pca

May 14, 2025

1 Improting the dataset

[58]:
```
import pandas as pd

url="https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data"
names=['sepal_length','sepal_width','petal_length','petal_width','species']

df=pd.read_csv(url,names=names);

print(df.head())
```

```
   sepal_length  sepal_width  petal_length  petal_width      species
0           5.1          3.5           1.4          0.2  Iris-setosa
1           4.9          3.0           1.4          0.2  Iris-setosa
2           4.7          3.2           1.3          0.2  Iris-setosa
3           4.6          3.1           1.5          0.2  Iris-setosa
4           5.0          3.6           1.4          0.2  Iris-setosa
```

2 Preprocessing

[69]:
```
from sklearn.preprocessing import StandardScaler

features=['sepal_length','sepal_width','petal_length','petal_width']

x=df.loc[:,features].values
y=df.loc[:,['species']].values

x=StandardScaler().fit_transform(x)

from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.
↪3,random_state=42)
```

3 Creating PCA

[70]:
```
from sklearn.decomposition import PCA
pca=PCA(n_components=3)
principalComponents=pca.fit_transform(x)
principalDf=pd.DataFrame(data=principalComponents,columns=['principal component␣
↪1', 'principal component 2', 'principal component 3'])
```

[71]:
```
finalDf=pd.concat([principalDf,df[['species']]],axis=1)
finalDf.head()
```

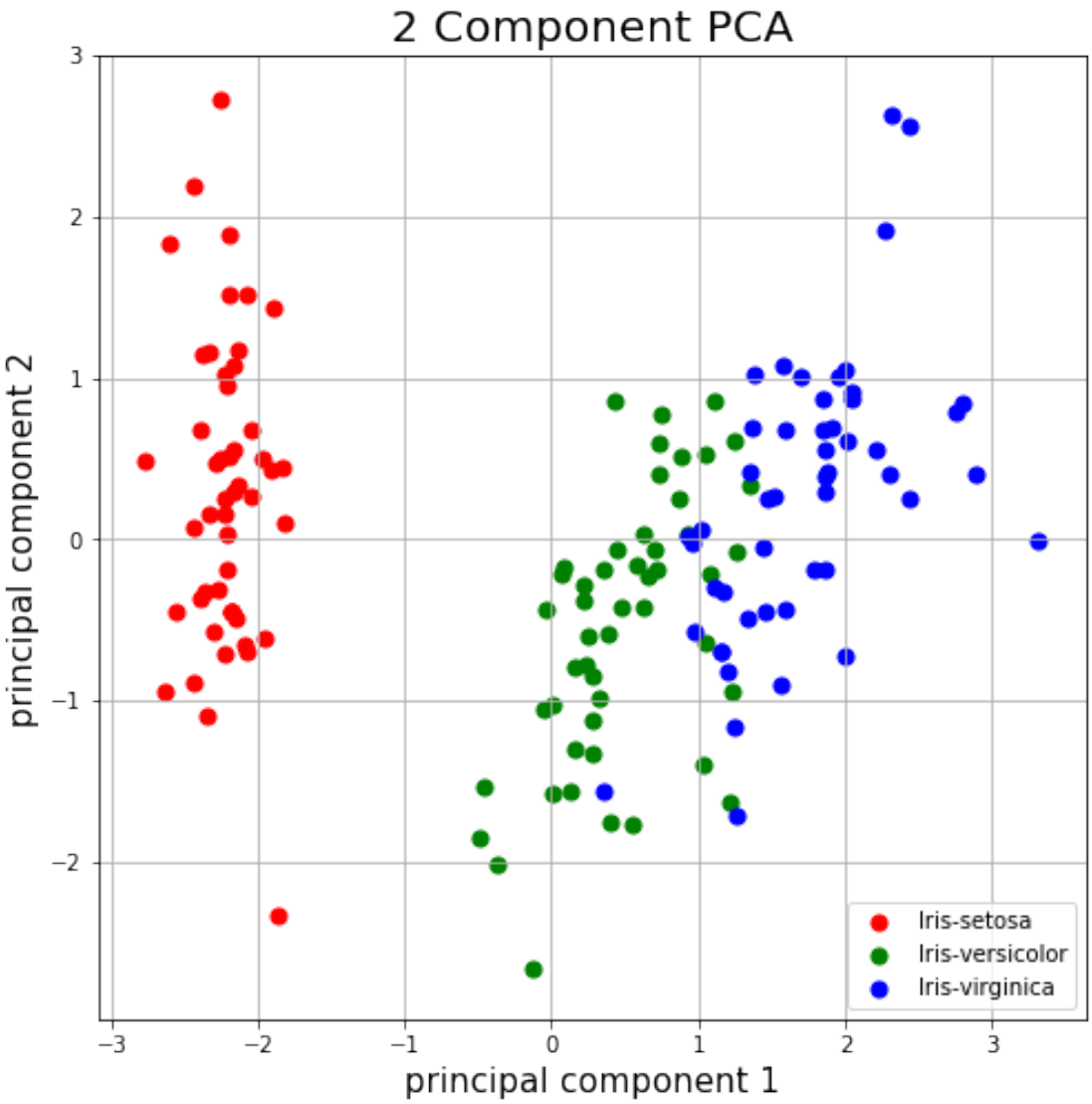
[71]:

