

Multivariate Statistik, Übung 9

HENRY HAUSTEIN

Aufgabe 1

- (a) Die Eigenwerte sind $\lambda_1 = 4$ und $\lambda_2 = 1$ ¹ und die dazugehörigen Eigenvektoren sind $v_1 = (0, 1)$ und $v_2 = (1, 0)$. Durch die erste Hauptkomponente wird dann

$$\frac{\lambda_1}{\lambda_1 + \lambda_2} = \frac{4}{4 + 1} = 80\%$$

der Streuung erklärt.

- (b) Selbiges Vorgehen hier: Die Eigenwerte sind $\lambda_1 = \lambda_2 = 1$ und die zugehörigen Eigenvektoren sind $v_1 = (1, 0)$ und $v_2 = (0, 1)$. Durch die erste Hauptkomponente wird dann

$$\frac{\lambda_1}{\lambda_1 + \lambda_2} = \frac{1}{1 + 1} = 50\%$$

der Streuung erklärt.

- (c) Ich würde die Hauptkomponentenanalyse der Varianz-Kovarianz-Matrix benutzen. Hier wird 80% der Streuung erklärt, im Gegensatz zu den 50% der Korrelationsmatrix.
- (d) Die erste Hauptkomponente ist $y_1 = (1, 0) \cdot x = 1 \cdot 2 + 0 \cdot 10 = 2$.

Aufgabe 2

- (a) Es handelt sich hier um eine 4×4 -Matrix, wir erwarten also 4 Eigenwerte.² Das charakteristische Polynom ist

$$\chi = \left((1 - \lambda)^2 - \rho_1^2 \right) \left((1 - \lambda)^2 - \rho_2^2 \right)$$

Nullsetzen des Polynoms liefert

$$\begin{aligned} (1 - \lambda)^2 - \rho_1^2 &= 0 \\ (1 - \lambda)^2 &= \rho_1^2 \\ \lambda_1 &= 1 + \rho_1 \\ \lambda_2 &= 1 - \rho_1 \end{aligned}$$

¹Wenn eine Matrix schon in Diagonalform ist, sind die Elemente auf der Hauptdiagonalen die Eigenwerte und die Einheitsvektoren sind die Eigenvektoren.

²Das muss nicht unbedingt sein; es ist durchaus möglich, dass eine $n \times n$ -Matrix weniger als n Eigenwerte hat. Allerdings sind in diesem Fall dann die sogenannten Eigenräume (der Vektorraum, aus denen der Eigenvektor kommt) mehrdimensional. Die Summe der sogenannten geometrischen Vielfachheiten (Dimension der Eigenräume) ist dann aber immer n . Eine solche Matrix lässt sich aber nicht mehr in Diagonalform bringen, das heißt die spezielle Form des Spektralsatzes aus der Vorlesung gilt dann nicht mehr. Ich denke, dass wir es deswegen in dieser Vorlesung nur mit Matrizen zu tun haben werden, die genau n Eigenwerte haben.

und

$$\begin{aligned}(1 - \lambda)^2 - \rho_2^2 &= 0 \\ (1 - \lambda)^2 &= \rho_2^2 \\ \lambda_3 &= 1 + \rho_2 \\ \lambda_4 &= 1 - \rho_2\end{aligned}$$

Wir wissen, dass $\rho_1 > \rho_2$ und weil es sich hier um eine Korrelationsmatrix handelt $\sum \lambda_i = 4$. Damit lassen sich die Eigenwerte ihrer Größe nach ordnen:

$$\lambda_1 > \lambda_2 > \lambda_4 > \lambda_3$$

Wenn mit der ersten Hauptkomponente mindestens 50% erklärt werden soll, so muss gelten:

$$\begin{aligned}\frac{\lambda_1}{4} &\geq 0.5 \\ \lambda_1 &\geq 2 \\ 1 + \rho_1 &\geq 2 \\ \rho_1 &\geq 1\end{aligned}$$

Da aber $\rho_1 \leq 1$, muss $\rho_1 = 1$ gelten. ρ_2 kann frei gewählt werden.

(b) Nun muss gelten:

$$\begin{aligned}\frac{\lambda_1}{4} + \frac{\lambda_2}{4} &\geq 0.8 \\ \lambda_1 + \lambda_2 &\geq 3.2 \\ (1 + \rho_1) + (1 + \rho_2) &\geq 3.2 \\ \rho_1 + \rho_2 &\geq 1.2\end{aligned}$$