blatt07_Guth_Venker_Jaekel

June 15, 2021

1 Abgabe SMD Blatt 07

1.0.1 von Nico Guth, David Venker, Jan Jäkel

[1]:
```
import numpy as np
import matplotlib.pyplot as plt
import matplotlib as mpl
import sklearn.datasets
import pandas as pd
```

[2]:
```
mpl.rcParams['font.size'] = 15
```

2 Aufgabe 15

2.1 a) Datensatz laden

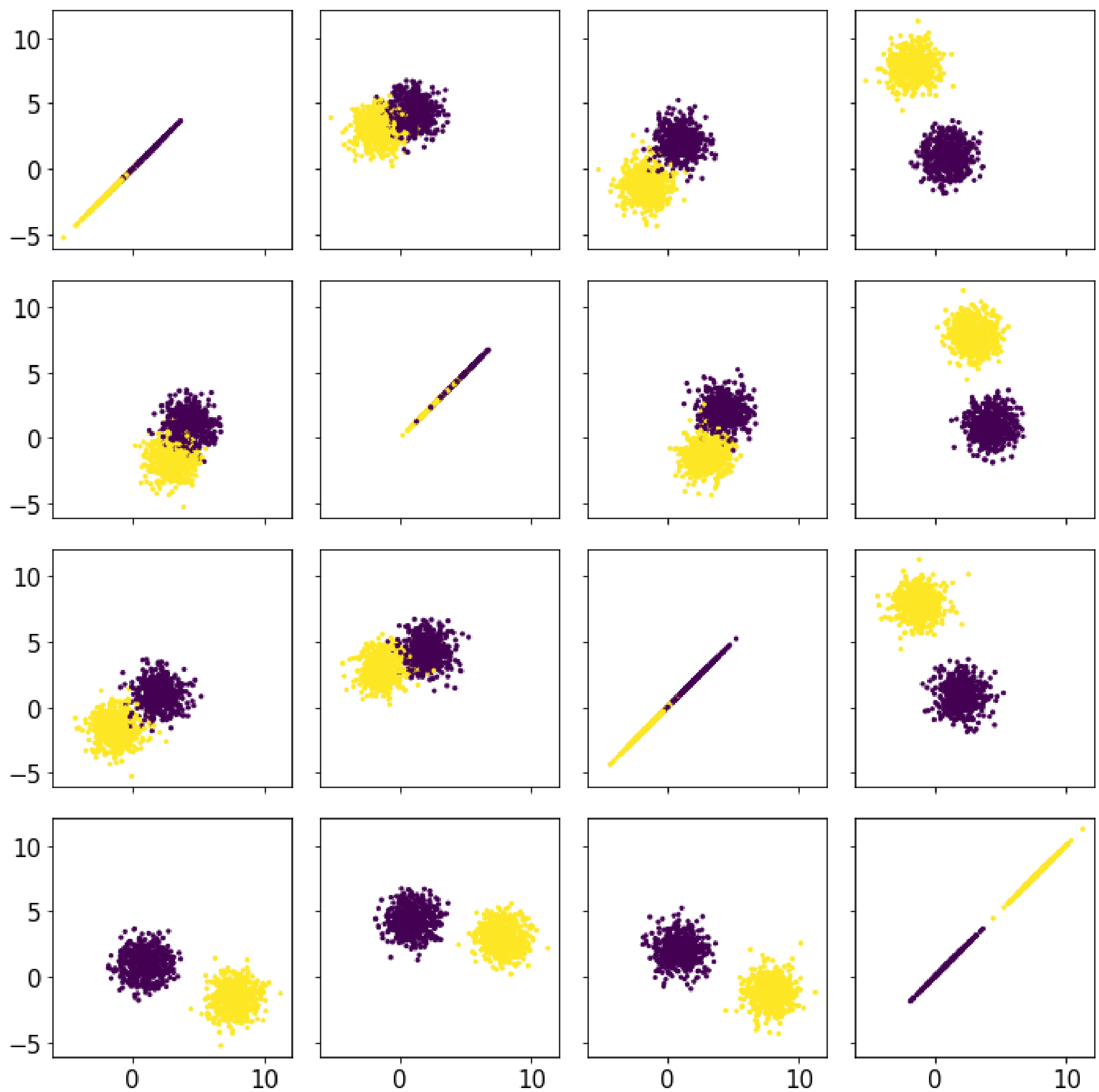
[3]:
```
X,y = sklearn.datasets.make_blobs(
    n_samples=1000,
    centers=2,
    n_features=4,
    random_state=0)
```

[4]:
```
X.shape,y.shape
```

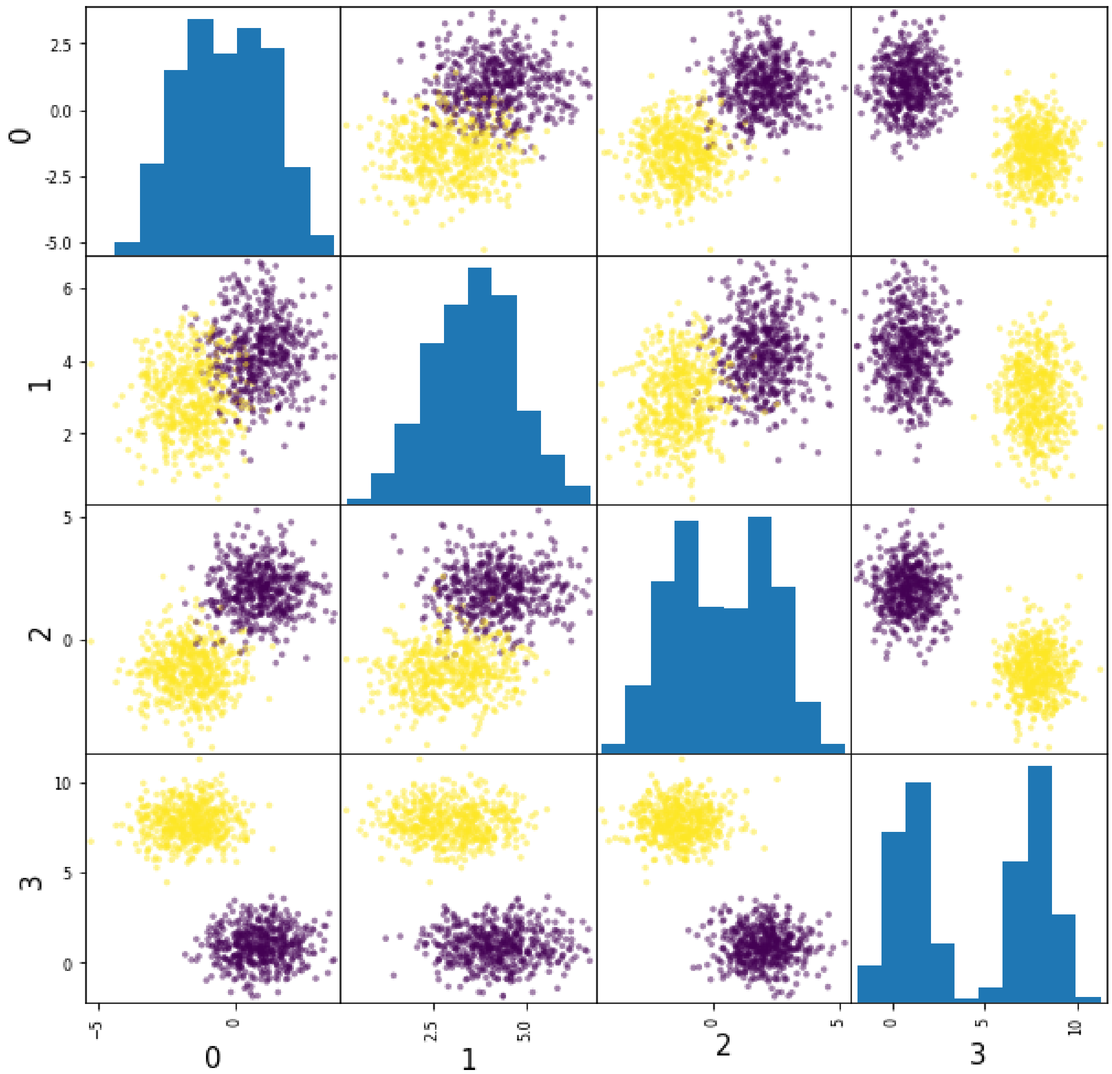
[4]:
```
((1000, 4), (1000,))
```

[5]:
```
# Scatter Matrix
fig, axes = plt.subplots(4,4,figsize=(10,10),sharex=True,sharey=True)

for i in range(4):
    for j in range(4):
        axes[i,j].scatter(X[:,i],X[:,j],s=5,c=y)
        axes[i,j].set_aspect('equal')
plt.tight_layout()
```



[6]:
```
# auch möglich:
pd.plotting.scatter_matrix(pd.DataFrame(X),figsize=(10,10),s=50,c=y);
```



