第七周作业

2024年10月14日　　18:19

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import MinMaxScaler
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error
file_path = r"C:\Users\ASUS\Documents\WeChat Files\wxid_gtfqqzwcu89o22\FileStorage
\File\2024-10\bike.csv"
data = pd.read_csv(file_path)
data = data.drop("id", axis=1)
data = data[data["city"] == 1]
data = data.drop("city", axis=1)
data["hour"] = data["hour"].apply(lambda x: 1 if 6 <= x <= 18 else 0)
y = data["y"].values.reshape(-1, 1)
data = data.drop("y", axis=1)
X = data.values
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
shuffle=True)
scaler = MinMaxScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)
y_train_scaled = scaler.fit_transform(y_train)
y_test_scaled = scaler.transform(y_test)
model = LinearRegression()
model.fit(X_train_scaled, y_train_scaled)
y_pred = model.predict(X_test_scaled)
mse = mean_squared_error(y_test_scaled, y_pred)
rmse = np.sqrt(mse)
print("均方根误差:", rmse)
```

```
PS C:\Users\ASUS> python -u "f:\programming\shujukexuedaolun\10_21.py"
均方根误差: 0.16412359338536198
```