

1 Lineare Algebra

1.1 Definitionen

Ein **Kern** ($\text{Ker}(A)$) existiert, wenn $\det(A) = 0$.

Der Kern einer Matrix A ist die Lösungsmenge von $A \cdot \vec{v} = \vec{0}$
→ LGS=0 durch elem. Zeilenoperationen lösen.

Das **Bild** ($\text{Im}(A)$) einer Matrix gibt an, welche Menge an Vektoren als Lösungen auftreten können (vgl. Wertebereich bei Funktionen).

Das Bild einer Matrix A ist die Lösungsmenge von $A \cdot \vec{v} = \vec{b}$

Der **Rang** ($\text{rank}(A)$) einer Matrix A ist die Anzahl der linear unabhängigen Spaltenvektoren

Ermittlung: Spalten von links nach rechts → ist die Spalte_i linear abhängig von den vorherigen? Rang = Anzahl der linear unabhängigen Spaltenvektoren

Verwendung z.B. zur Komprimierung von A:

$$A = \begin{pmatrix} 1 & 1 & 2 & 4 & 2 \\ 2 & 1 & 3 & 5 & 4 \\ 1 & 1 & 2 & 4 & 2 \\ 0 & 1 & 1 & 3 & 0 \end{pmatrix} \rightarrow \begin{array}{l} a_1 = 1 \cdot a_1 + 0 \cdot a_2 \\ a_2 = 0 \cdot a_1 + 1 \cdot a_2 \\ a_3 = 1 \cdot a_1 + 1 \cdot a_2 \\ a_4 = 1 \cdot a_1 + 3 \cdot a_2 \\ a_5 = 2 \cdot a_1 + 0 \cdot a_2 \end{array} \rightarrow A = \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 & 1 & 2 \\ 0 & 1 & 1 & 3 & 0 \end{pmatrix}$$

Die **Länge** eines Vektors \vec{v} ist die Wurzel aus dem Skalarprodukt mit sich selbst.

$$\rightarrow \|\vec{v}\| = \sqrt{\vec{v} \cdot \vec{v}} = \sqrt{v_1^2 + v_2^2 + \dots + v_n^2}$$

Das **Skalarprodukt** $\langle x, y \rangle$ zweier Vektoren ist die Summe der Produkte der jeweiligen Komponenten

→ $x_1y_1 + \dots + x_ny_n$; man kann damit den von x und y eingeschlossenen Winkel als Zahl $\theta \in [0, \pi]$ berechnen

mit: $\cos \theta = \frac{\langle x, y \rangle}{\|x\| \cdot \|y\|}$ → Zwei Vektoren stehen senkrecht zueinander, wenn $\langle x, y \rangle = 0$

1.2 Determinante

Spezielle Funktion, die einer quadratischen Matrix eine Zahl zuordnet. Diese gibt an, wie sich das Volumen bei der durch die Matrix beschriebenen linearen Abbildung ändert.

- Determinante = Null → Matrix ist nicht invertierbar
- $\det(I) = 1$

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$$

$$\det \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = aei + bfg + cdh - gec - hfa - ibd$$

1.3 Eigenwerte, Eigenvektoren und Eigenraum

Eine Zahl λ heißt Eigenwert der Matrix A, wenn es einen Vektor \vec{v} gibt, der nicht der Nullvektor ist, so dass gilt:

$$\begin{aligned} Av &= \lambda v \\ Av - \lambda v &= 0 \\ (A - \lambda I)v &= 0 \end{aligned}$$

1.3.1 Charakteristisches Polynom berechnen

Anstatt o.g. Gleichung zu lösen: Bestimmung der Nullstellen des charakteristischen Polynoms $p_A(\lambda)$ der Matrix A .

$$\begin{aligned} p_A(\lambda) &= \det(A - \lambda I) \\ &= \begin{vmatrix} a_{11} - \lambda & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} - \lambda \end{vmatrix} \stackrel{!}{=} 0 \end{aligned}$$

1.3.2 Eigenvektoren berechnen

Der zu einem Eigenwert λ_i gehörende Eigenvektor \vec{v}_i ist die Lösung der Gleichung:

$$\begin{aligned} A\vec{v}_i &= \lambda_i \vec{v}_i \\ (A - \lambda_i I) \cdot \vec{v}_i &= \vec{0} \end{aligned}$$

Rechenweg:

1. λ_i für λ in die Matrix $(A - \lambda I)$ einsetzen (siehe charakteristisches Polynom)
2. Das folgende LGS durch elementare Zeilenoperationen lösen:

$$\left(\begin{array}{ccc|c} a_{11} - \lambda & \cdots & a_{1n} & 0 \\ \vdots & \ddots & \vdots & 0 \\ a_{n1} & \cdots & a_{nn} - \lambda & 0 \end{array} \right)$$

3. Für Nullzeilen ergeben sich beliebige Lösungen, die gleich 1 gesetzt werden können.

1.3.3 Eigenraum berechnen

Der Eigenraum $E_A(\lambda_i)$ einer Matrix A zu einem Eigenwert λ_i ist die Menge aller Eigenvektoren \vec{v}_i zu λ_i .

Lösung: Vielfaches der Eigenvektoren in Mengenschreibweise festhalten:

$$E_A(\lambda_i) = \{k \cdot \vec{v}_i | k \in \mathbb{R}\}$$

1.3.4 algebraische vs. geometrische Vielfachheit von λ

- **algebraische Vielfachheit:** Anzahl gleicher Eigenwerte im charakteristischen Polynom
- **geometrische Vielfachheit:** Dimension (Anzahl der Vektoren) des Eigenraums $E(\lambda)$; \leq algebraische Vielfachheit

1.4 Orthogonale Matrizen

Zwei Vektoren sind orthogonal, wenn ihr **Skalarprodukt**

$$\langle a, b \rangle = a_1 b_1 + \dots + a_n b_n = 0$$

Äquivalente Aussagen:

- Matrix B ist orthogonal
- $B^T B = I$, d.h. B ist invertierbar mit $B^{-1} = B^T$.
- Die Spaltenvektoren von B definieren eine Orthonormalbasis von \mathbb{R}^n

1.4.1 Orthogonalen Vektor mit dem Kreuzprodukt finden

Für $\vec{a} \perp \vec{b}$ ergibt sich \vec{c} mit $\vec{c} \perp \vec{a}, \vec{c} \perp \vec{b}$ aus:

$$\vec{c} = \vec{a} \times \vec{b} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \times \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} a_2 b_3 - a_3 b_2 \\ a_3 b_1 - a_1 b_3 \\ a_1 b_2 - a_2 b_1 \end{pmatrix}$$

1.4.2 Gram-Schmidt-Verfahren

Ziel: Orthonormalbasis (ONB) zu einem Vektorraum $B = \{b_1, b_2, \dots, b_n\}$ finden.

