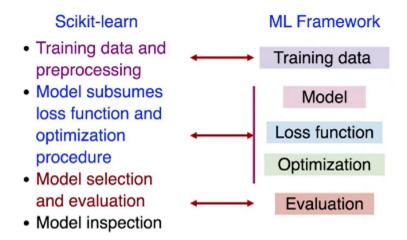
28 January 2023

11:30 PM

sklearn APIs are organized on the lines of our ML framework.



sklearn APIs are well designed with the following principles:

- Consistency: All APIs share a simple and consistent interface.
- Inspection: The learnable parameters as well as hyperparameters of all estimator's are accessible directly via public instance variables.
- Nonproliferation of classes: Datasets are represented as Numpy arrays or Scipy sparse matrix instead of custom designed classes.
- Composition: Existing building blocks are reduced as much as possible.
- Sensible defaults values are used for parameters that enables quick baseline building.

# Types of sklearn objects

#### Transformers

- · transforms dataset
- transform() for transforming dataset.
- fit() learns parameters.
- fit\_transform() fits parameters and transform() the dataset.

#### **Estimators**

- Estimates model parameters based on training data and hyper parameters.
- fit() method

#### **Predictors**

- Makes prediction on dataset
- predict() method that takes dataset as an input and returns predictions.
- score() method to measure quality of predictions.

Data Preprocessing

Training

Inference

# Data API

Provides functionality for loading, generating and preprocessing the training and test data.

Module	Functionality
sklearn.datasets	Loading datasets - custom as well as popular reference dataset.
sklearn.preprocessing	Scaling, centering, normalization and binarization methods
sklearn.impute	Filling missing values
sklearn.feature_selection	Implements feature selection algorithms
sklearn.feature_extraction	Implements feature extraction from raw data.

# Model API

Implements supervised and unsupervised models

### Regression

 sklearn.linear\_model (linear, ridge, lasso models)

### Classification

- sklearn.linear\_model
- sklearn.svm
- sklearn.trees

### Model API

Implements supervised and unsupervised models

#### Regression

 sklearn.linear\_model (linear, ridge, lasso models)
 sklearn.trees

#### Classification

- sklearn.linear model
- sklearn.svm
- sklearn.trees
- sklearn.neighbors
- sklearn.naive\_bayes
- sklearn.multiclass

sklearn.multioutput implements multi-output classification and regression.

sklearn.cluster implements many popular clustering algorithms

## Model selection API

sklearn.model\_selection implements various model selection strategies like cross-validation, tuning hyper-parameters and plotting learning curves.

# Model inspection API

sklearn.model\_inspection includes tools for model
inspection.