

Class period 17

บทที่ 7 การแสดงและการกระจายของข้อมูล
Visualize_Data_Distribution_PCA

1

Projection

- การฉายลงบนแกนตั้งหรือแกนนอนเรียกว่าการฉาย
- การฉายลงบนแกนตั้งเรียกว่าการฉายแนวตั้ง
- การฉายลงบนแกนนอนเรียกว่าการฉายแนวนอน

2

PCA (Principal component Analysis)

- PCA คือ การลดมิติของข้อมูลโดยการฉายข้อมูลลงบนแกนตั้งหรือแกนนอน
- PCA มี 2 ขั้นตอน 1. หาแกนตั้งหรือแกนนอนที่เหมาะสมที่สุด 2. หาแกนตั้งหรือแกนนอนที่เหมาะสมที่สุด

3

sklearn -> scikit-learn

- <https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html>
- เป็น package ที่รวม function ทั่วๆ ไป Data Science - Machine Learning มาไว้ใช้ทำงานได้สะดวก
- การนำเข้า PCA ใช้ sklearn
- from sklearn.decomposition import PCA

4

การใช้ PCA มี 3 ขั้นตอน

1. Import
 - from sklearn.decomposition import PCA
2. Define
 - pca = PCA()
3. Fit - Transform คือ คำสั่งที่ใช้สำหรับคำนวณแกนหลักในการอธิบายการกระจายตัวของข้อมูลที่มีอยู่
- new_axis = pca.fit_transform("ชื่อไฟล์ข้อมูลที่ต้องการใช้ PCA")

5

เตรียมข้อมูลดอกไม้อิริส

```
import pandas as pd
example_df = pd.read_csv("https://raw.githubusercontent.com/pandas-dev/pandas/master/pandas/tests/io/data/csv/iris.csv")
thisdata = example_df.iloc[:, :-1]
```