1. Ersten Basisvektor normieren: $\vec{q}_1 = \frac{\vec{q}_1}{\|\vec{q}_1\|}$
2. Falle das Lot von b_2 auf die von q_1 erzeugte Gerade: $l_2 = b_2 - \langle b_2, q_1 \rangle q_1$
3. Normiere das Lot: $\vec{q}_2 = \frac{l_2}{\|l_2\|}$
4. Wiederhole Schritte 2 und 3 fur alle Basisvektoren:
 $l_i = b_i - \langle b_i, q_1 \rangle q_1 - \langle b_i, q_2 \rangle q_2 - \dots - \langle b_i, q_{i-1} \rangle q_{i-1}$ und $\vec{q}_i = \frac{l_i}{\|l_i\|}$

1.5 Diagonalisierbarkeit

1.5.1 Diagonalisierbarkeit

A ist diagonalisierbar, wenn

- fur jeden Eigenwert von A die algebraische Vielfachheit gleich der geometrischen Vielfachheit ist, oder
- wenn alle Eigenwerte (λ_i) von A unterschiedlich sind.

Um die Diagonalmatrix $D = S^{-1}AS$ bzw. $A = SDS^{-1}$ zu bestimmen:

1. Eigenwerte λ_i von A bestimmen \rightarrow *Nullstellen char. Polynom*
2. Eigenvektoren \vec{v}_i zu λ_i bestimmen \rightarrow *Spalten der Matrix S*
3. Diagonalmatrix $D = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_i)$ bestimmen

1.5.2 Orthogonale Diagonalisierbarkeit

Eine Matrix $A \in \mathbb{R}^{n \times n}$ heit orthogonal diagonalisierbar, falls es eine orthogonale Matrix $S \in \mathbb{R}^{n \times n}$ gibt, so dass $D = S^T A S = S^{-1} A S$ eine Diagonalmatrix ist ($S^T S = I \Rightarrow$ Orthogonalitat $S^{-1} = S^T$).

Dies ist genau dann der Fall, wenn A symmetrisch ist:

$$A^T = (SDS^T)^T = (S^T)^T D^T S^T = SDS^T = A$$

Vorgehensweise analog zur Diagonalisierbarkeit; zusatzlich mussen die Eigenvektoren \vec{v}_i zu λ_i noch normiert werden ($\vec{v}_i = \frac{v_i}{\|v_i\|}$)

1.6 Pseudo-Inverse A^+

Approximation einer inversen Matrix Fur nicht-quadratische Matrizen mit Hilfe der Singularwertzerlegung (siehe 1.7).

$$A^+ = V \cdot \Sigma^{-1} \cdot U^T$$

wobei $\Sigma^{-1} = \text{diag}(\sigma_1^{-1}, \dots, \sigma_r^{-1})$

Eigenschaften:

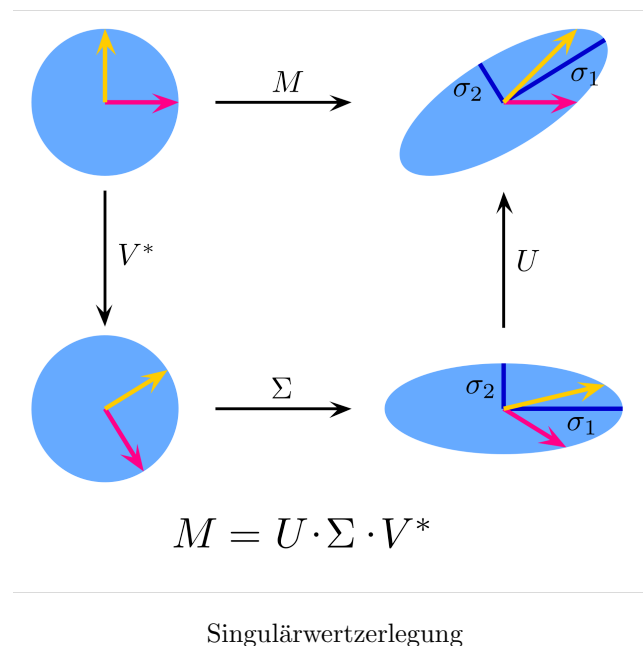
- $AA^+A = A$
- $A^+AA^+ = A^+$
- $(AA^+)^T = AA^+ \rightarrow AA^+$ ist symmetrisch

- $(A^+A)^T = A^+A \rightarrow A^+A$ ist symmetrisch
- $A^+ = A^{-1}$, wenn A invertierbar ist
- $A = U\Sigma V^T \Leftrightarrow A^T = V\Sigma U^T$
- $V^TV = VV^T = I$ und $U^TU = UU^T = I$

1.7 Singulärwertzerlegung

$$\underbrace{A}_{\mathbb{R}^{m \times n}} = \underbrace{U}_{\mathbb{R}^{m \times m}} \underbrace{\Sigma}_{\mathbb{R}^{m \times n}} \underbrace{V^T}_{\mathbb{R}^{n \times n}}$$

- U und V sind orthogonale/unitäre Matrizen
- U enthält die normierten Eigenvektoren von AA^T ; kann als eine Basis für den Spaltenraum von A betrachtet werden
- V enthält die normierten Eigenvektoren von A^TA ; kann als eine Basis für den Zeilenraum von A betrachtet werden
- Σ ist eine Diagonalmatrix mit den Singulärwerten $\sigma_1 \geq \sigma_2 \geq \dots \geq 0$ (sortiert) auf der Hauptdiagonalen. Die Singulärwerte sind die Wurzeln der Eigenwerte von A^TA und AA^T ($\sigma_i = \sqrt{\lambda_i}$, Rest = 0).
- Die Singulärwerte in Σ geben die Stärke der Korrelation zwischen den Spalten und Zeilen von A an. Die größten Singulärwerte in Σ geben die wichtigsten Merkmale von A an, während die kleinsten Singulärwerte in Σ die Rauschkomponenten von A darstellen.



