2013-11086 김찬규 과제 5

In [1]:

```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression, Lasso
```

In [2]:

```python
df = pd.read_csv("data/weather.csv")
df.head()
```

Out[2]:

	DateNumber	x1	x2	x3	x4	x5	x6	x7	x8	x9	...	x12	x1
0	WTH0001	0.123	0.156	0.173	0.179	0.108	0.118	0.181	0.167	0.149	...	0.190	0.13
1	WTH0002	0.145	0.138	0.173	0.197	0.131	0.183	0.194	0.191	0.167	...	0.200	0.19
2	WTH0003	0.190	0.124	0.110	0.102	0.126	0.104	0.189	0.180	0.146	...	0.108	0.14
3	WTH0004	0.168	0.200	0.169	0.162	0.190	0.152	0.164	0.123	0.116	...	0.166	0.10
4	WTH0005	0.178	0.126	0.159	0.165	0.121	0.129	0.112	0.189	0.197	...	0.177	0.11

5 rows × 22 columns

In [3]:

```python
df_x = df[[f"x{i}" for i in range(1, 21)]]
df_y = df["y"]
```

In [4]:

```python
x_train, x_test, y_train, y_test = train_test_split(df_x, df_y, random_state=0, test_size=0.3)
```

Linear Regression

In [5]:

```python
reg = LinearRegression()
```

In [6]:

```python
reg.fit(x_train, y_train)
```

Out[6]:

```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)
```

In [7]:

```
y_pred = reg.predict(x_test)
```

In [8]:

```
mape = np.mean(np.abs(y_test - y_pred) / y_test)
print(mape)
```

0.13275224292736143

(1) MAPE = 0.1328

Lasso Regression

In [9]:

```
reg_Lasso = Lasso(alpha=0.01)
```

In [10]:

```
reg_Lasso.fit(x_train, y_train)
```

Out[10]:

```
Lasso(alpha=0.01, copy_X=True, fit_intercept=True, max_iter=1000,
      normalize=False, positive=False, precompute=False, random_stat
e=None,
      selection='cyclic', tol=0.0001, warm_start=False)
```

In [11]:

```
y_pred_Lasso = reg_Lasso.predict(x_test)
```

In [12]:

```
np.sum(reg_Lasso.coef_ == 0)
```

Out[12]:

15

(2) 총 15개의 변수 탈락

In [ ]: