No. 14 a)

6,5 / 10

1. The data points are centralized to their mean values.
2. Then using the covariance matrix, eigenvalues and eigenvectors, the correlation of values is calculated.

Correlation of the values with what? −0,5?

3. Choice of the main correlation by picking the largest eigenvalue.
4. Filling the property matrix W with the eigenvectors
5. Application of the property matrix to the individual data points
6. The closer the points are on the primary axis the more likely they are correlating.

With what?

while using the word correlation is not completly wrong, the idea is to use the k principal comp. that incorporate the highest amount of information (to some extend the the highest correlation with the target variable)

b)

$x_1 : [1,3,1,2,3,2]$ $\qquad$ $x_2 : [1,0,3,0,1,1]$

Mean values:

$\bar{x}_1 = \frac{1}{6} \cdot (1+3+1+2+3+2) = 2$ $\qquad$ $\bar{x}_2 = \frac{1}{6}(1+0+3+0+1+1) = 1$

New data points: $[x_1' = x_1 - \bar{x}_1]$

$x_1' : [-1,1,-1,0,1,0]$ $\qquad$ $x_2' = [0,-1,2,-1,0,0]$

calculating Covariance Matrix:

$$\operatorname{cov}(x_i, x_j) = \frac{1}{N-1} \sum_{k=1}^{N} (x_{ik}' \cdot x_{jk}')$$

$$\operatorname{cov}(x_1, x_1) = \frac{1}{5} \cdot ((-1)^2 + 1^2 + (-1)^2 + 0^2 + 1^2 + 0^2) = \frac{4}{5}$$

$$\operatorname{cov}(x_2, x_2) = \frac{1}{5}((0^2 + (-1)^2 + (2)^2 + (-1)^2 + 0^2 + 0^2) = 6/5$$

BRUNNEN

$$\operatorname{cov}(x_1, x_2) = \operatorname{cov}(x_2, x_1) = \frac{1}{5} \cdot (-1 \cdot 0 + 1 \cdot (-1) + (-1) \cdot 2 + 0 \cdot (-1) + 1 \cdot 0 + 0 \cdot 0) = -3/5$$

$$S = \frac{1}{5}\begin{pmatrix} 4 & -3 \\ -3 & 6 \end{pmatrix}$$

Eigenvalues and -vectors:

$$\det(S - \lambda I) \stackrel{!}{=} 0$$

$$\frac{1}{5}\cdot\det\begin{pmatrix} 4-\lambda & -3 \\ -3 & 6-\lambda \end{pmatrix} \stackrel{!}{=} 0$$

$\frac{1}{5}\underline{\underline{X}} - \lambda \mathbb{1}$

$\uparrow \stackrel{!}{=} \frac{1}{5}(\underline{\underline{X}} - \lambda \mathbb{1})$ for some matrix $\underline{\underline{X}}$

-1 P

$$\Leftrightarrow \quad 24 - 4\lambda - 6\lambda + \lambda^2 - 9 = 0$$

$$\Leftrightarrow \quad \lambda^2 - 10\lambda + 15 = 0$$

$$\lambda_{1/2} = 5 \pm \sqrt{5^2 - 15}$$

$$= 5 \pm \sqrt{10}$$

$$\lambda_1 = 8{,}162 \qquad \lambda_2 = 1{,}838$$

$\uparrow$ larger

Eigenvector:

$$(S - \lambda_i \cdot I)\,\vec{v}_i = 0$$

$$\begin{pmatrix} (4-\lambda_i)\,u_{i1} & -3\,u_{i2} \\ -3\,u_{i1} & (6-\lambda_i)\,u_{i2} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\text{I}: \quad \frac{u_{i1}}{3} = \frac{u_{i2}}{4-\lambda_i}\,u_{i2} = t$$

for $t = 1$:

$$u_{i1} = 3 \quad \wedge \quad u_{i2} = 4 - \lambda_i$$

$$\vec{u}_i = \begin{pmatrix} 3 \\ 4-\lambda_i \end{pmatrix} \Rightarrow \vec{u}_1 = \begin{pmatrix} 3 \\ -4{,}162 \end{pmatrix} \Rightarrow e_1 = \begin{pmatrix} 0{,}5847 \\ -0{,}8112 \end{pmatrix}$$

$$\xrightarrow{\text{Normalize}}$$

$$\Rightarrow \vec{u}_2 = \begin{pmatrix} 3 \\ 2{,}162 \end{pmatrix} \Rightarrow e_2 = \begin{pmatrix} 0{,}8112 \\ 0{,}5847 \end{pmatrix}$$

Matrix W

$$W = \begin{pmatrix} | & | \\ e_1 & e_2 \\ | & | \end{pmatrix}^T = \begin{pmatrix} 0{,}5847 & 0{,}8112 \\ -0{,}8112 & 0{,}5847 \end{pmatrix}^T$$

Calculating new observations:

$$x' = x W$$

Therefore:

$$\left[P_{1i} = e_1^T \cdot \begin{matrix} x'_{1i} \\ x'_{2i} \end{matrix} \right]$$

$$P_{11} = e_1^T \cdot \begin{pmatrix} -1 \\ 0 \end{pmatrix} = -0{,}5847$$

$$P_{12} = e_1^T \cdot \begin{pmatrix} 1 \\ -1 \end{pmatrix} = 1{,}3959$$

$$P_{13} = e_1^T \begin{pmatrix} -1 \\ 2 \end{pmatrix} = -2{,}2071$$

$$P_{14} = e_1^T \begin{pmatrix} 0 \\ -1 \end{pmatrix} = 0{,}8112$$

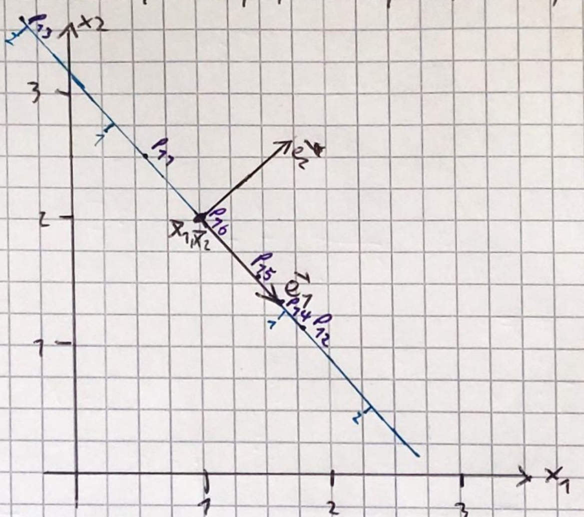
$$P_{15} = e_1^T \begin{pmatrix} 1 \\ 0 \end{pmatrix} = 0{,}5847$$