1.7.1 Einfaches Berechnungsverfahren (über Eigenvektoren)

1. AA^T und A^TA bestimmen
2. Für „kleinere“ Matrix aus 1) Eigenwerte λ_i bestimmen (char. Polynom)
3. Σ mit $\text{diag}(\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n)$ aufstellen; Singulärwerte $\sigma_i = \sqrt{\lambda_i}$ auf Hauptdiagonalen; Rest = 0
4. U aufstellen: normierte Eigenvektoren für AA^T für alle λ_i bestimmen
5. V aufstellen: normierte Eigenvektoren für A^TA für alle λ_i bestimmen; V^T bilden

1.7.2 Alternatives Berechnungsverfahren

1. Form prüfen: ist A „hochkant“?

→ sonst aufwendiger zu lösen

Umstellen zu A^T ist möglich, da

$$(A^T)^T = A; \text{ d.h. } A^T = V \Sigma^T U^T$$

2. Eigenwerte von $A^T A$ bestimmen

Eigenwerte (≥ 0) über *Nullstellen char.*

Polynom bestimmen, absteigend sortieren!

3. Σ aufstellen

Diagonalmatrix mit $\sigma_i = \sqrt{\lambda_i}$, Rest = 0

4. Spaltenvektoren für V ermitteln

Eigenvektoren zu λ aus 2. bestimmen,

normieren und in Matrix V eintragen

Für SVD: V^T bilden

5. U aufstellen

a) für vorhandene Singulärwerte:

$$u_i = \frac{1}{\sigma_i} A v_i$$

b) sonst: u_i so finden, dass u_i ONB sind

→ Kreuzprodukt (\mathbb{R}^3)

→ Gram-Schmidt

$$A = \begin{pmatrix} 1 & 0 \\ 2 & 2 \\ 0 & 1 \end{pmatrix}$$

$$A^T A = \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}$$

$$p_A(\lambda) = \det(A^T A - \lambda I) \stackrel{!}{=} 0$$

$$(5 - \lambda)^2 - 16 = 0$$

$$\lambda_1 = 9, \lambda_2 = 1$$

$$\Sigma = \begin{pmatrix} 3 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$$

für $\lambda_1 = 9$:

$$\left(\begin{array}{cc|c} 5-9 & 4 & 0 \\ 4 & 5-9 & 0 \end{array} \right) \Leftrightarrow \left(\begin{array}{cc|c} 1 & -1 & 0 \\ 0 & 0 & 0 \end{array} \right) \Rightarrow v_1^* = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$v_1 = \frac{v_1^*}{\|v_1^*\|} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

für $\lambda_2 = 1$:

...

$$v_2 = \frac{v_2^*}{\|v_2^*\|} = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

$$V = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$$

$$u_1 = \frac{1}{3\sqrt{2}} \begin{pmatrix} 1 \\ 4 \\ 1 \end{pmatrix}, u_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$$

$$\text{b) } u_3 = u_1 \times u_2 = \frac{1}{6} \begin{pmatrix} 4 \\ -2 \\ 4 \end{pmatrix}$$

2 Mehrdimensionale Wahrscheinlichkeitsrechnung

2.1 Formeln

2.1.1 Kovarianz, Korrelation

$$\text{Cov}[X, Y] = E[(X - E[X])(Y - E[Y])]$$

$$\text{Cov}[X, Y] = E[XY] - E[X]E[Y]$$

$$\text{Cov}[X, Y] = \text{Cov}[Y, X] \text{ und } \text{Cov}[X, X] = \text{Var}[X]$$

$$r_{XY} = \text{Cor}[X, Y] = \frac{\text{Cov}[X, Y]}{\sqrt{\text{Var}(X)\text{Var}(Y)}} \rightarrow [-1; +1]$$

Merke:

- statistische Unabhängigkeit \Rightarrow Unkorreliertheit
- Unkorreliertheit \Rightarrow statistische Unabhängigkeit, nur wenn X und Y normalverteilt sind

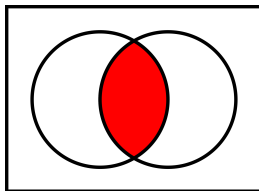
2.1.2 Linearkombinationen

$$E[aX + bY] = aE[X] + bE[Y]$$

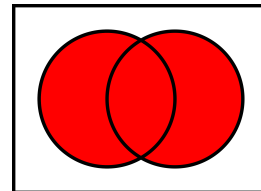
$$\text{Var}[aX + bY] = a^2\text{Var}[X] + b^2\text{Var}[Y] + 2ab\text{Cov}[X, Y]$$

$$\text{Var}[X_1 + \dots + X_n] = \sum_{i=1}^n \text{Var}[X_i] + 2 \sum_{1 \leq i < j \leq n} \text{Cov}[X_i, X_j] \rightarrow \text{Satz von Bienaymé}$$

2.1.3 Mengen



Schnittmenge $A \cap B$



Vereinigung $A \cup B$

$$A \cup B = A + B - A \cap B$$

2.1.4 Konvergenz

- in Wahrscheinlichkeit: $X_n \xrightarrow{P} X$ wenn $\forall \epsilon > 0 : \lim_{n \rightarrow \infty} P(|X_n - X| > \epsilon) = 0$
- im p-ten Mittel/in \mathcal{L}^p : $X_n \xrightarrow{\mathcal{L}^p} X$ wenn $\lim_{n \rightarrow \infty} E(|X_n - X|^p) = 0$
- fast sicher: $X_n \xrightarrow{fs} X$ wenn $P(\lim_{n \rightarrow \infty} X_n = X) = 1$

2.2 Bedingte Wahrscheinlichkeit

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

2.3 Bayes-Theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

3 Optimierung

3.1 Konvexe Funktionen/Mengen

Eine Menge $G \subseteq \mathbb{R}^n$ heißt *konvex*, wenn für alle $x, y \in G$ und $t \in [0, 1]$ gilt:

$$tx + (1 - t)y \in G$$

D.h. die Verbindungsstrecke zwischen zwei Punkten der Menge liegt komplett in der Menge.

Eine Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$ heißt *konvex* (\leq) bzw. *strikt konvex* ($<$), wenn für alle $x, y \in \mathbb{R}^n$ und $t \in [0, 1]$ gilt:

$$f(tx + (1 - t)y) \leq tf(x) + (1 - t)f(y)$$

Eine Funktion $f(x)$ ist (strikt) konvex, wenn $f''(x)$ überall ≥ 0 (bzw. > 0) ist.

3.1.1 Vorgehensweise bei mehrdimensionalen Funktionen:

1. Hesse-Matrix ($H_f(x)$, =symmetrisch) bestimmen: $H_f(x) = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix}$
2. Definitheit bestimmen:
 - (a) über *Eigenwerte*
 - Nullstellen charakteristisches Polynom bestimmen (\rightarrow 1.3.1)
 - Interpretation:
 - alle $\lambda > 0 \rightarrow H_f(x) = \text{pos. definit}$
 - alle $\lambda \geq 0 \rightarrow H_f(x) = \text{pos. semidefinit}$
 - alle $\lambda < 0 \rightarrow H_f(x) = \text{neg. definit}$
 - alle $\lambda \leq 0 \rightarrow H_f(x) = \text{neg. semidefinit}$
 - λ positiv und negativ $\rightarrow H_f(x) = \text{indefinit}$
 - (b) über *Diagonaldominanz*: Ist $H_f(x)$ Diagonaldominant und alle Diagonalelemente > 0 , so ist $H_f(x)$ positiv definit.
 \rightarrow für alle Zeilen: $||\text{Diagonalelement}|| > \sum ||\text{übrige Zeilenelemente}||$
 - (c) *Choleskyzerlegung* ist möglich = $H_f(x)$ ist positiv definit
3. Konvexität bestimmen:
 - $H_f(x)$ positiv definit $\Leftrightarrow f$ strikt konvex
 - $H_f(x)$ positiv semidefinit $\Leftrightarrow f$ konvex
 - $H_f(x)$ negativ definit $\Leftrightarrow f$ strikt konkav
 - $H_f(x)$ negativ semidefinit $\Leftrightarrow f$ konkav

3.1.2 L-glatt und Lipschitz-stetig

Eine Funktion $f(x)$ heißt *Lipschitz-stetig*, wenn $||f(x) - f(y)|| \leq L||x - y||$ für alle x, y gilt.

Eine Lipschitz-stetige Funktion ist eine stetige Funktion, deren Steigung beschränkt ist:

$$\left\| \frac{f(x) - f(y)}{x - y} \right\| \leq L \rightarrow \text{Jede Sekantensteigung} \leq L$$

Implikation: Wenn zwei eingesetzte Punkte (x, y) näher zusammenrücken, dann nähern sich auch die Funktionswerte $f(x)$ und $f(y)$ an.

Eine Funktion $f(x)$ heißt *L-glatt*, wenn f differenzierbar ist und $||\nabla f(x)|| \leq L$ für alle x gilt bzw. falls der Gradient L-Lipschitz-stetig ist, d.h. wenn

$$||\nabla f(x) - \nabla f(y)|| \leq L||x - y||$$

3.2 Optimierung ohne NB - Gradientenverfahren

Beginnend an einer Stelle x_0 :

1. Berechne Gradienten $\nabla f(x_n)$ (im ersten Schritt mit x_0)
2. Setze $x_{n+1} = x_n - \alpha \nabla f(x_n)$, $n \geq 0$ mit Schrittweite α
3. Wiederhole Schritt 1 und 2 bis Abbruchkriterium erfüllt; $x^* = x_k$

Approximation von $\nabla f(x_0)$ durch $\frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}}$ möglich.

3.3 Optimierung unter Nebenbedingungen (KKT)

Erweiterung des Lagrange-Verfahrens um Nebenbedingungen. Es gelten folgende KKT-Bedingungen:

- (I) $g_i(x^*) \leq 0 \rightarrow$ Ungleichungs-NB nach 0 umformen
- (II) $h_j(x^*) = 0 \rightarrow$ Gleichungs-NB nach 0 umformen
- (III) $\lambda_i^* \geq 0$
- (IV) $\lambda_i^* g_i(x^*) = 0$
- (V) $\mathcal{L}'(x, \lambda, v) = \nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla g_i(x^*) + \sum_{j=1}^p v_j^* \nabla h_j(x^*) = 0$

mit den Lagrangemultiplikatoren (λ, v) als *duale Variablen* und x als *primale Variable*.

