1 Aufgabe 14

1.1 Aufgabe 14a

Scatterplot für die ersten zwei Dimensionen des Datensatzes:

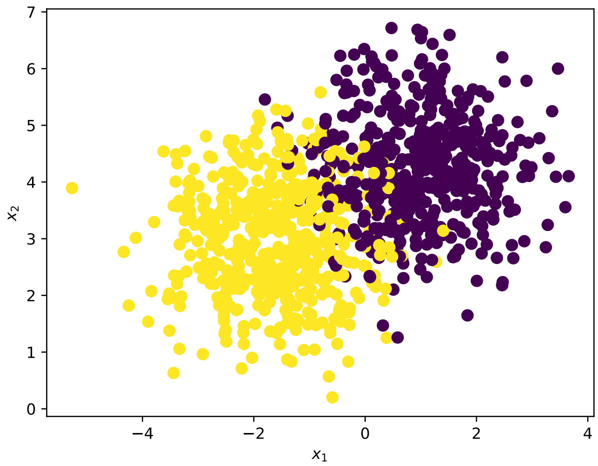

Abbildung 1: Scatterplot von x_1 und x_2.

1.2 Aufgabe 14b

Die Hauptkomponentenanalyse sucht nach einer Basis im Raum indem die Varianz entlang der Basisvektoren maximiert wird.
Gegeben seien also N Datenpunkte mit d Dimensionen.

1. Zentrierung
   a) Mittelwertvektor μ bilden.

$$\mu = \begin{pmatrix} \bar{x_1} \\ \bar{x_2} \\ \bar{x_3} \\ \bar{x_4} \end{pmatrix}$$

   b) $x_i = x_i - \mu$
2. Kovarianz
   a) Kovarianzmatrix $\mathrm{Cov}(\boldsymbol{X})$ bilden
3. Eigenwerte und Vektoren
   a) Die 4 Eigenwerte und Eigenvektoren von $\mathrm{Cov}(X)$ bestimmen, und der Größe nach sortieren.

4. Transformierung
    a) Den Datensatz $\boldsymbol{X}$ mit der Transformationsmatrix $\boldsymbol{W}$ aus den Eigenvektoren multiplizieren.

$$\boldsymbol{X} = \boldsymbol{X}\boldsymbol{W}$$

Es ergibt sich die transformierte Matrix $\boldsymbol{X}$

1.3 Aufgabe 14c

Die Eigenwerte der Kovarianzmatrix ergeben sich zu:

$$\lambda_1 = 17.519$$
$$\lambda_2 = 0.999$$
$$\lambda_3 = 0.988$$
$$\lambda_4 = 0.899$$

Es ist deutlich dass der erste Eigenwert eine wesentlich höhere Korrelation als die anderen beschreibt.

1.4 Aufgabe 14d

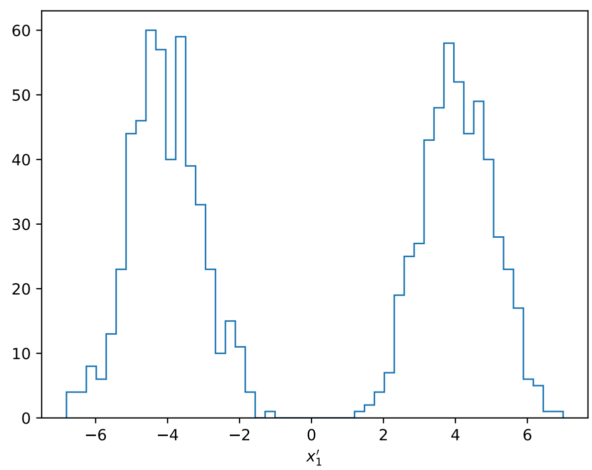

Abbildung 2: Histogramm von x_1'.