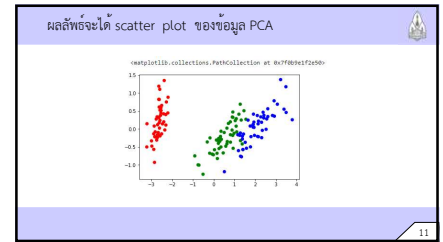
Species	Sepal length	Petal length	Petal width	Height
Setosa	4.3	1.5	0.2	0.1
Setosa	4.7	1.4	0.3	0.1
Setosa	4.0	1.2	0.1	0.1
Setosa	4.9	1.6	0.4	0.1
Setosa	4.4	1.3	0.2	0.1
Setosa	5.0	1.6	0.4	0.1
Setosa	4.4	1.4	0.2	0.1
Setosa	4.5	1.5	0.2	0.1
Setosa	4.7	1.4	0.3	0.1
Setosa	4.0	1.2	0.1	0.1
Setosa	4.9	1.6	0.4	0.1
Setosa	4.4	1.3	0.2	0.1
Setosa	5.0	1.6	0.4	0.1
Setosa	4.4	1.4	0.2	0.1
Setosa	4.5	1.5	0.2	0.1
Setosa	4.7	1.4	0.3	0.1
Setosa	4.0	1.2	0.1	0.1
Setosa	4.9	1.6	0.4	0.1
Setosa	4.4	1.3	0.2	0.1
Setosa	5.0	1.6	0.4	0.1
Setosa	4.4	1.4	0.2	0.1
Setosa	4.5	1.5	0.2	0.1
Setosa	4.7	1.4	0.3	0.1
Setosa	4.0	1.2	0.1	0.1
Setosa	4.9	1.6	0.4	0.1
Setosa	4.4	1.3	0.2	0.1
Setosa	5.0	1.6	0.4	0.1
Setosa	4.4	1.4	0.2	0.1
Setosa	4.5	1.5	0.2	0.1
Setosa	4.7	1.4	0.3	0.1
Setosa	4.0	1.2	0.1	0.1
Setosa	4.9	1.6	0.4	0.1
Setosa	4.4	1.3	0.2	0.1
Setosa	5.0	1.6	0.4	0.1
Setosa	4.4	1.4	0.2	0.1
Setosa	4.5	1.5	0.2	0.1
Setosa	4.7	1.4	0.3	0.1
Setosa	4.0	1.2	0.1	0.1
Setosa	4.9	1.6	0.4	0.1
Setosa	4.4	1.3	0.2	0.1
Setosa	5.0	1.6	0.4	0.1
Setosa	4.4	1.4	0.2	0.1
Setosa	4.5	1.5	0.2	0.1
Setosa	4.7	1.4	0.3	0.1
Setosa	4.0	1.2	0.1	0.1
Setosa	4.9	1.6	0.4	0.1
Setosa	4.4	1.3	0.2	0.1
Setosa	5.0	1.6	0.4	0.1
Setosa	4.4	1.4	0.2	0.1
Setosa	4.5	1.5	0.2	0.1
Setosa	4.7	1.4	0.3	0.1
Setosa	4.0	1.2	0.1	0.1
Setosa	4.9	1.6	0.4	0.1
Setosa	4.4	1.3	0.2	0.1
Setosa	5.0	1.6	0.4	0.1
Setosa	4.4	1.4	0.2	0.1
Setosa	4.5	1.5	0.2	0.1
Setosa	4.7	1.4	0.3	0.1
Setosa	4.0	1.2	0.1	0.1
Setosa	4.9	1.6	0.4	0.1
Setosa	4.4	1.3	0.2	0.1
Setosa	5.0	1.6	0.4	0.1
Setosa	4.4	1.4	0.2	0.1
Setosa	4.5	1.5	0.2	0.1
Setosa	4.7	1.4	0.3	0.1
Setosa	4.0	1.2	0.1	0.1
Setosa	4.9	1.6	0.4	0.1
Setosa	4.4	1.3	0.2	0.1
Setosa	5.0	1.6	0.4	0.1
Setosa	4.4	1.4	0.2	0.1
Setosa	4.5	1.5	0.2	0.1
Setosa	4.7	1.4	0.3	0.1
Setosa	4.0	1.2	0.1	0.1
Setosa	4.9	1.6	0.4	0.1
Setosa	4.4	1.3	0.2	0.1
Setosa	5.0	1.6	0.4	0.1
Setosa	4.4	1.4	0.2	0.1
Setosa	4.5	1.5	0.2	0.1
Setosa	4.7	1.4	0.3	0.1
Setosa	4.0	1.2	0.1	0.1
Setosa	4.9	1.6	0.4	0.1
Setosa	4.4	1.3	0.2	0.1
Setosa	5.0	1.6	0.4	0.1
Setosa	4.4	1.4	0.2	0.1
