

17

21회 기출 분석 (3)

[수업 목표]

00:00

16 21회 기출분석 (3)

2. 통계분석 (50점)
 2. 연속형 독립변수 여러개의 소규모 데이터. (총 29점) 변수명은 순서대로 $x_1 \sim x_{10}$, 의미는 없는 데이터
 2-1. 데이터 8:2로 분할하고 선형회귀 적용하시오. 결정계수와 $rmse$ 구하시오

2-2. 데이터 8:2로 분할하고 릿지 회귀 적용하시오.
 α 값을 0부터 1까지 0.1단위로 모두 탐색해서 결정계수가 가장 높을때의 알파를 찾고, 해당 알파로 다시 모델을 학습해서 결정계수와 $rmse$ 를 계산

2-3. 데이터 8:2로 분할하고 라쏘 회귀 적용하시오.
 α 값을 0부터 1까지 0.1단위로 모두 탐색해서 결정계수가 가장 높을때의 알파를 찾고, 해당 알파로 다시 모델을 학습해서 결정계수와 $rmse$ 를 계산

```

In [1]: import pandas as pd
import numpy as np
import mglearn
X, y = mglearn.datasets.load_extended_boston()
# 훈련, 테스트 셋 분리
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=0)

In [3]: from sklearn.model_selection import GridSearchCV
from sklearn.metrics import mean_squared_error
from sklearn.linear_model import LinearRegression
from sklearn.linear_model import Ridge
from sklearn.linear_model import Lasso
  
```

이번 시간에는 통계분석 파트를 풀어보겠습니다

[선형 회귀]

01:37

2. 통계분석 (50점)
 2. 연속형 독립변수 여러개의 소규모 데이터. (총 29점) 변수명은 순서대로 $x_1 \sim x_{10}$, 의미는 없는 데이터
 2-1. 데이터 8:2로 분할하고 선형회귀 적용하시오. 결정계수와 $rmse$ 구하시오

2-2. 데이터 8:2로 분할하고 릿지 회귀 적용하시오.
 α 값을 0부터 1까지 0.1단위로 모두 탐색해서 결정계수가 가장 높을때의 알파를 찾고, 해당 알파로 다시 모델을 학습해서 결정계수와 $rmse$ 를 계산

2-3. 데이터 8:2로 분할하고 라쏘 회귀 적용하시오.
 α 값을 0부터 1까지 0.1단위로 모두 탐색해서 결정계수가 가장 높을때의 알파를 찾고, 해당 알파로 다시 모델을 학습해서 결정계수와 $rmse$ 를 계산

```

In [1]: import pandas as pd
import numpy as np
import mglearn
X, y = mglearn.datasets.load_extended_boston()
# 훈련, 테스트 셋 분리
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=0)

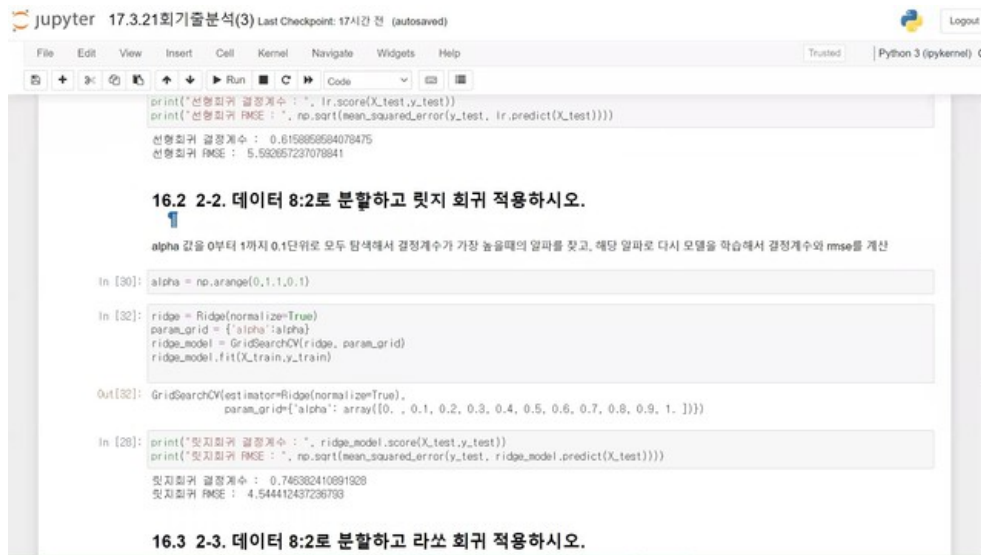
In [3]: from sklearn.model_selection import GridSearchCV
from sklearn.metrics import mean_squared_error
from sklearn.linear_model import LinearRegression
from sklearn.linear_model import Ridge
from sklearn.linear_model import Lasso
  
```

16.1 2-1. 데이터 8:2로 분할하고 선형회귀 적용하시오. 결정계수와 $rmse$ 구하시오

[릿지, 라쏘 회귀분석]

02:43

#매개변수를 조정할 수 있는가를 물어보는 문제였습니다.



The image shows a Jupyter Notebook interface with the following content:

Header: jupyter 17.3.21회기출분석(3) Last Checkpoint: 17시간 전 (autosaved) Python 3 (ipykernel)

Code cell 1:

```
print("선형회귀 결정계수 : ", lr.score(X_test, y_test))
print("선형회귀 RMSE : ", np.sqrt(mean_squared_error(y_test, lr.predict(X_test))))
```

Output 1:

```
선형회귀 결정계수 : 0.6158858584078475
선형회귀 RMSE : 5.592857237078841
```

Section 16.2 2-2. 데이터 8:2로 분할하고 릿지 회귀 적용하시오.

alpha 값을 0부터 1까지 0.1단위로 모두 탐색해서 결정계수가 가장 높을때의 알파를 찾고, 해당 알파로 다시 모델을 학습해서 결정계수와 rmse를 계산

Code cell 2:

```
In [30]: alpha = np.arange(0,1,0.1)

In [32]: ridge = Ridge(normalize=True)
param_grid = {'alpha':alpha}
ridge_model = GridSearchCV(ridge, param_grid)
ridge_model.fit(X_train, y_train)

Out[32]: GridSearchCV(estimator=Ridge(normalize=True),
    param_grid={'alpha': array([0. , 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1. ])}))

In [28]: print("릿지회귀 결정계수 : ", ridge_model.score(X_test, y_test))
print("릿지회귀 RMSE : ", np.sqrt(mean_squared_error(y_test, ridge_model.predict(X_test))))
```

Output 2:

```
릿지회귀 결정계수 : 0.746382410691928
릿지회귀 RMSE : 4.544412437236793
```

Section 16.3 2-3. 데이터 8:2로 분할하고 라쏘 회귀 적용하시오.

[다음 수업 예고]

다음 시간에는 통계분석 파트2 를 풀어보겠습니다