2.2 b) Hauptkomponentenanalyse

[7]:
```
import sklearn.decomposition
```

[8]:
```
pca = sklearn.decomposition.PCA()
X_ = pca.fit_transform(X)
```

[9]:
```
X_.shape
```

[9]: (1000, 4)

[10]:
```
covariance = pca.get_covariance()
covariance
```

[10]: array([[2.62583302, 0.80302067, 2.12234349, -4.3783228],
        [0.80302067, 1.33533433, 1.09428645, -2.18876646],
        [2.12234349, 1.09428645, 3.69087992, -5.66291375],
        [-4.3783228 , -2.18876646, -5.66291375, 12.75375474]])

[11]:
```
l, W = np.linalg.eigh(covariance)
print(f'Eigenwerte=\n{l}\n\nEigenvektoren=\n{W}')
```

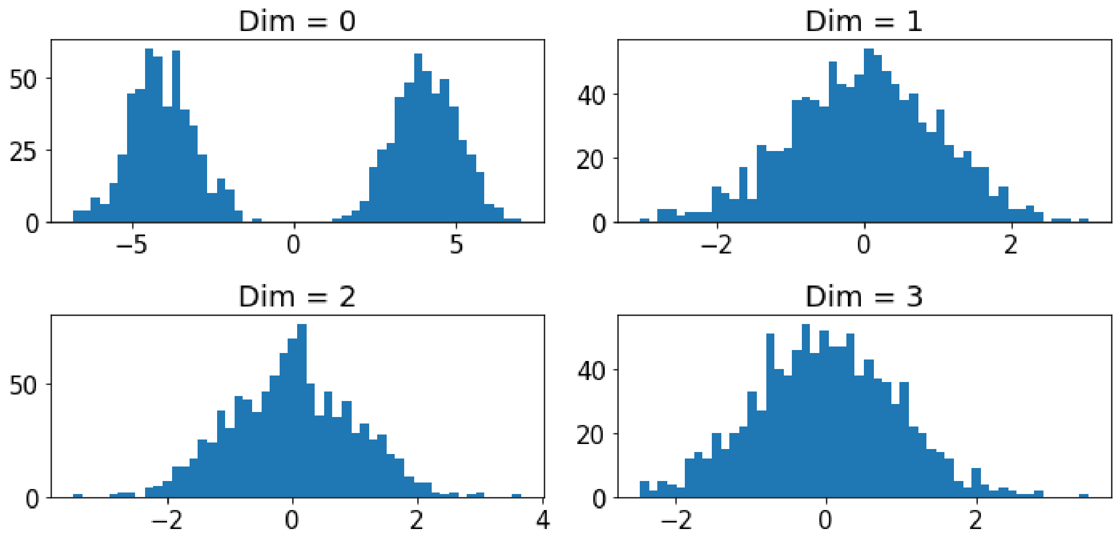
```
Eigenwerte=
[ 0.89875061  0.98813673  0.99958442 17.51933024]

Eigenvektoren=
[[-0.1888122  -0.45174256 -0.81331229 -0.31432743]
 [-0.82630524  0.53883501 -0.04672846 -0.15713986]
 [ 0.52973932  0.66484304 -0.33530128 -0.40611949]
 [ 0.03075479  0.25212938 -0.47319407  0.84354745]]
```

Ein Eigenwert ist deutlich größer, also scheint es zu reichen auf nur diese Diemension zu reduzieren.

