

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()

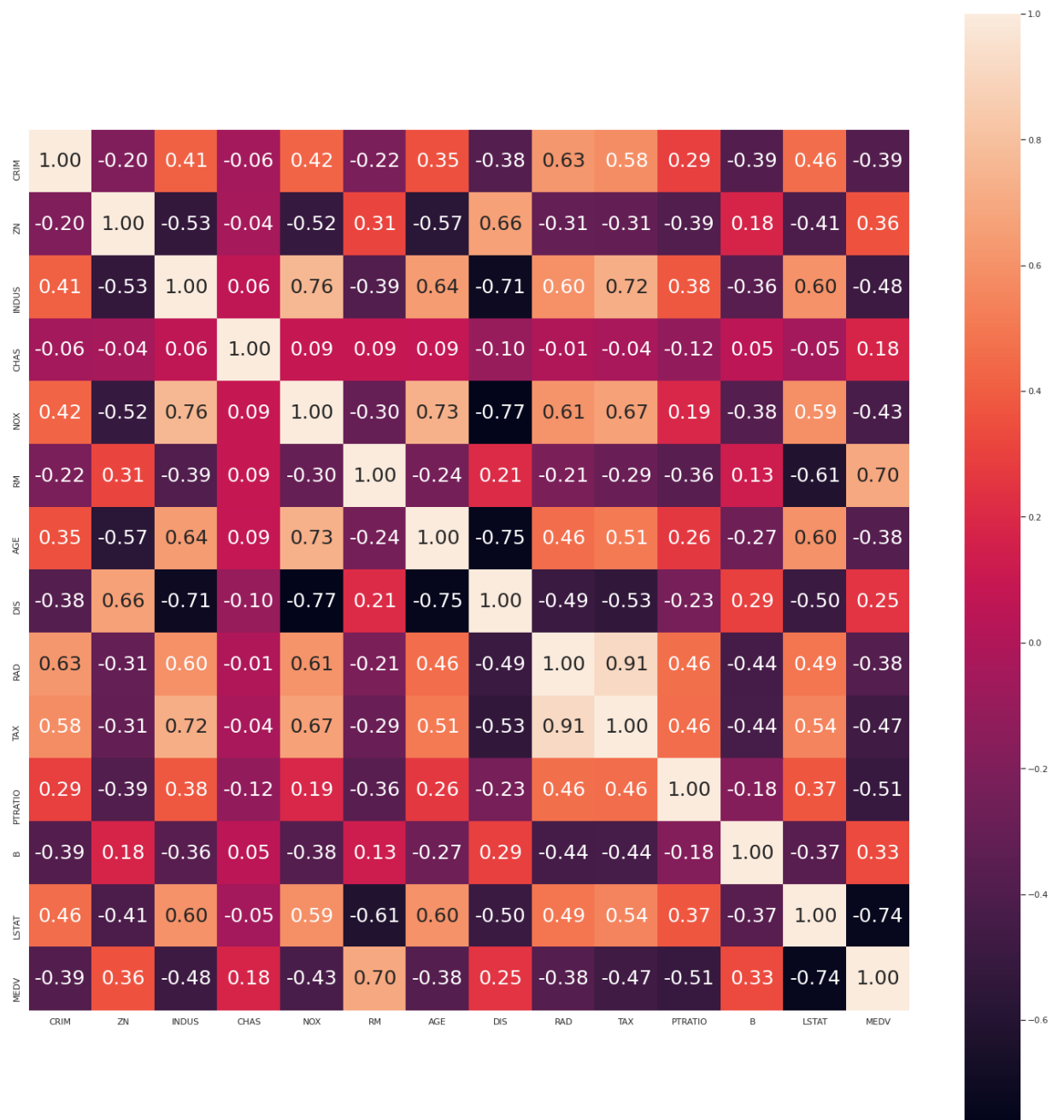
        boston_df = pd.DataFrame(boston.data, columns = boston.feature_names)
        boston_df['MEDV'] = boston.target
```

```
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()
```



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

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

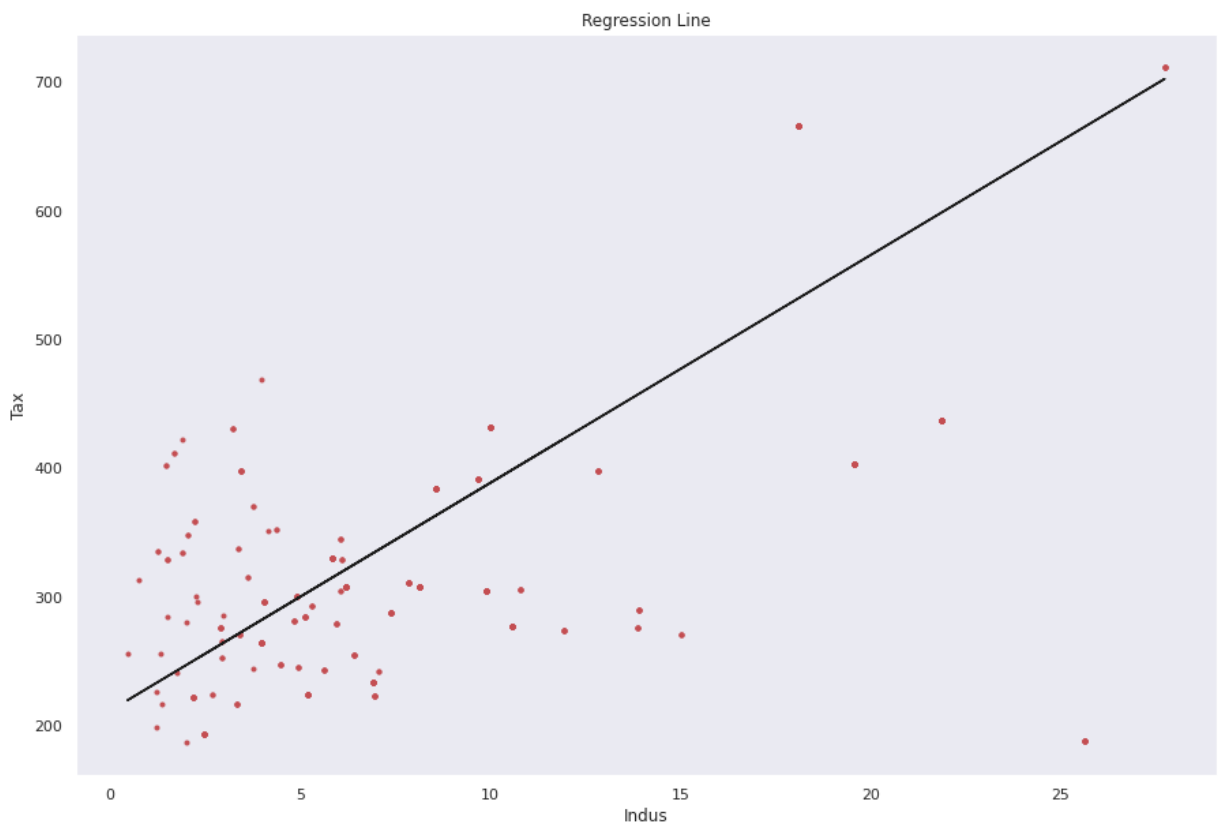
lr.fit(X, Y)
```

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

```
In [23]: #plot
Y_pred = lr.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()
```

```
plt.show()
```



```
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

lr.fit(X1, Y1)
print(lr.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、外れ値についての工学的な判断も必要であると感じた。