

handson-m12-02.ipynb의 사본

array([45893.36082829, 49774.46796717])

z-분포를 이용한 신뢰구간은 다음과 같다.

```
zscore = stats.norm.ppf((1 + confidence) / 2)
zmargin = zscore * squared_errors.std(ddof=1) / np.sqrt(m)
np.sort(mean - zmargin), np.sort(mean + zmargin)
```

**NameError** Traceback (most recent call last):  
File "python-input-107-23bfeab5408a" in <cell line: 2>():  
1 zscore = stats.norm.ppf((1 + confidence) / 2)  
----> 2 zmargin = zscore \* squared\_errors.std(ddof=1) / np.sqrt(m)  
3 np.sort(mean - zmargin), np.sort(mean + zmargin)  
  
**NameError: name 'm' is not defined**

STACK OVERFLOW 검색

부록: 추가 내용

전처리와 예측을 포함한 전체 파이프라인

```
[108] full_pipeline_with_predictor = Pipeline([
    ("preparation", full_pipeline),
    ("linear", LinearRegression())
])

full_pipeline_with_predictor.fit(housing, housing_labels)
full_pipeline_with_predictor.predict(some_data)

array([ 85657.90192014, 305492.60737488, 152056.46122456, 186095.70946094,
        244550.67966089])

joblib를 사용한 모델 저장



```
[109] my_model = full_pipeline_with_predictor

[110] import joblib
```


```

과제 실습 중 오류가 발생하였습니다. “m”에 대한 정의가 되어있지 않아서 발생한 오류였습니다. 이어서 “mean”도 정의되어 있지 않아 두 변수에 대한 정의를 선행 후 코드를 진행하였습니다.

handson-m12-02.ipynb의 사본

colab.research.google.com/drive/1Ceux7TMAg5zc66sonloP7Mb5On-J0xyH#scr...

handson-m12-02.ipynb의 사본

파일 수정 보기 삽입 런타임 도구 도움말 모든 변경사항이 저장됨

+ 코드 + 텍스트

z-분포를 이용한 신뢰구간은 다음과 같다.

```
[139] m = len(squared_errors)
      mean = squared_errors.mean()

      zscore = stats.norm.ppf((1 + confidence) / 2)
      zmargin = zscore * squared_errors.std(ddof=1) / np.sqrt(m)
      np.sqrt(mean - zmargin), np.sqrt(mean + zmargin)

(45993.9540110131, 49773.921030650374)
```

부록: 추가 내용

전처리와 예측을 포함한 전체 파이프라인

```
[108] full_pipeline_with_predictor = Pipeline([
      ("preparation", full_pipeline),
      ("linear", LinearRegression())
  ])

full_pipeline_with_predictor.fit(housing, housing_labels)
full_pipeline_with_predictor.predict(some_data)

array([ 85657.90192014, 305492.60737488, 152056.46122456, 186095.70946094,
       244550.67966089])
```

joblib를 사용한 모델 저장

```
[109] my_model = full_pipeline_with_predictor

[110] import joblib
      joblib.dump(my_model, "my_model.pkl") # DIFF
      #...
      my_model_loaded = joblib.load("my_model.pkl") # DIFF
```

