Pandas 라이브러리 IRIS 데이터 셋 실습해보기

학습 내용

- scikit-learn를 활용한 머신러닝 모델 구축 실습

01 데이터 준비

In [18]:
```
import pandas as pd
import seaborn as sns
import numpy as np

print(pd.__version__)
iris = sns.load_dataset("iris")
iris
```

```
1.1.3
```

Out[18]:

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa
...	...	...	...	...	...
145	6.7	3.0	5.2	2.3	virginica
146	6.3	2.5	5.0	1.9	virginica
147	6.5	3.0	5.2	2.0	virginica
148	6.2	3.4	5.4	2.3	virginica
149	5.9	3.0	5.1	1.8	virginica

150 rows × 5 columns

02 라이브러리 불러오기

In [19]:
```
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import train_test_split
```

03 모델 구축 및 예측, 평가

In [20]:
```
model = RandomForestClassifier(max_depth=5, n_estimators=10)

X = iris.iloc[:, 0:4]
y = iris.iloc[:, 4]

X.shape, y.shape
```

Out[20]:
```
((150, 4), (150,))
```

In [21]:
```
y
```

Out[21]:
```
0          setosa
1          setosa
2          setosa
3          setosa
4          setosa
          ...
145    virginica
146    virginica
147    virginica
148    virginica
149    virginica
Name: species, Length: 150, dtype: object
```

In [22]:
```
# test 30%, train 70% 로 분할
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size= 0.3)
model.fit(X_train, y_train)

y_pred = model.predict(X_test)
print( len(y_pred) )
print( y_pred[0:10] )
```

```
45
['versicolor' 'setosa' 'versicolor' 'virginica' 'versicolor' 'setosa'
 'setosa' 'versicolor' 'setosa' 'setosa']
```

In [23]:
```
df_iris = pd.DataFrame(list(zip(y_pred, y_test)), columns=['pred_val', 'actua
df_iris['correct'] = df_iris.apply(lambda x:
                                   1 if x['pred_val'] == x['actual'] else 0,
df_iris.head(10)
```

Out[23]:

	pred_val	actual	correct
0	versicolor	versicolor	1
1	setosa	setosa	1
2	versicolor	versicolor	1
3	virginica	virginica	1
4	versicolor	versicolor	1
5	setosa	setosa	1
6	setosa	setosa	1
7	versicolor	versicolor	1
8	setosa	setosa	1
9	setosa	setosa	1

In [24]:
```
np.mean( df_iris['correct'] )
```

Out[24]:
```
0.9555555555555556
```