Setosa	4.5	1.5	0.2	0.1
Setosa	4.7	1.4	0.3	0.1
Setosa	4.0	1.2	0.1	0.1
Setosa	4.9	1.6	0.4	0.1
Setosa	4.4	1.3	0.2	0.1
Setosa	5.0	1.6	0.4	0.1
Setosa	4.4	1.4	0.2	0.1
Setosa	4.5	1.5	0.2	0.1
Setosa	4.7	1.4	0.3	0.1
Setosa	4.0	1.2	0.1	0.1
Setosa	4.9	1.6	0.4	0.1
Setosa	4.4	1.3	0.2	0.1
Setosa	5.0	1.6	0.4	0.1
Setosa	4.4	1.4	0.2	0.1
Setosa	4.5	1.5	0.2	0.1
Setosa	4.7	1.4	0.3	0.1
Setosa	4.0	1.2	0.1	0.1
Setosa	4.9	1.6	0.4	0.1
Setosa	4.4	1.3	0.2	0.1
Setosa	5.0	1.6	0.4	0.1
Setosa	4.4	1.4	0.2	0.1
Setosa	4.5	1.5	0.2	0.1
Setosa	4.7	1.4	0.3	0.1
Setosa	4.0	1.2	0.1	0.1
Setosa	4.9	1.6	0.4	0.1
Setosa	4.4	1.3	0.2	0.1
Setosa	5.0	1.6	0.4	0.1
Setosa	4.4	1.4	0.2	0.1
Setosa	4.5	1.5	0.2	0.1
Setosa	4.7	1.4	0.3	0.1
Setosa	4.0	1.2	0.1	0.1
Setosa	4.9	1.6	0.4	0.1
Setosa	4.4	1.3	0.2	0.1
Setosa	5.0	1.6	0.4	0.1
Setosa	4.4	1.4	0.2	0.1
Setosa	4.5	1.5	0.2	0.1
Setosa	4.7	1.4	0.3	0.1
Setosa	4.0	1.2	0.1	0.1
Setosa	4.9	1.6	0.4	0.1
Setosa	4.4	1.3	0.2	0.1
Setosa	5.0	1.6	0.4	0.1
Setosa	4.4	1.4	0.2	0.1
Setosa	4.5	1.5	0.2	0.1
Setosa	4.7	1.4	0.3	0.1
Setosa	4.0	1.2	0.1	0.1
Setosa	4.9	1.6	0.4	0.1
Setosa	4.4	1.3	0.2	0.1
Setosa	5.0	1.6	0.4	0.1
Setosa	4.4	1.4	0.2	0.1
Setosa	4.5	1.5	0.2	0.1
Setosa	4.7	1.4	0.3	0.1
Setosa	4.0	1.2	0.1	0.1
Setosa	4.9	1.6	0.4	0.1
Setosa	4.4	1.3	0.2	0.1
Setosa	5.0	1.6	0.4	0.1
Setosa	4.4	1.4	0.2	0.1
Setosa	4.5	1.5	0.2	0.1
Setosa	4.7	1.4	0.3	0.1
Setosa	4.0	1.2	0.1	0.1
Setosa	4.9	1.6	0.4	0.1
Setosa	4.4	1.3	0.2	0.1
Setosa	5.0	1.6	0.4	0.1
Setosa	4.4	1.4	0.2	0.1
Setosa	4.5	1.5	0.2	0.1
Setosa	4.7	1.4	0.3	0.1
Setosa	4.0	1.2	0.1	0.1
Setosa	4.9	1.6	0.4	0.1
Setosa	4.4	1.3	0.2	0.1
Setosa	5.0	1.6	0.4	0.1
Setosa	4.4	1.4	0.2	0.1
Setosa	4.5	1.5	0.2	0.1
Setosa	4.7	1.4	0.3	0.1
Setosa	4.0	1.2	0.1	0.1
Setosa	4.9	1.6	0.4	0.1
Setosa	4.4	1.3	0.2	0.1
Setosa	5.0	1.6	0.4	0.1
Setosa	4.4	1.4	0.2	0.1
Setosa	4.5	1.5	0.2	0.1
Setosa	4.7	1.4	0.3	0.1
Setosa	4.0	1.2	0.1	0.1
Setosa	4.9	1.6	0.4	0.1
Setosa	4.4	1.3	0.2	0.1
Setosa	5.0	1.6	0.4	0.1
Setosa	4.4	1.4	0.2	0.1
Setosa	4.5	1.5	0.2	0.1
Setosa	4.7	1.4	0.3	0.1
Setosa	4.0	1.2	0.1	0.1
Setosa	4.9	1.6	0.4	0.1