Vorgehensweise:

- Gleichungs-NB vorhanden \rightarrow (II) nach einer Variablen auflösen und in (I) und (V) einsetzen
- 1. Ableitungen (Gradient ∇) von (I), (II) und Zielfunktion $f(x^*)$ bilden und (V) aufstellen
- (IV) aufstellen und Fallgruppen bilden: Was muss für λ_i und $g_i(x^*)$ gelten, damit Produkt = 0 wird? Fallgruppen über Kreuz bilden!
- Fallgruppen in (V) einsetzen und prüfen: Ist Lösung möglich? Sind alle Bedingungen erfüllt?
 - Wenn ja: KKT-Punkt = Lösung gefunden
 - Wenn nein: Kombination nicht zulässig \rightarrow verwerfen!
- Lösung aufschreiben: KKT-Punkt (x, y) , Zielfunktionswert (p^*) , duale Variablen (Lagrange-Multiplikatoren λ_i, v_j)

Beispiel:

$$\begin{aligned} p^* &= \min 3x^2 + y^2 \\ \text{unter } x - y &\leq -8 \\ -y &\leq 0 \end{aligned}$$

1. $g_1(x^*) = x - y + 8 \leq 0 \rightarrow \nabla g_1(x^*) = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$
 $g_2(x^*) = -y \leq 0 \rightarrow \nabla g_2(x^*) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$
 $\nabla f(x^*) = \begin{pmatrix} 6x \\ 2y \end{pmatrix}$
2. (V): $\begin{pmatrix} 6x \\ 2y \end{pmatrix} + \lambda_1^* \begin{pmatrix} 1 \\ -1 \end{pmatrix} + \lambda_2^* \begin{pmatrix} 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$$3. \lambda_1^* g_1(x^*) = \lambda_1^*(x - y + 8) = 0 \rightarrow \underline{\lambda_1^* = 0} \text{ und } x - y + 8 = 0 \Leftrightarrow \underline{x = y - 8}$$

$$\lambda_2^* g_2(x^*) = -\lambda_2^* y = 0 \rightarrow \underline{\lambda_2^* = 0} \text{ und } \underline{y = 0}$$

4. Fallkombinationen prüfen:

- Fall 1: $\lambda_1^* = \lambda_2^* = 0 \rightarrow \begin{pmatrix} 6x \\ 2y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Leftrightarrow x = y = 0 \nrightarrow$ wegen $x - y \leq -8$
- Fall 2: $\lambda_1^* = y = 0 \rightarrow \begin{pmatrix} 6x \\ 0 \end{pmatrix} + \lambda_2^* \begin{pmatrix} 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Leftrightarrow x = \lambda_2^* = 0 \nrightarrow$ wegen $x - y \leq -8$
- Fall 3: $x = y - 8; \lambda_2^* = 0 \rightarrow \begin{pmatrix} 6(y - 8) \\ 2y \end{pmatrix} + \lambda_1^* \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$$6y - 48 + \lambda_1^* = 0$$

$$2y - \lambda_1^* = 0$$

$$8y = 48 \Leftrightarrow y = 6 \checkmark$$

$$x = y - 8 \Leftrightarrow x = -2 \checkmark$$

$$\lambda_1^* = 2y = 12 \checkmark$$

Es handelt sich um einen KKT-Punkt

- Fall 4: $x = y - 8; y = 0 \rightarrow \begin{pmatrix} 6(-8) \\ 0 \end{pmatrix} + \lambda_1^* \begin{pmatrix} 1 \\ -1 \end{pmatrix} + \lambda_2^* \begin{pmatrix} 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Leftrightarrow \lambda_1^* = 48 \text{ und } -\lambda_1^* - \lambda_2^* = 0 \Leftrightarrow \lambda_2^* = -48 \nrightarrow$ wegen (III)

5. Lösung: Der Punkt $(x^*, y^*, \lambda_1^*, \lambda_2^*) = (-2, 6, 12, 0)$ erfüllt die KKT-Bedingungen (KKT-Punkt). $(x, y) = (-2, 6)$ löst das *primale Problem* und $(\lambda_1, \lambda_2) = (12, 0)$ löst das *duale Problem*. Damit ist $p^* = 3 \cdot (-2)^2 + 6^2 = 48$

3.4 Support Vector Machines (SVM)

To be done

4 Statistik

4.1 Verteilungen

4.1.1 Binomialverteilung ($X \sim \text{Bin}(n, \pi)$)

Diskrete Wahrscheinlichkeitsverteilung; beschreibt die Anzahl an Erfolgen in einer Serie von unabhängigen Versuchen, die jeweils genau zwei Ergebnisse haben (Erfolg/Misserfolg); z.B. beim Münzwurf, Ziehen mit Zurücklegen.

- n =Anzahl der Versuche/Ziehungen, π =Erfolgswahrscheinlichkeit (z.B. 0.3)
- **Wahrscheinlichkeit:** $P(X = k) = \binom{n}{k} \pi^k (1 - \pi)^{n-k}$; k = Anzahl der gewünschten Erfolge
- **Erwartungswert:** $E(X) = n\pi$
- **Varianz:** $\text{Var}(X) = n\pi(1 - \pi)$
- **Normalapproximation:** $X \sim N(n\pi, n\pi(1 - \pi))$
- **Binomialkoeffizient:** $\binom{n}{k} = \frac{n!}{k!(n-k)!}$
- **Bernoulli-Verteilung:** Spezialfall der Binomialverteilung mit $n = 1$

4.1.2 Poisson-Verteilung ($X \sim \text{Poi}(\mu)$)

Diskrete Wahrscheinlichkeitsverteilung; beschreibt die Anzahl an Ereignissen in einem festgelegten Zeitintervall, wenn die Ereignisse mit einer konstanten Rate und unabhängig von der Zeit auftreten; z.B. Anzahl der Anrufe in einer Stunde, Anzahl der Kunden in einer Schlange, Anzahl der Fehler in einem Text. Für kleine μ zeigt die Poisson-Verteilung eine starke Asymmetrie (Rechtsschiefe).

- Θ =Erwartungswert (z.B. 0.3)
- **Wahrscheinlichkeit:** $P(X = n) = \frac{\Theta^n}{n!} e^{-\Theta}$; n = Anzahl der gewünschten Ereignisse
- **Erwartungswert:** $E(X) = \Theta$
- **Varianz:** $\text{Var}(X) = \Theta$
- **Normalapproximation:** $X \sim N(\Theta, \Theta)$

4.1.3 Hypergeometrische Verteilung ($X \sim H(n, N, m)$)

Ähnlich wie Binomialverteilung, aber ohne Zurücklegen; z.B. beim Ziehen ohne Zurücklegen aus einer Urne mit N Kugeln, davon m mit Erfolgsmarkierung.

