306281422_stats101c_hw4

November 27, 2024

[ ]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.linear_model import Lasso
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
```

[ ]:
```
df = pd.read_csv('DF_LASSO')
X = df.iloc[:,1:7]
y = df.iloc[:,7]

scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)
```

[ ]:
```
X_train, X_test, y_train, y_test = train_test_split(X_scaled, y, test_size=0.3,␣
↪random_state=42)
```

[ ]:
```
lasso = Lasso(alpha=0.1)
lasso.fit(X_train, y_train)
coeff_lasso = pd.DataFrame({
    'Predictor': ['X_1', 'X_2', 'X_3', 'X_4', 'X_5', 'X_6'],
    'Coefficient': lasso.coef_
})
print(coeff_lasso)
```

```
  Predictor  Coefficient
0       X_1     5.582536
1       X_2     5.590277
2       X_3     0.000000
3       X_4     0.045882
4       X_5     5.577791
5       X_6     0.000000
```

From the result above, we can see that predictor X_3, X_4, X_6 achieved sparsity. These 3 predictors should not be included in the regression model.