Setosa	4.4	1.3	0.2	0.1
Setosa	5.0	1.6	0.4	0.1
Setosa	4.4	1.4	0.2	0.1
Setosa	4.5	1.5	0.2	0.1
Setosa	4.7	1.4	0.3	0.1
Setosa	4.0	1.2	0.1	0.1
Setosa	4.9	1.6	0.4	0.1
Setosa	4.4	1.3	0.2	0.1
Setosa	5.0	1.6	0.4	0.1
Setosa	4.4	1.4	0.2	0.1
Setosa	4.5	1.5	0.2	0.1
Setosa	4.7	1.4	0.3	0.1
Setosa	4.0	1.2	0.1	0.1
Setosa	4.9	1.6	0.4	0.1
Setosa	4.4	1.3	0.2	0.1
Setosa	5.0	1.6	0.4	0.1
Setosa	4.4	1.4	0.2	0.1
Setosa	4.5	1.5	0.2	0.1
Setosa	4.7	1.4	0.3	0.1
Setosa	4.0	1.2	0.1	0.1
Setosa	4.9	1.6	0.4	0.1
Setosa	4.4	1.3	0.2	0.1
Setosa	5.0	1.6	0.4	0.1
Setosa	4.4	1.4	0.2	0.1
Setosa	4.5	1.5	0.2	0.1
Setosa	4.7	1.4	0.3	0.1
Setosa	4.0	1.2	0.1	0.1
Setosa	4.9	1.6	0.4	0.1
Setosa	4.4	1.3	0.2	0.1
Setosa	5.0	1.6	0.4	0.1
Setosa	4.4	1.4	0.2	0.1
Setosa	4.5	1.5	0.2	0.1
Setosa	4.7	1.4	0.3	0.1
Setosa	4.0	1.2	0.1	0.1
Setosa	4.9	1.6	0.4	0.1
Setosa	4.4	1.3	0.2	0.1
Setosa	5.0	1.6	0.4	0.1
Setosa	4.4	1.4	0.2	0.1
Setosa	4.5	1.5	0.2	0.1
Setosa	4.7	1.4	0.3	0.1
Setosa	4.0	1.2	0.1	0.1
Setosa	4.9	1.6	0.4	0.1
Setosa	4.4	1.3	0.2	0.1
Setosa	5.0	1.6	0.4	0.1
Setosa	4.4	1.4	0.2	0.1
Setosa	4.5	1.5	0.2	0.1
Setosa	4.7	1.4	0.3	0.1
Setosa	4.0	1.2	0.1	0.1
Setosa	4.9	1.6	0.4	0.1
Setosa	4.4	1.3	0.2	0.1
Setosa	5.0	1.6	0.4	0.1
Setosa	4.4	1.4	0.2	0.1
Setosa	4.5	1.5	0.2	0.1
Setosa	4.7	1.4	0.3	0.1
Setosa	4.0	1.2	0.1	0.1
Setosa	4.9	1.6	0.4	0.1
Setosa	4.4	1.3	0.2	0.1
Setosa	5.0	1.6	0.4	0.1
Setosa	4.4	1.4	0.2	0.1
Setosa	4.5	1.5	0.2	0.1
Setosa	4.7	1.4	0.3	0.1
Setosa	4.0	1.2	0.1	0.1
Setosa	4.9	1.6	0.4	0.1
Setosa	4.4	1.3	0.2	0.1
Setosa	5.0	1.6	0.4	0.1
Setosa	4.4	1.4	0.2	0.1
Setosa	4.5	1.5	0.2	0.1
Setosa	4.7	1.4	0.3	0.1
Setosa	4.0	1.2	0.1	0.1
Setosa	4.9	1.6	0.4	0.1
Setosa	4.4	1.3	0.2	0.1
Setosa	5.0	1.6	0.4	0.1
Setosa	4.4	1.4	0.2	0.1
Setosa	4.5	1.5	0.2	0.1
Setosa	4.7	1.4	0.3	0.1
Setosa	4.0	1.2	0.1	0.1
Setosa	4.9	1.6	0.4	0.1
Setosa	4.4	1.3	0.2	0.1
Setosa	5.0	1.6	0.4	0.1
Setosa	4.4	1.4	0.2	0.1
Setosa	4.5	1.5	0.2	0.1
Setosa	4.7	1.4	0.3	0.1
Setosa	4.0	1.2	0.1	0.1
Setosa	4.9	1.6	0.4	0.1
Setosa	4.4	1.3	0.2	0.1
Setosa	5.0	1.6	0.4	0.1



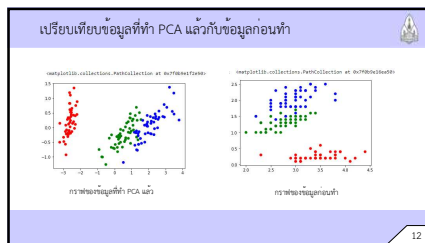
10



11



12



13