Class2_dataset

January 8, 2025

[1]:

```python
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn import metrics

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

[2]:

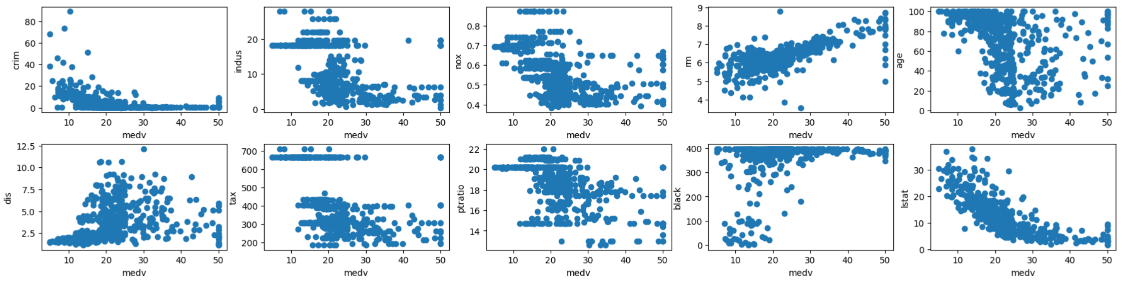
```python
dataset = pd.read_csv(r"./dataset/Boston.csv")
dataset = dataset.drop(columns=["Unnamed: 0"]) # Drop Index
dataset = dataset.drop(columns=["chas", "rad", "zn"]) # Drop Discrete data
label = dataset["medv"]
dataset = dataset.drop(columns=["medv"])
dataset.head()
```

[2]:

```
      crim  indus    nox     rm   age     dis  tax  ptratio   black  lstat
0  0.00632   2.31  0.538  6.575  65.2  4.0900  296     15.3  396.90   4.98
1  0.02731   7.07  0.469  6.421  78.9  4.9671  242     17.8  396.90   9.14
2  0.02729   7.07  0.469  7.185  61.1  4.9671  242     17.8  392.83   4.03
3  0.03237   2.18  0.458  6.998  45.8  6.0622  222     18.7  394.63   2.94
4  0.06905   2.18  0.458  7.147  54.2  6.0622  222     18.7  396.90   5.33
```

[3]:

```python
from utils.plot_tools import multiple_plot
multiple_plot(
    2, 5,
    [(label ,dataset[i]) for i in dataset.columns],
    figsize=(22, 5),
    show_axis=True,
    func_1="scatter"
    )
```



[4]:

```
x_train, x_test, y_train, y_test = train_test_split(
    dataset["rm"].to_numpy().reshape(-1, 1),
    label,
    shuffle=True,
    test_size=0.2
    )
```

[5]:

```
model = LinearRegression()
model.fit(x_train, y_train)
```

[5]:

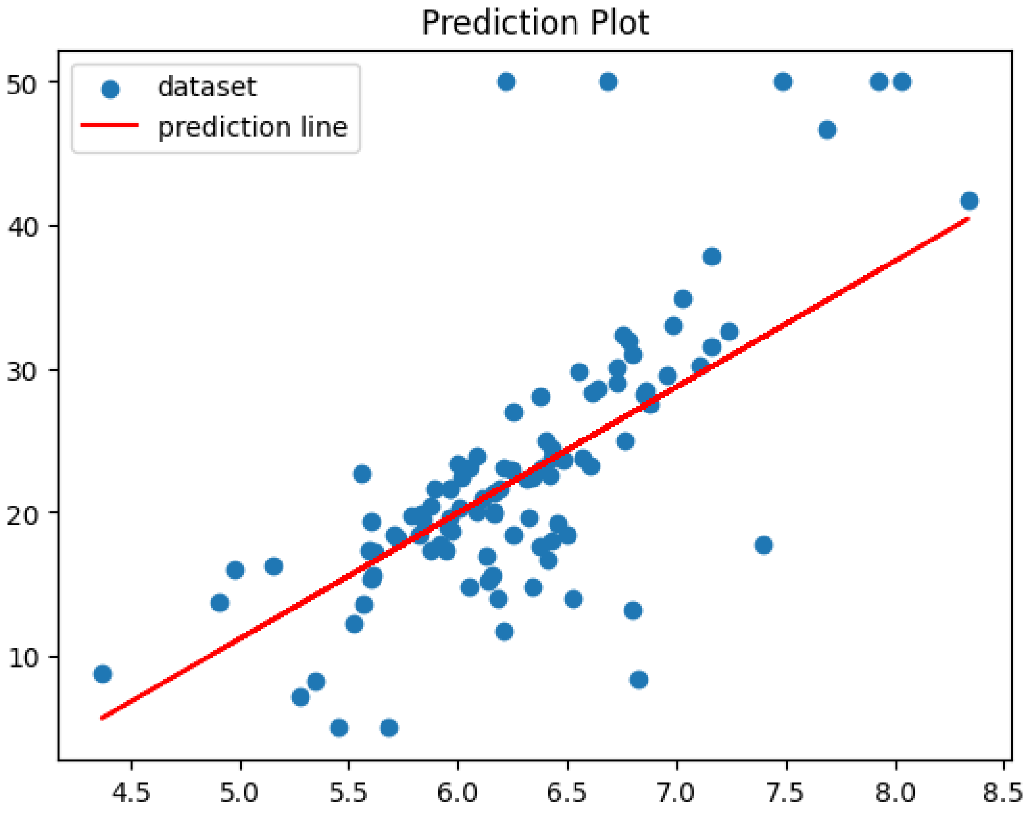
```
LinearRegression()
```

[12]:

```
y_predict = model.predict(x_test)

plt.scatter(x_test, y_test, label="dataset")
plt.plot(x_test, y_predict, c="r", label="prediction line")
plt.legend()
plt.title("Prediction Plot")
plt.figure(figsize=(4, 4))

plt.show()
```



```
<Figure size 400x400 with 0 Axes>
```

[7]:

```
print("mae:", metrics.mean_absolute_error(y_test, y_predict))
print("mse:", metrics.mean_squared_error(y_test, y_predict))
print("rmse:", np.sqrt(metrics.mean_squared_error(y_test, y_predict)))
```

```
mae: 4.280140981435683
mse: 43.07300777821382
rmse: 6.563002954304822
```

[ ]: