



GPT-4



OK

Pythonでscikit-learnのcalifornia housing datasetを使って、カリフォルニアの住宅価格の単回帰分析をしてもらえますか。目的変数はpriceで説明変数は、MedInc (median income、「所得」の中央値) にしてもらって構いません。単回帰直線のplotと、相関係数をお願いします。



以下にPythonとscikit-learnを用いた単回帰分析のコードを記載します。ここでは、目的変数としてカリフォルニアの住宅価格を、説明変数として「所得」の中央値(MedInc)を使用しています。

python

Copy code

```
import matplotlib.pyplot as plt
import numpy as np
from sklearn import datasets
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from scipy.stats import pearsonr

# データのロード
california = datasets.fetch_california_housing()

# MedInc (median income) を説明変数に、target (price) を目的変数にセット
X = california.data[:, np.newaxis, 0] # MedInc
y = california.target

# データの分割
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# モデルの作成と学習
model = LinearRegression()
model.fit(X_train, y_train)

# プロット
plt.scatter(X_test, y_test, color='black')
plt.plot(X_test, model.predict(X_test), color='blue', linewidth=3)
plt.xlabel('Median Income')
plt.ylabel('House Price')
plt.title('Regression Line for California Housing Prices')
plt.show()

# 相関係数の計算
correlation, _ = pearsonr(X_train.reshape(-1), y_train)
print(f'Correlation coefficient: {correlation}')
```



Send a message

ChatGPT may produce inaccurate information about people, places, or facts. [ChatGPT August 3 Version](#)