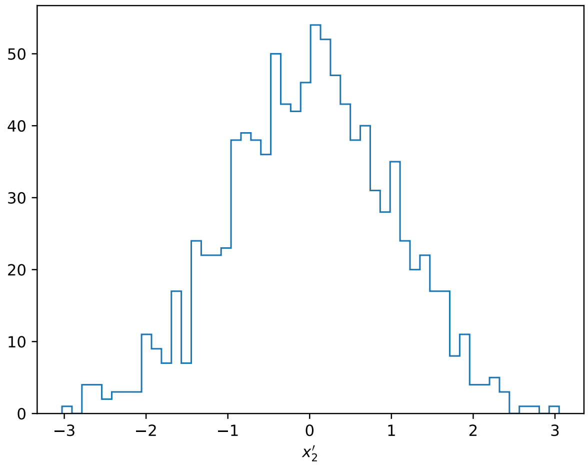

Abbildung 3: Histogramm von x_2'.

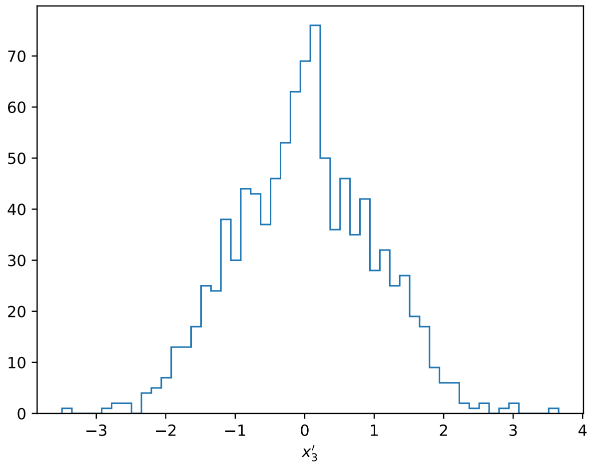

Abbildung 4: Histogramm von x_3'.

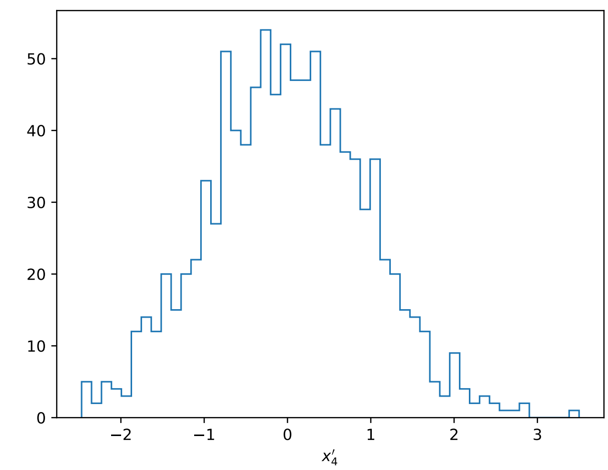

Abbildung 5: Histogramm von x_4'.

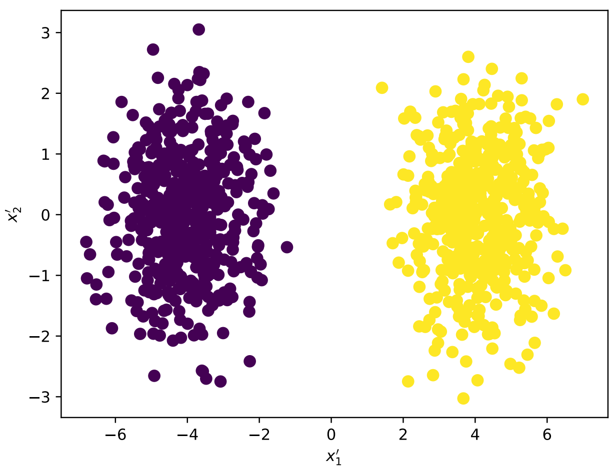

Abbildung 6: Scatterplot von x_1' und x_2'.