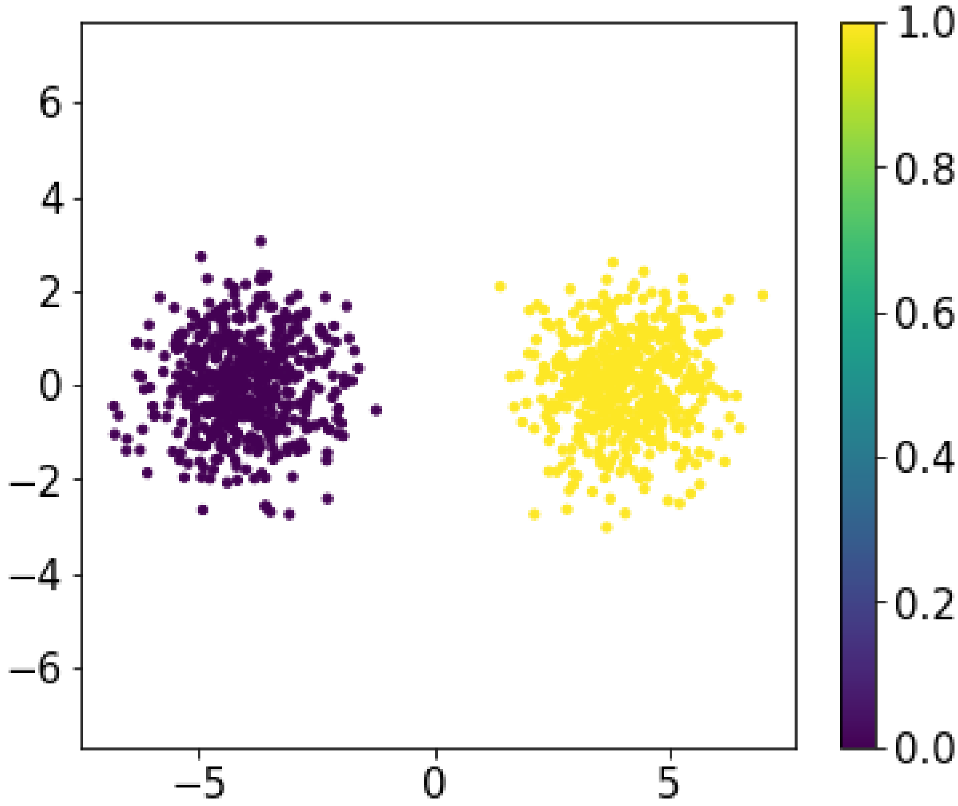
2.3 c) PCA darstellen

[12]:
```
fig,axes = plt.subplots(2,2,figsize=(10,5))
for i,ax in enumerate(axes.flatten()):
    ax.hist(X_[:,i],bins=50)
    ax.set_title(f'Dim = {i}')
plt.tight_layout()
```



[13]:
```
plt.figure(figsize=(6,5))
plt.scatter(X_[:,0],X_[:,1],s=10,c=y)
plt.axis('equal')
```

```
plt.colorbar()
plt.show()
```



[ ]:

```
```

Blatt 7

Freitag, 11. Juni 2021 13:38

Aufgabe 14:

a) Beschreibung der PCA

Ziel: Dimensionsreduktion

gegeben: $N \times d$ Datenmatrix X

jede Reihe stellt einen Punkt dar

Vorgehensweise:

1) Zentriere Daten um Mittelwert
2) Berechne Kovarianzmatrix
3) Berechne Eigenwerte und Eigenvektoren der Kovarianzmatrix
4) Bilde die $d \times k$ Matrix W mit den Eigenvektoren der k größten Eigenwerte als Spalten
5) Wende W auf jede Zeile vom zentrierten X an

$$X = \begin{pmatrix} 1 & 3 & 1 & 2 & 3 & 2 \\ 1 & 0 & 3 & 0 & 1 & 1 \end{pmatrix}^T$$

1)

b)

$X_1 : [1, 3, 1, 2, 3, 2]$ $\qquad \vec{x}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

$X_2 : [1, 0, 3, 0, 1, 1]$ $\qquad \vec{x}_2 = \begin{pmatrix} 3 \\ 0 \end{pmatrix}$

$d = 2 \ , N = 6$ $\qquad \vdots$

1)

$$\vec{\mu} = \frac{1}{N} \sum_{i=1}^{N} \vec{x}_i$$

$$\vec{\mu} = \frac{1}{6} \begin{pmatrix} 12 \\ 6 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

zentrierte Daten: $\vec{x}'_i = \vec{x}_i - \vec{\mu}$

$$X' = \begin{pmatrix} -1 & 1 & -1 & 0 & 1 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 \end{pmatrix}^T$$

2) Kovarianzmatrix:

$$\mathrm{Cov}(X) = E\left[(X - E[X])(X - E[X])^T\right]$$

$$\mathrm{cov}(X) = E[\vec{x}' \cdot \vec{x}'^T]$$

$$= \frac{1}{N} \sum_{i=1}^{N} \vec{x}'_i \cdot \vec{x}'^T_i$$

$$= \frac{1}{6} \Bigg[\begin{pmatrix} -1 \\ 0 \end{pmatrix} (-1 \ \ 0) + \begin{pmatrix} 1 \\ -1 \end{pmatrix} (1 \ \ -1) + \begin{pmatrix} -1 \\ 2 \end{pmatrix} (-1 \ \ 2)$$

$$+ \begin{pmatrix} 0 \\ -1 \end{pmatrix} (0 \ \ -1) + \begin{pmatrix} 1 \\ 0 \end{pmatrix} (1 \ \ 0) + \begin{pmatrix} 0 \\ 0 \end{pmatrix} (0 \ \ 0) \Bigg]$$

$$= \frac{1}{6}\left[\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix} + \begin{pmatrix}1 & -1\\ -1 & 1\end{pmatrix} + \begin{pmatrix}1 & -2\\ -2 & 4\end{pmatrix} + \begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix} + \begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix} + \begin{pmatrix}0 & 0\\ 0 & 0\end{pmatrix}\right]$$

$$= \frac{1}{6}\begin{pmatrix}4 & -3\\ -3 & 6\end{pmatrix} = \begin{pmatrix}\frac{2}{3} & -\frac{1}{2}\\ -\frac{1}{2} & 1\end{pmatrix}$$

