Iris 데이터셋 PCA 이용하여 차원 축소

라이브러리 및 패키지 불러오기

In [1]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

from sklearn.datasets import load_iris
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import accuracy_score

from sklearn.decomposition import PCA
```

```
C:\Users\user\AppData\Local\Temp\ipykernel_18216\205787104.py:1: DeprecationWarning:
Pyarrow will become a required dependency of pandas in the next major release of pandas (pandas 3.0),
(to allow more performant data types, such as the Arrow string type, and better interoperability with other lib
raries)
but was not found to be installed on your system.
If this would cause problems for you,
please provide us feedback at https://github.com/pandas-dev/pandas/issues/54466

  import pandas as pd
```

데이터셋 불러오기

In [2]:

```
# 데이터셋 로드
iris = load_iris()
df = pd.DataFrame(data= np.c_[iris.data])
```

In [3]:

```
# 데이터셋 정규화
scaler = StandardScaler()
scaler.fit(df)
df_scaled = scaler.transform(df)

# 데이터 프레임으로 자료형 변환 및 target class 정보 추가
df_scaled = pd.DataFrame(df_scaled, columns= ['sepal length', 'sepal width', 'petal length', 'petal width'])
df_scaled['target'] = iris.target

df_scaled.head()
```

Out[3]:

	sepal length	sepal width	petal length	petal width	target
0	-0.900681	1.019004	-1.340227	-1.315444	0
1	-1.143017	-0.131979	-1.340227	-1.315444	0
2	-1.385353	0.328414	-1.397064	-1.315444	0
3	-1.506521	0.098217	-1.283389	-1.315444	0
4	-1.021849	1.249201	-1.340227	-1.315444	0

PCA 차원 축소 적용

In [4]:

```
# 2차원으로 차원 축소, target 정보는 제외
pca = PCA(n_components = 2)
pca.fit(df_scaled.iloc[:,:-1])

# pca transform 후 데이터프레임으로 자료형 변경
df_pca = pca.transform(df_scaled.iloc[:,:-1])
df_pca = pd.DataFrame(df_pca, columns = ['component 0', 'component 1'])

df_pca
```

Out[4]:

	component 0	component 1
0	-2.264703	0.480027
1	-2.080961	-0.674134
2	-2.364229	-0.341908
3	-2.299384	-0.597395
4	-2.389842	0.646835
...	...	...
145	1.870503	0.386966
146	1.564580	-0.896687
147	1.521170	0.269069
148	1.372788	1.011254
149	0.960656	-0.024332

150 rows × 2 columns

In [5]:

```
# PCA 주성분 설명력 출력
print(pca.explained_variance_ratio_)
```

```
[0.72962445 0.22850762]
```

PCA 차원축소 결과 시각화

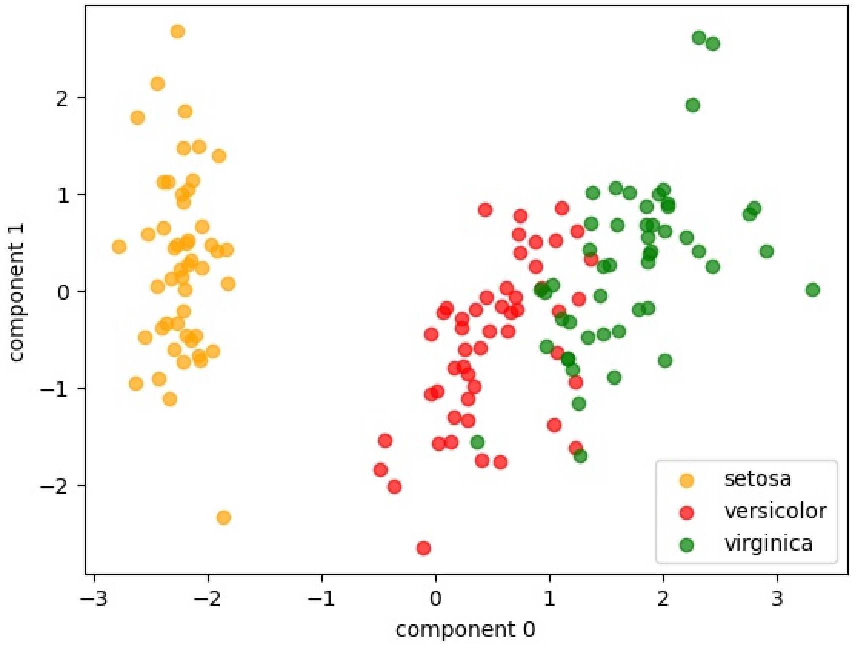
In [6]:

```
# class target 정보 불러오기
df_pca['target'] = df_scaled['target']

# target 별 분리
df_pca_0 = df_pca[df_pca['target'] == 0]
df_pca_1 = df_pca[df_pca['target'] == 1]
df_pca_2 = df_pca[df_pca['target'] == 2]

# target 별 시각화
plt.scatter(df_pca_0['component 0'], df_pca_0['component 1'], color = 'orange', alpha = 0.7, label = 'setosa')
plt.scatter(df_pca_1['component 0'], df_pca_1['component 1'], color = 'red', alpha = 0.7, label = 'versicolor')
plt.scatter(df_pca_2['component 0'], df_pca_2['component 1'], color = 'green', alpha = 0.7, label = 'virginica'

plt.xlabel('component 0')
plt.ylabel('component 1')
plt.legend()
plt.show()
```



In [ ]:

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js