Beispiel

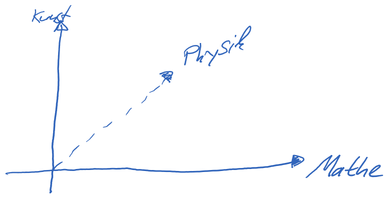
Freitag, 5. April 2019 12:24

Dataframe:

Student	Mathe	Physik	Kunst
1	90	60	90
2	90	90	30
3	60	60	60
4	60	60	90
5	30	30	30

$$X_1 = \begin{pmatrix} 90 \\ 60 \\ 90 \end{pmatrix}$$

Visualisierung in $\mathbb{R}^3$



- Welche Variablen sind korreliert?
  - Mathe, Physik?

$$\vec{A} = \begin{array}{c} \begin{array}{ccc} \text{Mathe} & \text{Physik} & \text{Kunst} \end{array} \\ \begin{pmatrix} 90 & 60 & 90 \\ 90 & 90 & 30 \\ 60 & 60 & 60 \\ 60 & 60 & 90 \\ 30 & 30 & 30 \end{pmatrix} \end{array}$$

Mean of $\vec{A}$ (mean of each column):

$$\bar{\vec{A}} = \begin{pmatrix} 66 & 60 & 60 \end{pmatrix}$$

Die Cross-Covarianz Matrix (Variance-covariance Matrix)

$$\vec{S} = \begin{array}{cc} \begin{array}{ccc} \text{Mathe} & \text{Physik} & \text{Kunst} \end{array} & \\ \begin{pmatrix} 630 & 450 & 225 \\ 450 & 450 & 0 \\ 225 & 0 & 900 \end{pmatrix} & \begin{array}{l} \text{Mathe} \\ \text{Physik} \\ \text{Kunst} \end{array} \end{array}$$

- Auf der Diagonalen ist die Variance
  - Kunst hat höhere Variance als andere Fächer

- covariance(Physik, Mathe) = 450 (Physik ↑ ⇒ Mathe ↑)

- covariance (Kunst, Mathe) = 225

- covariance (Physik, Kunst) = 0 $\Rightarrow$ keine Beziehung!

$\longrightarrow$ next step: EV und EW von $\vec{S}$

EV und EW

$$\det\left(\begin{pmatrix} 630 & 450 & 225 \\ 450 & 450 & 0 \\ 225 & 0 & 900 \end{pmatrix} - \underbrace{\lambda \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{\begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix}}\right) = \vec{0}$$

$$= \det\begin{pmatrix} 630-\lambda & 450 & 225 \\ 450 & 450-\lambda & 0 \\ 225 & 0 & 900-\lambda \end{pmatrix}$$

$$\boxed{\begin{array}{l} \det\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{matrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{matrix} \\ = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{31}a_{22}a_{13} - \ldots \end{array}}$$

$$= -\lambda^3 + 1980\lambda^2 - 1002375\lambda + 50118750 = 0$$

$$\Rightarrow \left.\begin{array}{l} \lambda_1 \approx 56 \\ \lambda_2 \approx 1137 \\ \lambda_3 \approx 786 \end{array}\right\} \text{EW}$$

$$\left.\begin{matrix} \Rightarrow EV_1 = \begin{pmatrix} 0.65 \\ -0.7 \\ -0.17 \end{pmatrix} \\ EV_2 = \begin{pmatrix} -0.65 \\ -0.43 \\ -0.62 \end{pmatrix} \\ EV_3 = \begin{pmatrix} -0.38 \\ -0.5 \\ 0.76 \end{pmatrix} \end{matrix}\right\} \left.\text{principal components}\right\} \text{sind immer orthogonal}$$

How to calculate eigenvectors:
https://www.mathebibel.de/eigenvektoren-berechnen

- λ_2 hat größten Effekt
- Die 2te principal component hat

$$\frac{\lambda_2}{\lambda_1+\lambda_2+\lambda_3} = \boxed{57\,\%}$$ der variance in den Daten

$\Rightarrow$ hat größten Einfluss auf Klausurergebnisse

- $$\frac{\lambda_3}{\lambda_1+\lambda_2+\lambda_3} = \boxed{40\,\%}$$

- $$\frac{\lambda_1}{\lambda_1+\lambda_2+\lambda_3} = \boxed{3\,\%}$$

$$\Rightarrow \vec{W} = \overbrace{\begin{pmatrix} 630 & 450 & 225 \\ 450 & 450 & 0 \\ 225 & 0 & 900 \end{pmatrix}}^{\text{cross-covarianz Matrix}} \cdot \underbrace{\begin{pmatrix} 0.65 & -0.65 & -0.38 \\ -0.7 & -0.43 & -0.5 \\ -0.17 & -0.62 & 0.76 \end{pmatrix}}_{\text{die wichtigsten EV}}$$

$\underbrace{\qquad\qquad}_{\vec{v}}$

Mittels Matrix-Multiplikation können die Daten in den Unterraum (supspace) projeziert werden:

$\boxed{\vec{W}^T \times \vec{A}}$ nicht ganz sicher

Transpose of W

$\boxed{\vec{A} \cdot \vec{v}}$

https://plot.ly/ipython-notebooks/principal-component-analysis/

Next

- watch statquest udacity } um zu verstehen wie transformieren auf subspace

- Alternativen und Anwendungen