- n =Anzahl der Versuche/Ziehungen, N =Anzahl der Kugeln in der Urne, m =Anzahl der Kugeln mit Erfolgsmarkierung
- **Wahrscheinlichkeit:** $P(X = k) = \frac{\binom{m}{k} \binom{N-m}{n-k}}{\binom{N}{n}}$; k = Anzahl der gewünschten Erfolge
- **Erwartungswert:** $E(X) = n \frac{m}{N}$
- **Varianz:** $\text{Var}(X) = n \frac{m}{N} \left(1 - \frac{m}{N}\right) \left(\frac{N-n}{N-1}\right)$
- **Normalapproximation:** $X \sim N\left(n \frac{m}{N}, n \frac{m}{N} \left(1 - \frac{m}{N}\right) \left(\frac{N-n}{N-1}\right)\right)$

4.1.4 Normalverteilung ($X \sim N(\mu, \sigma^2)$)

Stetige Wahrscheinlichkeitsverteilung; beschreibt viele natürliche Vorgänge (z.B. Körpergröße, IQ, Fehler in Messungen); zentrales Grenzwerttheorem: Summe von unabhängigen Zufallsvariablen strebt gegen Normalverteilung; symmetrisch um μ , σ^2 -bestimmte Breite; μ =Erwartungswert, σ^2 =Varianz, σ =Standardabweichung.

- **Erwartungswert:** $E(X) = \mu$
- **Varianz:** $\text{Var}(X) = \sigma^2$
- **Standardnormalverteilung:** $Z \sim N(0, 1)$

4.2 Parameterschätzung

Zentrale Größen:

- Kovarianz: $cov(X, Y) = E[(X - E[X])(Y - E[Y])] = E(XY) - E(X)E(Y)$
- Korrelation: $corr(X, Y) = \frac{cov(X, Y)}{\sigma_X \sigma_Y}$ (liegt zwischen -1 und +1)

Kriterien für gute Schätzer:

- **Konsistenz:** Schätzungen werden genauer, je größer die Stichprobe ist
- **Erwartungstreue/Unverzerrtheit/unbiased:** Schätzer liegt im Mittel richtig.

4.2.1 Maximum-Likelihood Schätzer (ML-Schätzer)

- **Binomialverteilung** ($X \sim \text{Bin}(n, \pi)$, n =Länge der Versuchsreihe, π =Wahrscheinlichkeit für Erfolg):
 - Erwartungswert $\hat{\pi} = T(x) = \frac{x}{n} \rightarrow$ Anzahl Erfolge / Anzahl Versuche
 - Varianz $\hat{\sigma}^2 = \frac{\pi(1-\pi)}{n} \rightarrow$ ggf. mit $\hat{\pi}$ rechnen
- **Normalverteilung** ($X \sim N(\mu, \sigma^2)$, μ =Erwartungswert, σ^2 =Varianz):
 - Erwartungswert $\hat{\mu} = T(x) = \frac{1}{n} \sum_{i=1}^n x_i \rightarrow$ arithmetisches Mittel
 - Varianz $S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu})^2 \rightarrow$ empirische Varianz
 - korrigierte Stichprobenvarianz $S^{*2} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \hat{\mu})^2$

Beispiel für Herleitung ($X \sim N(\mu, \sigma)$):

1. Aufstellen der ML-Funktion: $L(\mu, \sigma^2) = \prod_{i=1}^n f(x_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}}$
2. Logarithmierung:

$$\begin{aligned} \ln(L(\mu, \sigma^2)) &= \sum_{i=1}^n \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}}\right) \\ &= \sum_{i=1}^n \left[\ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{(x_i - \mu)^2}{2\sigma^2} \right] \\ &= -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \end{aligned}$$

3. Ableiten nach μ und σ^2 und Nullsetzen:

$$\begin{aligned}
\frac{\partial \ln(L(\mu, \sigma^2))}{\partial \mu} &= \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0 \\
&\Rightarrow \sum_{i=1}^n (x_i - \mu) = 0 \\
&\Rightarrow \sum_{i=1}^n x_i - \sum_{i=1}^n \mu = 0 \\
&\Rightarrow \sum_{i=1}^n x_i - n\mu = 0 \\
&\Rightarrow \sum_{i=1}^n x_i = n\mu \\
&\Rightarrow \mu = \frac{1}{n} \sum_{i=1}^n x_i \\
\frac{\partial \ln(L(\mu, \sigma^2))}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^n (x_i - \mu)^2 = 0 \\
&\Rightarrow \frac{n}{2\sigma^2} = \frac{1}{2(\sigma^2)^2} \sum_{i=1}^n (x_i - \mu)^2 \\
&\Rightarrow \frac{n}{2} = \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \\
&\Rightarrow \sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 \\
&\Rightarrow \sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2}
\end{aligned}$$

4.2.2 Bayes-Schätzer

4.3 Konfidenzintervalle

Schätzung eines Intervalls, in dem sich der wahre Wert (z.B. der Erwartungswert μ) mit einer gewissen Wahrscheinlichkeit befindet.

4.3.1 Vorgehensweise bei Normalverteilung und bekanntem σ

1. Punktschätzung des Erwartungswerts aus n Stichproben (x_i)

$$M(x) = \frac{1}{n} \sum_{i=1}^n x_i$$

→ dieser entspricht i.d.R. nicht dem wahren Wert μ der Grundgesamtheit.

2. Konfidenzniveau ($1 - \alpha$; α = Irrtumsniveau) festlegen und $z_{1-\frac{\alpha}{2}}$ aus Tabelle zur Normalverteilung ablesen

- 90% $\rightarrow 1 - \alpha \rightarrow \alpha = 0.1 \rightarrow z_{0.95} = 1.65$
- 95% $\rightarrow \alpha = 0.05 \rightarrow z_{0.975} = 1.96$
- 96% $\rightarrow \alpha = 0.04 \rightarrow z_{0.980} = 2.06$
- 97% $\rightarrow \alpha = 0.03 \rightarrow z_{0.985} = 2.17$
- 98% $\rightarrow \alpha = 0.02 \rightarrow z_{0.99} = 2.33$
- 99% $\rightarrow \alpha = 0.01 \rightarrow z_{0.995} = 2.58$

3. Berechnung des Konfidenzintervalls

$$\mathcal{I}(x) = \left[M(x) - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} ; M(x) + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \right]$$

ist ein Konfidenzintervall zum Sicherheitsniveau $1 - \alpha$.