$$P_{16} = e_1^T \begin{pmatrix} 0 \\ 0 \end{pmatrix} = 0$$

The task did not ask for a dimension reduction, second component missing, -0,5P

$$\Rightarrow x' = [-0{,}5847; 1{,}3959; -2{,}2071; 0{,}8112; 0{,}5847; 0]$$

#



3/5

Exercise 15

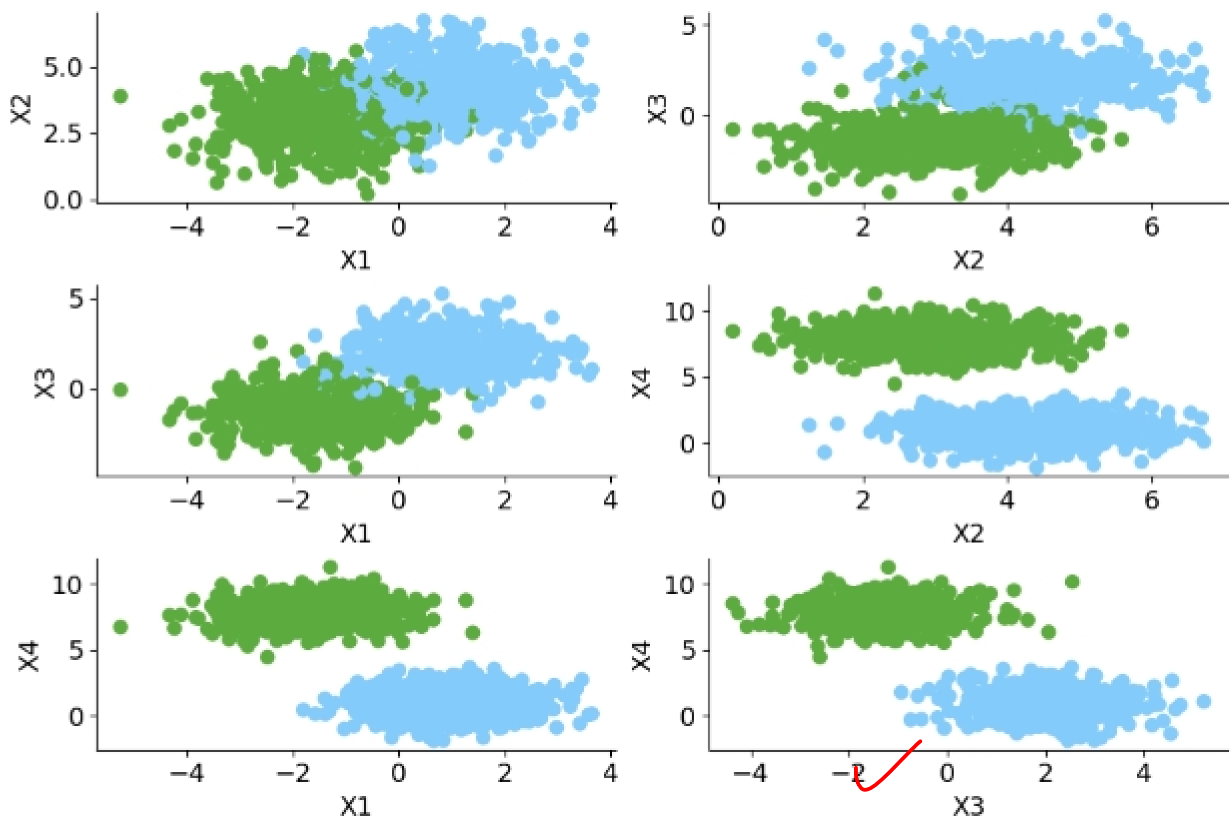
a)

This is hardly readable

In [49]:

```
from ml import plots
import matplotlib
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
import itertools

plots.set_plot_style()

%matplotlib widget

colors = plots.colors
cmap = plots.cmap

from sklearn.datasets import make_blobs

X,y = make_blobs(n_samples=1000, centers=2, n_features=4, random_state=0)

fig, ax = plt.subplots(3,2)


feature_combinations = list(itertools.combinations(range(4), 2))

plot_pos = [(0,0), (1,0), (2,0), (0,1), (1,1), (2,1)]


for i in range(len(feature_combinations)):
    ax[plot_pos[i]].scatter(X[:,feature_combinations[i][0]], X[:,feature_combinations[i][1]], c=y, s=50, cmap=cmap)
    ax[plot_pos[i]].set_xlabel('X'+str(feature_combinations[i][0]+1))
    ax[plot_pos[i]].set_ylabel('X'+str(feature_combinations[i][1]+1))
```

out [49]:



b)

In [41]:

```python
from sklearn.decomposition import PCA

k = 3

pca = PCA(n_components = k)

X_prime_sklearn = pca.fit_transform(X)


cov_mat = pca.get_covariance()

l, W = np.linalg.eigh(cov_mat)
l = l[::-1]


print('Eigenvalues', l)
```