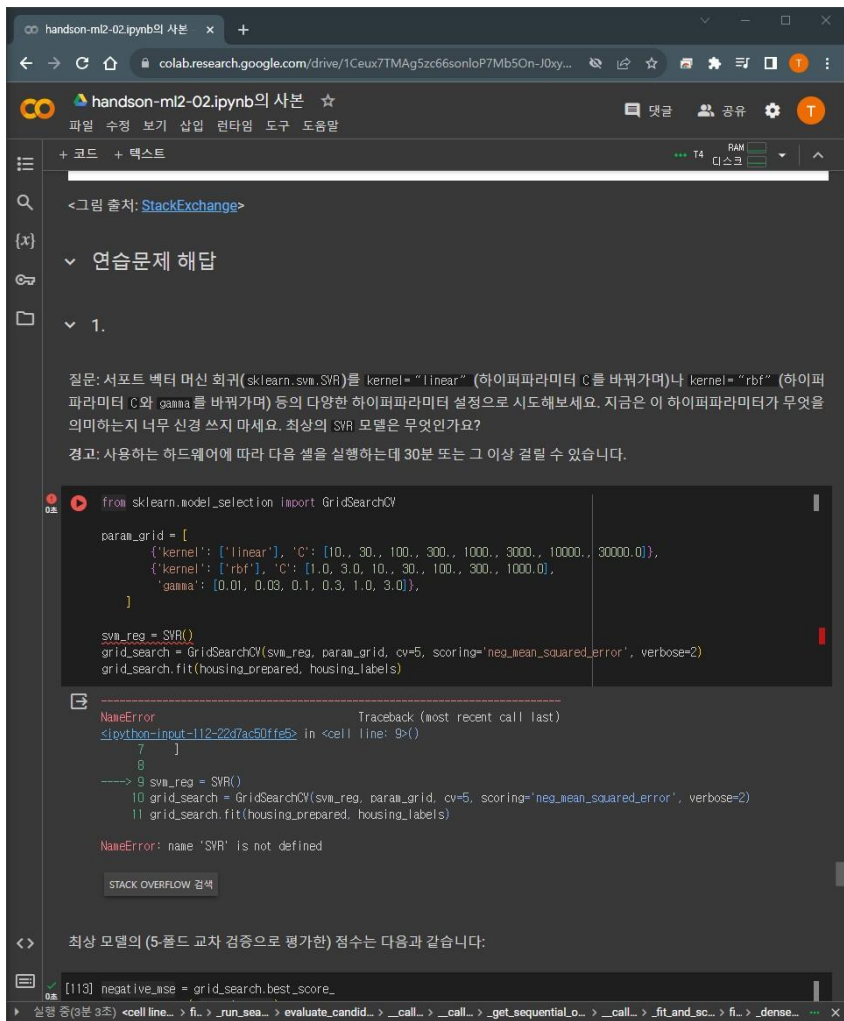
RandomizedSearchCV를 위한 Scipy 분포 함수

랜덤 탐색을 하려면 특정 옵션 변수값들을 무작위로 선택해주는 확률분포함수를 지정해야 한다. 앞서 랜덤 탐색을 이용하여

0초 오후 12:01에 완료됨

다음과 같이

`m = len(squared_errors)`로,  
`mean = squared_errors.mean()`으로  
정의하였습니다.



이번에는 ‘SVR’이 정의되지 않아있었습니다. 아래 모든 연습문제들도 마찬가지였습니다.

```
[141]: from sklearn.model_selection import GridSearchCV
from sklearn.svm import SVR

param_grid = [
    ('kernel': ['linear'], 'C': [10., 30., 100., 300., 1000., 3000., 10000., 30000.0]),
    ('kernel': ['rbf'], 'C': [1.0, 3.0, 10., 30., 100., 300., 1000.0],
     'gamma': [0.01, 0.03, 0.1, 0.3, 1.0, 3.0]),
]

svm_reg = SVR()
grid_search = GridSearchCV(svm_reg, param_grid, cv=5, scoring='neg_mean_squared_error', verbose=2)
grid_search.fit(housing_prepared, housing_labels)
```

Fitting 5 folds for each of 50 candidates, totalling 250 fits

CV	END	.....	C=10.0, kernel=linear: total time=	6.1s
CV	END	.....	C=10.0, kernel=linear: total time=	7.4s
CV	END	.....	C=10.0, kernel=linear: total time=	6.4s
CV	END	.....	C=10.0, kernel=linear: total time=	7.5s
CV	END	.....	C=10.0, kernel=linear: total time=	6.4s
CV	END	.....	C=30.0, kernel=linear: total time=	7.2s
CV	END	.....	C=30.0, kernel=linear: total time=	6.9s
CV	END	.....	C=30.0, kernel=linear: total time=	6.8s
CV	END	.....	C=30.0, kernel=linear: total time=	7.1s
CV	END	.....	C=30.0, kernel=linear: total time=	6.4s
CV	END	.....	C=100.0, kernel=linear: total time=	7.4s
CV	END	.....	C=100.0, kernel=linear: total time=	5.9s
CV	END	.....	C=100.0, kernel=linear: total time=	7.3s
CV	END	.....	C=100.0, kernel=linear: total time=	6.0s
CV	END	.....	C=100.0, kernel=linear: total time=	7.3s
CV	END	.....	C=300.0, kernel=linear: total time=	6.2s
CV	END	.....	C=300.0, kernel=linear: total time=	7.2s
CV	END	.....	C=300.0, kernel=linear: total time=	6.0s
CV	END	.....	C=300.0, kernel=linear: total time=	7.3s
CV	END	.....	C=300.0, kernel=linear: total time=	6.1s
CV	END	.....	C=1000.0, kernel=linear: total time=	7.6s
CV	END	.....	C=1000.0, kernel=linear: total time=	6.2s
CV	END	.....	C=1000.0, kernel=linear: total time=	7.4s
CV	END	.....	C=1000.0, kernel=linear: total time=	6.3s

44분 45초    오후 12:55에 완료됨

from sklearn.svm import SVR  
위 코드로 SVR을 호출하여 문제를 해결하였습니다.



