線形回帰モデル 実装演習

```
In [ ]:
          from sklearn.datasets import load_boston
          from sklearn.linear_model import LinearRegression
          from sklearn.metrics import mean_squared_error
          import pandas as pd
          import numpy as np
          import matplotlib.pyplot as plt
          import seaborn as sns
In [ ]:
          boston = load_boston()
          boston_df = pd. DataFrame(boston. data, columns = boston. feature_names)
          boston_df['MEDV'] = boston. target
In [ ]:
          boston_df. columns. values
Out[]: array(['CRIM', 'ZN', 'INDUS', 'CHAS', 'NOX', 'RM', 'AGE', 'DIS', 'RAD', 'TAX', 'PTRATIO', 'B', 'LSTAT', 'MEDV'], dtype=object)
In [21]:
          #相関関係の模索
          plt.figure(figsize=(20, 20))
          cm = np. corrcoef(boston_df. values. T)
          sns. set (font_scale=1)
          hm = sns. heatmap(cm,
                             cbar=True,
                             annot=True,
                             square=True,
                             fmt='.2f',
                             annot_kws={'size': 25},
                             yticklabels=boston_df.columns.values,
                             xticklabels=boston_df. columns. values)
          plt. tight_layout()
          plt. show()
```

```
-0.20 0.41 -0.06 0.42 -0.22 0.35 -0.38
                                                   0.63 0.58 0.29 -0.39
                                                                           0.46 -0.39
 1.00
  -0.20 1.00
              -0.53 -0.04 -0.52
                                0.31
                                      -0.57
                                             0.66
                                                  -0.31 -0.31 -0.39 0.18 -0.41 0.36
NZ
        -0.53 1.00
                    0.06
                          0.76
                                -0.39
                                      0.64 -0.71
                                                   0.60 0.72
                                                               0.38
                                                                    -0.36 0.60
                                                                                 -0.48
  0.41
  -0.06 -0.04 0.06
                    1.00
                          0.09
                                0.09 0.09 -0.10 -0.01 -0.04 -0.12 0.05 -0.05 0.18
CHAS
                                            -0.77
                                                  0.61 0.67
                                                               0.19 -0.38
        -0.52
              0.76
                    0.09
                          1.00
                                 -0.30
                                      0.73
                                                                                 -0.43
XON
  -0.22 0.31
              -0.39 0.09 -0.30 1.00
                                      -0.24 0.21 -0.21 -0.29 -0.36 0.13 -0.61
                                                                                 0.70
ĕ
        -0.57
              0.64
                    0.09
                          0.73
                                -0.24 1.00 -0.75
                                                               0.26 -0.27
                                                                           0.60
                                                                                 -0.38
  0.35
AGE
              -0.71 -0.10 -0.77
                                0.21
                                      -0.75 1.00
                                                  -0.49 -0.53 -0.23 0.29
                                                                          -0.50 0.25
  -0.38
        0.66
Si
              0.60 -0.01 0.61
                                -0.21 0.46 -0.49 1.00 0.91
  0.63
        -0.31
                                                               0.46
                                                                    -0.44 0.49 -0.38
3AD
        -0.31
              0.72
                    -0.04 0.67
                                 -0.29
                                      0.51 -0.53
                                                   0.91 1.00
                                                               0.46 -0.44
                                                                           0.54 -0.47
TAX
  0.29
        -0.39
              0.38
                    -0.12 0.19
                                -0.36
                                      0.26
                                            -0.23
                                                   0.46 0.46
                                                               1.00
                                                                     -0.18 0.37
                                                                                -0.51
  -0.39
       0.18 -0.36 0.05 -0.38 0.13 -0.27
                                            0.29
                                                  -0.44 -0.44 -0.18 1.00
                                                                           -0.37 0.33
                                -0.61
                                                   0.49 0.54 0.37
                                                                           1.00
                                                                                 -0.74
        -0.41
              0.60 -0.05
                                      0.60
                                            -0.50
                                                                    -0.37
  -0.39
        0.36
              -0.48
                    0.18
                          -0.43
                                0.70
                                      -0.38
                                            0.25
                                                  -0.38 -0.47 -0.51
                                                                     0.33
                                                                           -0.74
                                                                                 1.00
                                                                            LSTAT
```

```
In [22]: #taxとindusの線形回帰モデルの作成
Ir = LinearRegression()

X = boston_df[['INDUS']]. values
Y = boston_df['TAX']. values

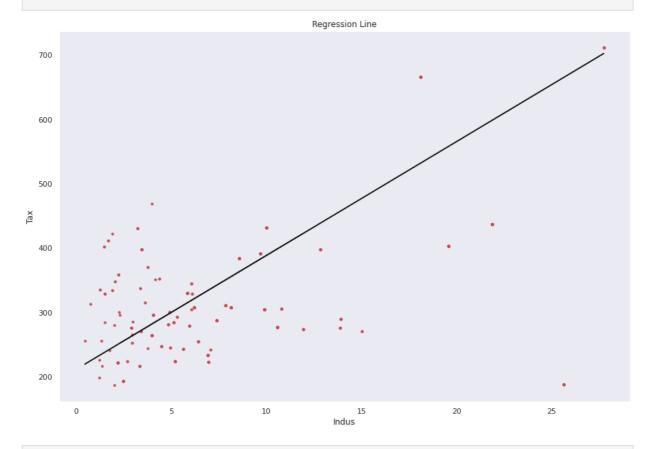
Ir. fit(X, Y)
```

Out[22]: LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)

```
In [23]: #plot
Y_pred = Ir. predict(X)
fig = plt. figure(figsize=(15, 10))
plt. plot(X, Y, 'r.')
plt. plot(X, Y_pred, 'k-')

plt. title('Regression Line')
plt. xlabel('Indus')
plt. ylabel('Tax')
plt. grid()
```





In [24]:

#評価
print('MSE: ', mean_squared_error(Y, Y_pred))

MSE: 13621.64866409399

In [25]:

```
#課題
X1 = boston_df[['CRIM','RM']]. values
Y1 = boston_df['MEDV']. values

Ir. fit(X1, Y1)
print(Ir. predict([[0.3, 4]]))
```

[4. 24007956]

ボストンデータセットから相関の強いデータを探し、線形回帰を試みた。

MSEで評価を行ったが、単体の値だけでは意味が薄く、パラメトリックスタディなどをして比較する際に求めるべきだったと感じた。

また、相関が強いからと言って必ずしも線形回帰が適切であるとは限らず、外れ値に関しての 工学的な判断も必要であると感じた。