The task did again not ask for a dimension reduction

Out [41]:

```
Eigenvalues [17.51933024  0.99958442  0.98813673  0.89875061]
```

✓ This was meant as „interpret the eigenvalues you have just computed"

The eigenvalues describe the variance of the datapoints projected on the fisher discriminant (eigenvector) in that dimension

What does the eigenvalues tell you about the „information" present in the 4 principal components?

c)

In [57]:

```python
fig, ax = plt.subplots(3)

for k in range(3):
    ax[k].hist(X_prime_sklearn[:,k])
    ax[k].set_xlabel("$x'$"+str(k+1))
    ax[k].set_ylabel('# of occurences')

None
```

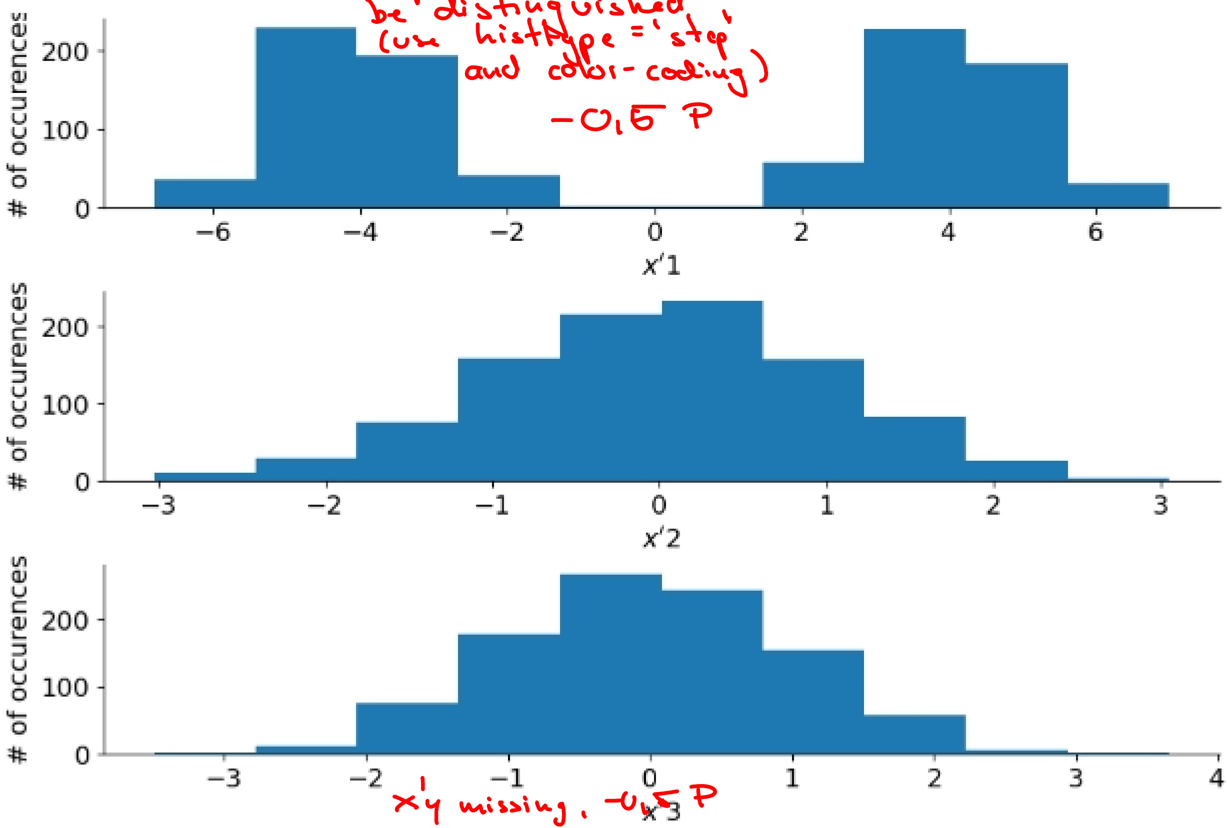
⇒ First picipal component contains nearly all information