+ 코드 + 텍스트

✓ RAM  
다스크

```
scores = scorer(estimator, x_test, y_test)
File ~/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_scorer.py, line 234, in _call__
return self._score(
File ~/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_scorer.py, line 276, in _score
y_pred = method_caller(estimator, "predict", X)
File ~/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_scorer.py, line 73, in _cached_call
return getattr(estimator, method)(*args, **kwargs)
File ~/usr/local/lib/python3.10/dist-packages/sklearn/pipeline.py, line 480, in predict
Xt = transform.transform(Xt)
File ~/usr/local/lib/python3.10/dist-packages/sklearn/utils/_set_output.py, line 140, in wrapped
data_to_wrap = f(self, X, *args, **kwargs)
File ~/usr/local/lib/python3.10/dist-packages/sklearn/compose/_column_transformer.py, line 800, in transform
```

[1] grid\_search\_prep.best\_params\_

```
NameError                                Traceback (most recent call last)
<ipython-input-1-f50815ca8f92> in <cell line: 1>()
----> 1 grid_search_prep.best_params_

NameError: name 'grid_search_prep' is not defined
```

STACK OVERFLOW 검색

다음은 오류라고 할 수는 없고, “SVR”을 import하여 마지막 연습문제까지 실습을 하였으나, 마지막에 실수로 연결이 끊긴 후 코드를 실행하여 생긴 오류입니다. 이것을 제대로 실행하기 위해 다시 몇 시간을 기다려 위의 코드를 실행할 자신이 없어 오류가 발생한 채로 저장하였습니다.

최상의 imputer 정책은 most\_frequent 이고 거의 모든 특성이 유용합니다(16개 중 15개). 마지막 특성(ISLAND)은 잡음이 추가될 뿐입니다.



handson\_ml2\_03.ipynb

파일 수정 보기 삽입 런타임 도구 도움말 모든 변경사항이 저장됨



+ 코드 + 텍스트

+ 코드 plot\_confusion\_matrix() 함수

사이킷런 0.22 버전부터는 `sklearn.metrics.plot_confusion_matrix()` 함수를 사용할 수 있다. 예를 들어, 지금까지 학습된 SGD 분류기의 오차행렬은 다음과 같이 구한다.

- `include_values=False`: 오차 행렬에 사용된 값 함께 표시 여부 지정. 여기서는 표시 안함.
- `cmap='gray'`: 흑백 이미지로 표현. 지정하지 않으면 컬러 이미지 보여줌.

```
!pip install --upgrade scikit-learn
```

```
Requirement already satisfied: scikit-learn in /usr/local/lib/python3.10/dist-packages (1.2.2)
Collecting scikit-learn
  Downloading scikit_learn-1.3.2-cp310-cp310-manylinux_2_17_x86_64_manylinux2014_x86_64.whl (10.8 MB)
    10.8/10.8 MB 22.7 MB/s eta 0:00:00
Requirement already satisfied: numpy<2.0,>=1.17.3 in /usr/local/lib/python3.10/dist-packages (from scikit-learn) (1.23.5)
Requirement already satisfied: scipy>=1.5.0 in /usr/local/lib/python3.10/dist-packages (from scikit-learn) (1.11.4)
Requirement already satisfied: joblib>=1.1.1 in /usr/local/lib/python3.10/dist-packages (from scikit-learn) (1.3.2)
Requirement already satisfied: threadpoolctl>=2.0.0 in /usr/local/lib/python3.10/dist-packages (from scikit-learn) (3.2.0)
Installing collected packages: scikit-learn
  Attempting uninstall: scikit-learn
    Found existing installation: scikit-learn 1.2.2
    Uninstalling scikit-learn-1.2.2:
      Successfully uninstalled scikit-learn-1.2.2
Successfully installed scikit-learn-1.3.2
```

**WARNING: The following packages were previously imported in this runtime:**  
[sklearn]

You must restart the runtime in order to use newly installed versions.

[RESTART SESSION](#)

cannot import name 'plot\_confusion\_matrix' from 'sklearn.metrics'

위 오류를 해결하기 위해 “!pip install --upgrade scikit-learn” 명령어로

업그레이드도(사진 속에는 !pip가 아닌 pip로 돼있는데, !pip로 다시

업그레이드하였습니다.) 해보고, 다른 방법들을 찾아보았지만 결국 해결하지

못 했습니다.

```
[73] from sklearn.metrics import plot_confusion_matrix
```

```
plot_confusion_matrix(sgd_clf, X_train, y_train, include_values=False, cmap='gray')
```

```
ImportError                                Traceback (most recent call last)
<ipython-input-73-4dc5917d8995> in <cell line: 1>()
----> 1 from sklearn.metrics import plot_confusion_matrix
      2
      3 plot_confusion_matrix(sgd_clf, X_train, y_train, include_values=False, cmap='gray')
```

```
ImportError: cannot import name 'plot_confusion_matrix' from 'sklearn.metrics' (/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_init_.py)
```

NOTE: If your import is failing due to a missing package, you can manually install dependencies using either !pip or !apt.

To view examples of installing some common dependencies, click the