

module sklearn → i) various ML algos implemented

ii) sample datasets

(already well cleaned)



Boston Dataset : House pricing dataset

① Loading Data

from sklearn import datasets

boston = datasets.load_boston() → i) DESCR



type : sklearn.utils.Bunch

ii) data

iii) feature_names

iv) target

X = boston.data

Y = boston.target

import pandas as pd.

df = pd.DataFrame(X)

print(boston.feature_names)

df.columns = boston.feature_names

*) Using through the description provided with the dataset may help gain important insights.

df.describe()

boston.DESCR

② Splitting data into train and test sets

from sklearn import model-selection

X_train, X_test, Y_train, Y_test = model-selection.

train-test-split (X, Y)
↓
numpy arrays

③ Getting the algorithm to fit data on.

```
from sklearn.linear_model import LinearRegression
```

```
alg1 = LinearRegression()
```

```
alg1.fit(X_train, Y_train)
```

④ Using the fitted algorithm to get prediction

```
Y_pred = alg1.predict(X_test)
```

⑤ Comparing Y_{pre} and Y_{test}

↓

Method 1 : plotting Y_{pre} and Y_{test} . The closer the values are to the line $y = x$, the better the predictions are.

↓

```
import matplotlib.pyplot as plt  
plt.scatter(Y_pred, Y_test)
```

↓ better

```
plt.scatter(Y_test, Y_pred)
```

```
plt.axis([0, 40, 0, 40])
```

```
plt.grid()
```

```
plt.show()
```