

Principle Component Analysis

Bei der Principle Component Analysis wird eine Datenmatrix $\underline{\underline{X}}$ per linearer Transformation in ein neues Koordinatensystem überführt:

$$\underline{\underline{Z}} = \underline{\underline{X}} \cdot \underline{\underline{P}} \quad (1)$$

Die Transformationsmatrix $\underline{\underline{P}}$ wird dabei so gewählt, dass es im transformierten Datensatz $\underline{\underline{Z}}$ keine linearen Korrelationen zwischen den neuen Variablen gibt. In diesem Fall ist die Stichproben-Kovarianzmatrix $\underline{\underline{S}}_Z$ von $\underline{\underline{Z}}$ eine *Diagonalmatrix*.

Die Stichproben-Kovarianzmatrix $\underline{\underline{S}}_Z$ lässt sich für einen zentrierten Datensatz mit M Datenpunkten berechnen über:

$$\underline{\underline{S}}_Z = \frac{1}{M-1} \underline{\underline{Z}}^T \cdot \underline{\underline{Z}} \quad (2)$$

Mit Gleichung (1) folgt aus Gleichung (2):

$$\underline{\underline{S}}_Z = \frac{1}{M-1} (\underline{\underline{X}} \cdot \underline{\underline{P}})^T \cdot \underline{\underline{X}} \cdot \underline{\underline{P}} \quad (3)$$

$$\underline{\underline{S}}_Z = \frac{1}{M-1} \underline{\underline{P}}^T \cdot \underline{\underline{X}}^T \cdot \underline{\underline{X}} \cdot \underline{\underline{P}} \quad (4)$$

$$\underline{\underline{S}}_Z = \underline{\underline{P}}^T \cdot \frac{1}{M-1} \underline{\underline{X}}^T \cdot \underline{\underline{X}} \cdot \underline{\underline{P}} \quad (5)$$

$$\underline{\underline{S}}_Z = \underline{\underline{P}}^T \cdot \underline{\underline{S}}_X \cdot \underline{\underline{P}} \quad (6)$$

Es gilt: $\underline{\underline{S}}_Z$ ist genau dann eine Diagonalmatrix, wenn $\underline{\underline{P}}$ aus den (Spalten-)Eigenvektoren von $\underline{\underline{S}}_X$ besteht.

Beweis: Wir betrachten folgendes Eigenwertproblem:

$$\underline{\underline{S}}_X \cdot \underline{\underline{e}}_j = \lambda_j \underline{\underline{e}}_j \quad (7)$$

wobei $\underline{\underline{e}}_j$ ein (Spalten-)Eigenvektor von $\underline{\underline{S}}_X$ und λ_j der zugehörige Eigenwert ist.

Wir können für die Gesamtheit aller Eigenvektoren/Eigenwerte schreiben:

$$\underline{\underline{S}}_X \cdot \underline{\underline{E}} = \underline{\underline{E}} \cdot \underline{\underline{\Lambda}} \quad (8)$$

mit

$$\underline{\underline{E}} = [\underline{\underline{e}}_1, \underline{\underline{e}}_2, \dots, \underline{\underline{e}}_N] \quad (9)$$

Maschinelles Lernen in der Verfahrenstechnik

und

$$\underline{\underline{\Lambda}} = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_N \end{pmatrix} \quad (10)$$

Aus Gleichung (8) folgt durch Umstellen:

$$\underline{\underline{S}}_x = \underline{\underline{E}} \cdot \underline{\underline{\Lambda}} \cdot \underline{\underline{E}}^{-1} \quad (11)$$

Da $\underline{\underline{E}}$ orthogonal ist, können wir auch schreiben:

$$\underline{\underline{S}}_x = \underline{\underline{E}} \cdot \underline{\underline{\Lambda}} \cdot \underline{\underline{E}}^T \quad (12)$$

Einsetzen von (12) in (6) liefert:

$$\underline{\underline{S}}_z = \underline{\underline{P}}^T \cdot \underline{\underline{E}} \cdot \underline{\underline{\Lambda}} \cdot \underline{\underline{E}}^T \cdot \underline{\underline{P}} \quad (13)$$

bzw. mit $\underline{\underline{P}} = \underline{\underline{E}}$:

$$\underline{\underline{S}}_z = \underline{\underline{E}}^T \cdot \underline{\underline{E}} \cdot \underline{\underline{\Lambda}} \cdot \underline{\underline{E}}^T \cdot \underline{\underline{E}} \quad (14)$$

Da die Eigenvektoren sogar orthonormal sind, gilt $\underline{\underline{E}}^T \cdot \underline{\underline{E}} = \underline{\underline{I}}$ mit der Einheitsmatrix $\underline{\underline{I}}$ und damit:

$$\underline{\underline{S}}_z = \underline{\underline{\Lambda}} = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_N \end{pmatrix} \quad (15)$$

Damit ist in diesem Fall $\underline{\underline{S}}_z$ eine Diagonalmatrix.

Anmerkung: In der Regel werden die Eigenvektoren und zugehörigen Eigenwerte so sortiert, dass λ_1 am größten ist, λ_2 am zweitgrößten etc. Dann gibt λ_1 die Varianz des Datensatzes in Richtung der 1. Hauptkomponente an, λ_2 die Varianz des Datensatzes in Richtung der 2. Hauptkomponente etc.