	principal component 1	principal component 2	principal component 3	\
0	-2.264542	0.505704	-0.121943	
1	-2.086426	-0.655405	-0.227251	
2	-2.367950	-0.318477	0.051480	
3	-2.304197	-0.575368	0.098860	
4	-2.388777	0.674767	0.021428	

	species
0	Iris-setosa
1	Iris-setosa
2	Iris-setosa
3	Iris-setosa
4	Iris-setosa

4 Plot the graph

[72]:
```
import matplotlib.pyplot as plt
fig=plt.figure(figsize=(8,8))
ax=fig.add_subplot(1,1,1)
ax.set_xlabel('principal component 1',fontsize=15)
ax.set_ylabel('principal component 2',fontsize=15)
ax.set_title('2 Component PCA',fontsize=20)
targets=['Iris-setosa', 'Iris-versicolor', 'Iris-virginica']
colors=['r','g','b']
for target,color in zip(targets,colors):
    ind=finalDf['species']==target
    ax.scatter(finalDf.loc[ind,'principal component 1']
               ,finalDf.loc[ind,'principal component 2']
               ,c=color
               ,s=50
              )
ax.legend(targets)
ax.grid()
```



[73]:
```
pca.explained_variance_ratio_
```

[73]:
```
array([0.72770452, 0.23030523, 0.03683832])
```

5 Using Perceptron without PCA

[74]:
```
from sklearn.linear_model import Perceptron
model=Perceptron(max_iter=1000,random_state=43,tol=1e-3)
model.fit(x_train,y_train.ravel())
pred=model.predict(x_test)
print(pred)
```

```
['Iris-versicolor' 'Iris-setosa' 'Iris-virginica' 'Iris-versicolor'
 'Iris-versicolor' 'Iris-setosa' 'Iris-versicolor' 'Iris-virginica'
 'Iris-versicolor' 'Iris-versicolor' 'Iris-virginica' 'Iris-setosa'
 'Iris-setosa' 'Iris-setosa' 'Iris-setosa' 'Iris-virginica'
 'Iris-virginica' 'Iris-versicolor' 'Iris-versicolor' 'Iris-virginica'
 'Iris-setosa' 'Iris-virginica' 'Iris-setosa' 'Iris-virginica'
 'Iris-virginica' 'Iris-virginica' 'Iris-virginica' 'Iris-virginica'
 'Iris-setosa' 'Iris-setosa' 'Iris-setosa' 'Iris-setosa' 'Iris-setosa'
 'Iris-setosa' 'Iris-setosa' 'Iris-virginica' 'Iris-setosa' 'Iris-setosa'
 'Iris-setosa' 'Iris-setosa' 'Iris-virginica' 'Iris-setosa'
 'Iris-virginica' 'Iris-setosa' 'Iris-setosa']
```

[75]:
```
from sklearn.metrics import accuracy_score
print(accuracy_score(y_test,pred))
```

```
0.8888888888888888
```

6 Seperating the features

[76]:
```
x1=finalDf.drop(['species'],axis=1)
x1=StandardScaler().fit_transform(x1)
x1_train,x1_test,y_train,y_test=train_test_split(x1,y,test_size=0.
 ↪3,random_state=42)
```

7 Using Perceptron with PCA

[77]:
```
model=Perceptron(max_iter=1000,random_state=43,tol=1e-3)
model.fit(x1_train,y_train.ravel())
pred=model.predict(x1_test)
print(pred)
```

```
['Iris-versicolor' 'Iris-setosa' 'Iris-virginica' 'Iris-virginica'
 'Iris-versicolor' 'Iris-setosa' 'Iris-setosa' 'Iris-virginica'
 'Iris-virginica' 'Iris-versicolor' 'Iris-virginica' 'Iris-setosa'
 'Iris-setosa' 'Iris-setosa' 'Iris-setosa' 'Iris-virginica'
 'Iris-virginica' 'Iris-versicolor' 'Iris-virginica' 'Iris-virginica'
 'Iris-setosa' 'Iris-virginica' 'Iris-setosa' 'Iris-virginica'
 'Iris-virginica' 'Iris-virginica' 'Iris-virginica' 'Iris-virginica'
 'Iris-setosa' 'Iris-setosa' 'Iris-setosa' 'Iris-setosa' 'Iris-versicolor'
 'Iris-setosa' 'Iris-setosa' 'Iris-virginica' 'Iris-virginica'
 'Iris-setosa' 'Iris-setosa' 'Iris-setosa' 'Iris-virginica'
 'Iris-virginica' 'Iris-virginica' 'Iris-setosa' 'Iris-setosa']
```

[78]:
```
accuracy_score(y_test,pred)
```

[78]:
```
0.8222222222222222
```