Vorgehensweise: $1 - \frac{\alpha}{2}$ berechnen und innerhalb der Tabelle zur Normalverteilung diesen Wert suchen. Der gesuchte z -Wert ergibt sich dann aus den Zeilen- und Spaltenberschriften.

4.3.2 Vorgehensweise bei Normalverteilung und unbekanntem σ

Grds. analog zu oben, wobei Werte für $t_{n-1; 1-\frac{\alpha}{2}}$ aus der t -Verteilung verwendet werden (Studentsche (t -) Verteilung mit $n - 1$ Freiheitsgraden). Nur bei $n \geq 30$.

1. Punktschätzung des Erwartungswerts aus n Stichproben (x_i)

$$M(x) = \frac{1}{n} \sum_{i=1}^n x_i$$

→ dieser entspricht i.d.R. nicht dem wahren Wert μ der Grundgesamtheit.

2. Berechnung des Standardfehlers der Stichprobenmittelwerte

$$\text{korrigierte Stichprobenvarianz: } V^*(x) = \frac{1}{n-1} \sum_{i=1}^n (x_i - M(x))^2$$

$$\text{Standardfehler: } s^* = \sqrt{\frac{V^*(x)}{n}}$$

→ dieser entspricht i.d.R. nicht dem wahren Wert σ der Grundgesamtheit.

3. Berechnung des Konfidenzintervalls

$$\mathcal{I}(x) = [M(x) - t_{n-1; 1-\frac{\alpha}{2}} \cdot s^* ; M(x) + t_{n-1; 1-\frac{\alpha}{2}} \cdot s^*]$$

ist ein Konfidenzintervall zum Sicherheitsniveau $1 - \alpha$ mit $n - 1$ Freiheitsgraden

Vorgehensweise: Richtige Zeile für Freiheitsgrade $n - 1$ suchen. In Spalte $1 - \frac{\alpha}{2}$ suchen. Der gewünschte t -Wert ergibt sich aus der Tabelle.

4.3.3 Vorgehensweise im Binominalmodell

Wenn X binominalverteilt ist ($X \sim B(n, p)$, n =Anzahl gezogene Versuche, p =Erfolgswahrscheinlichkeit), n groß und die Varianz nicht zu klein ist (Faustregel: $np(1-p) > 9$), gilt die Approximation durch die Normalverteilung mit:

- Erwartungswert: $\mu = np$
- Standardabweichung: $\sigma = \sqrt{np(1-p)}$
- Standardfehler: $s = \sqrt{\frac{p(1-p)}{n}}$
- Punktschätzung: $\hat{p} = \frac{x}{n}$
- Konfidenzintervall für p : $\mathcal{I}(p) \approx [\hat{p} - z_{1-\frac{\alpha}{2}} \cdot s; \hat{p} + z_{1-\frac{\alpha}{2}} \cdot s]$
- Konfidenzintervall für μ : $\mathcal{I}(\mu) \approx [\hat{\mu} - z_{1-\frac{\alpha}{2}} \cdot s; \hat{\mu} + z_{1-\frac{\alpha}{2}} \cdot s]$

Vorgehensweise: $1 - \frac{\alpha}{2}$ berechnen und innerhalb der Tabelle zur Normalverteilung diesen Wert suchen. Der gesuchte z -Wert ergibt sich dann aus den Zeilen- und Spaltenberschriften.

4.4 Tests

- **Nullhypothese (H_0):** Annahme, die geprüft werden soll
- **Gegenhypothese (H_1):** Gegenteil der Nullhypothese
- **Signifikanzniveau (α):** Wahrscheinlichkeit, mit der die Nullhypothese fälschlicherweise abgelehnt wird
- **Teststatistik:** Funktion der Stichprobenwerte, die zur Entscheidung über die Annahme oder Ablehnung der Nullhypothese herangezogen wird
- **Ablehnungsbereich:** Bereich der Teststatistik, in dem die Nullhypothese abgelehnt wird
- **p-Wert:** Wahrscheinlichkeit, mit der die Nullhypothese verworfen werden kann
- **Fehler 1. Art:** H_0 wird fälschlicherweise abgelehnt
- **Fehler 2. Art:** H_0 wird fälschlicherweise nicht abgelehnt

4.4.1 χ^2 -Anpassungstest

Vergleicht die beobachtete Verteilung einer Stichprobe mit einer theoretischen (erwarteten) Verteilung. Es wird geprüft, ob die beobachtete Häufigkeitsverteilung von Kategorien mit der erwarteten Häufigkeitsverteilung übereinstimmt.

1. Voraussetzung: Zufallsvariable X (z.B. Ergebnis eines Würfelwurfs) mit s Ausprägungen (Kategorien; z.B. 1-6 Würfelaugen) und N Beobachtungen (Stichprobenumfang).

2. Nullhypothese (H_0): Die tatsächliche Verteilung entspricht der erwarteten Verteilung ($P(X \in A_i) = \rho_i = \frac{1}{6}$)

3. Gegenhypothese (H_1): Nicht H_0

4. χ^2 -Teststatistik:

$$D_\rho = \sum_{i=1}^s \frac{(h(i) - N\rho(i))^2}{N\rho(i)} = \left(\sum_{i=1}^s \frac{h(i)^2}{N\rho(i)} \right) - N = N \left(\sum_{i=1}^s \frac{L(i)^2}{\rho(i)} \right) - N$$

wobei

- N = Stichprobengröße (z.B. 50 Würfe mit Würfel)
 - s = Anzahl der Kategorien (z.B. 6 Würfelaugen)
 - A_i = Kategorie i (z.B. Würfelaugen 1-6)
 - $h(i)$ = Anzahl der Beobacht. in Kategorie A_i (z.B. 8 Würfe 1er Würfe)
 - $L(i) = \frac{h(i)}{N}$ = relative Häufigkeit der Beobachtungen in Kategorie A_i (z.B. 8/50 Würfe mit Würfelaugen 1 usw.)
 - $\rho(i)$ = Wahrscheinlichkeit der Kategorie A_i (z.B. $\frac{1}{6}$ für Würfelauge 1 usw.)
 - α = Irrtumswahrscheinlichkeit/Signifikanzniveau (z.B. 5%)
- 5. Ablehnungsbereich:** H_0 ablehnen, wenn: $D_\rho > \chi_{s-1; 1-\alpha}^2$

4.4.2 χ^2 -Unabhängigkeitstest

Prüft die Unabhängigkeit zweier Merkmale, d.h. ob das Vorkommen einer Variable von der anderen abhängt oder nicht.