−0,5 P

Population 1 and 2 cannot be distinguished (use histtype='step' and color-coding)

−0,5 P

out [57]:

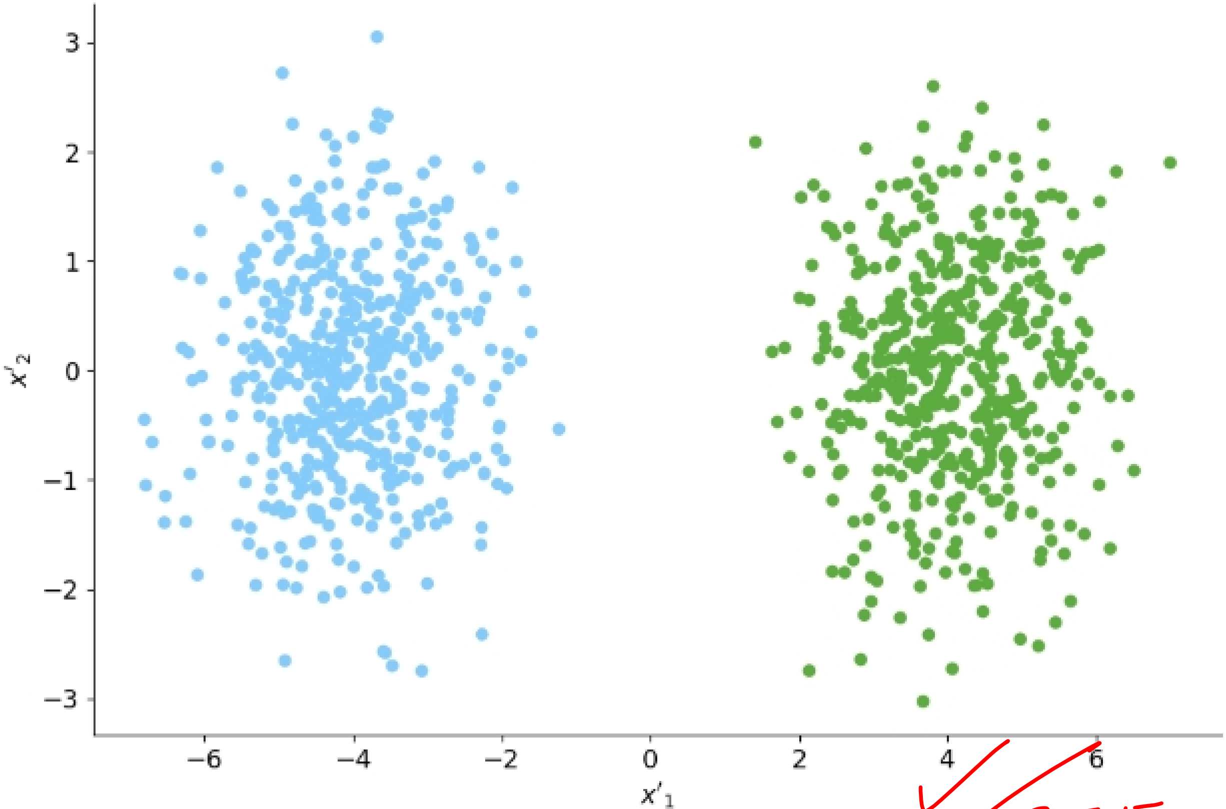


x'4 missing. −0,5 P

In [58]:

```python
plt.figure()
plt.scatter(X_prime_sklearn[:,0], X_prime_sklearn[:,1], c = y, cmap =cmap)
plt.xlabel("$x'_1$")
plt.ylabel("$x'_2$")
None
```

out [58]:



3,5 15

New Feature x'_1 clearly delivers better information than the orginal x_1, a separation between the two populations is easier now