3)

Eigenwerte von $Cov(X)$

$$\det(cov(X) - \mathbb{1}\lambda) = 0$$

$$= \left(\frac{2}{3} - \lambda\right)(1-\lambda) - \left(-\frac{1}{2}\right)\left(-\frac{1}{2}\right)$$

$$= \frac{2}{3} - \frac{5}{3}\lambda + \lambda^2 - \frac{1}{4}$$

$$= \lambda^2 - \frac{5}{3}\lambda + \frac{5}{12} = 0$$

pq:

$$\lambda_\pm = -\left(-\frac{5}{6}\right) \pm \sqrt{\left(\frac{5}{6}\right)^2 - \frac{5}{12}}$$

$$\lambda_\pm = \frac{5}{6} \pm \sqrt{\frac{5}{18}} = \frac{1}{6}\left(5 \pm \sqrt{10}\right)$$

$$\max(\lambda) = \frac{1}{6}\left(5 \pm \sqrt{10}\right)$$

Eigenvektoren der größten $k = 1$ Eigenwerte

$$\lambda = \frac{1}{6}\left(5 + \sqrt{10}\right)$$

$$(cov(X) - \lambda\mathbb{1})\vec{v} = 0$$

$$\begin{pmatrix}\frac{2}{3} - \lambda & -\frac{1}{2}\\ & \end{pmatrix}\vec{v} = 0$$

$$\begin{pmatrix} \frac{2}{3}-\lambda & -\frac{1}{2} \\ -\frac{1}{2} & 1-\lambda \end{pmatrix} \vec{v} = 0$$

$$\begin{pmatrix} \frac{1}{6}(-1-\sqrt{10}) & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{6}(1-\sqrt{10}) \end{pmatrix} \vec{V} = 0$$

$$\text{I} \quad -\frac{1}{6}(1+\sqrt{10})\, V_1 - \frac{1}{2} V_2 = 0$$

$$\text{II} \quad -\frac{1}{2} V_1 - \frac{1}{6}(1-\sqrt{10})\, V_2 = 0$$

$$\text{I} \Leftrightarrow \quad V_2 = 2 \cdot \left(-\frac{1}{6}\right)(1+\sqrt{10})\, V_1$$

$$V_2 = -\frac{1}{3}(1+\sqrt{10})\, V_1$$

wähle $V_1 = 3$

$$\Rightarrow V_2 = -1-\sqrt{10}$$

$\vec{V}$ normieren

$$\hat{\vec{V}} = \frac{1}{|\vec{V}|} \vec{v} = \frac{1}{\sqrt{3^2+(1+\sqrt{10})^2}} \begin{pmatrix} 3 \\ -1-\sqrt{10} \end{pmatrix}$$

$$W = \frac{1}{\sqrt{3^2+(1+\sqrt{10})^2}} \begin{pmatrix} 3 \\ -1-\sqrt{10} \end{pmatrix}$$

5) Transformiere X

$$X' = X \cdot W$$

$$= \begin{pmatrix} 1 & 3 & 1 & 2 & 3 & 2 \\ 1 & 0 & 3 & 0 & 1 & 1 \end{pmatrix}^T \begin{pmatrix} 3 \\ -1-\sqrt{10} \end{pmatrix} \frac{1}{\sqrt{3^2+(1+\sqrt{10})^2}}$$

$$= \begin{pmatrix} 3-1-\sqrt{10} \\ 9 \\ 3-3-3\sqrt{10} \\ 6 \\ 9-1-\sqrt{10} \\ 6-1-\sqrt{10} \end{pmatrix} \frac{1}{\sqrt{3^2+(1+\sqrt{10})^2}}$$

$$= \begin{pmatrix} 2-\sqrt{10} \\ 9 \\ -3\sqrt{10} \\ 6 \\ 8-\sqrt{10} \\ 5-\sqrt{10} \end{pmatrix} \frac{1}{\sqrt{3^2+(1+\sqrt{10})^2}}$$

$$= \begin{pmatrix} -0.227 \\ 1.754 \\ -1.849 \\ 1.169 \\ 0.943 \\ 0.358 \end{pmatrix}$$