1. Voraussetzung: Zufallsvariable X und Y nehmen jeweils zwei Werte an (z.B. X =männlich/weiblich, Y =raucht/nicht raucht).

2. Nullhypothese (H_0): X und Y sind stochastisch unabhängig

3. Gegenhypothese (H_1): Nicht H_0

4. χ^2 -Teststatistik:

Aufstellen einer Vierfeldertafel:

	Y	\bar{Y}	Σ
X	N_{11}	N_{12}	$N_{1\cdot}$
\bar{X}	N_{21}	N_{22}	$N_{2\cdot}$
Σ	$N_{\cdot 1}$	$N_{\cdot 2}$	n

	Nichtraucher	Raucher	Σ
männlich	170	30	200
weiblich	250	150	400
Σ	420	180	600

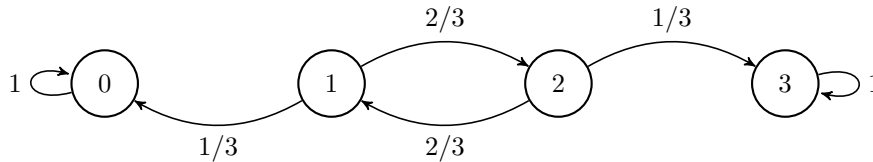
Daraus Berechnung der Teststatistik:

$$D_p = n \frac{(N_{11}N_{22} - N_{12}N_{21})^2}{N_{1\cdot}N_{2\cdot}N_{\cdot 1}N_{\cdot 2}}$$

Gedankenstütze: Determinante hoch 2 geteilt durch Produkt aus allen Spalten- und Zeilensummen

5. Ablehnungsbereich: H_0 ablehnen, wenn: $T > \chi^2_{1;1-\alpha}$

5 Markov-Ketten



Eine homogene, **irreduzible**, **aperiodische** Markov-Kette mit endlichem Zustandsraum ist immer **stationär**, d.h. sie konvergiert gegen ihr statistisches Gleichgewicht.

5.1 Übergangsmatrix

5.2 Stationäre Verteilung

Ein Zustandsvektor π heißt *stationäre Verteilung* einer Markov-Kette, wenn gilt:

$$\pi \cdot P = \pi$$

Gleichungen lösen, ggf. mit Hilfe von Parametern t , wenn es keine eindeutige Lösung gibt. Wert für t bestimmen, indem die Summe der Komponenten des Vektors π gleich 1 gesetzt wird.

Beispiel:

$$P = \begin{pmatrix} 0.5 & 0.5 \\ 0.3 & 0.7 \end{pmatrix}$$

$$(\pi_1 \quad \pi_2) \cdot \begin{pmatrix} 0.5 & 0.5 \\ 0.3 & 0.7 \end{pmatrix} = (\pi_1 \quad \pi_2)$$

Da $\sum \pi_i = 1$ gilt

$$\pi_1 + \pi_2 = 1 \Leftrightarrow \pi_1 = 1 - \pi_2 \Leftrightarrow \pi_2 = 1 - \pi_1$$

π_1 und π_2 in die folgenden Gleichungen einsetzen:

$$\pi_1 \cdot 0.5 + \pi_2 \cdot 0.3 = \pi_1 \implies \pi_2 = \frac{5}{8}$$

$$\pi_1 \cdot 0.5 + \pi_2 \cdot 0.7 = \pi_2 \implies \pi_1 = \frac{3}{8}$$

Daraus folgt der stationäre Zustandsvektor: $\pi = \left(\frac{3}{8} \quad \frac{5}{8}\right)$

5.3 Irreduzibilität

Es ist von jedem Zustand aus möglich, jeden anderen Zustand zu erreichen. Die Prüfung kann manuell erfolgen.

Beispiel: $P = \begin{pmatrix} 1-p & p/2 & p/2 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$ ist irreduzibel für $p \in (0, 1]$, da für $p = 0$ Zustände 2 und 3 nicht mehr erreichbar sind.

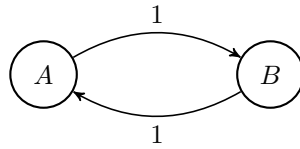
5.4 Aperiodizität

Die Periode eines Zustands ist die größte gemeinsame Teiler aller Pfade, die zu diesem Zustand zurück führen.

Starte in einem Zustand i und gehe in einen Zustand j . Die Periode von i ist die größte gemeinsame Teiler aller Pfade, die von j nach i führen. Wenn die Periode von jedem Zustand i gleich 1 ist, ist die Markov-Kette aperiodisch. Das heißt:

$$d(z) = \text{ggT}\{n \in \mathbb{N} | P_{ii}^{(n)} > 0\} = 1$$

Beispiel: $P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ ist *nicht* aperiodisch, da $d(z) = 2$; man springt immer zwischen den beiden Zuständen mit einer geraden Anzahl hin und her.



Beispiel: $P = \begin{pmatrix} 1-p & p/2 & p/2 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$ ist aperiodisch für $p \in [0, 1)$, da bei $p = 1$ immer von Zustand 1 in 2 oder 3 gesprungen wird und wieder zurück. Für alle anderen Werte wird in zwei Fällen auch hin- und zurückgesprungen. Allerdings gibt es auch die Möglichkeit, dass vom Zustand A nicht gewechselt wird und man dort bleibt.

