Errata

(Mathematische Einführung in Data Science von Sven-Ake Wegner)

15. Oktober 2024

- Seite 10, Zeile -1:
$$\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y}) = \sum_{i=1}^{n} x_iy_i - \overline{y}\sum_{i=1}^{n} x_i - \overline{x}\sum_{i=1}^{n} y_i + n\overline{x}\,\overline{y} = \Big(\sum_{i=1}^{n} x_iy_i\Big) - n\overline{x}\,\overline{y}.$$
- Seite 11, Zeile -1:
$$0 = \sum_{i=1}^{n}(ax_i+b-y_i) = a\sum_{i=1}^{n} x_i + nb - \sum_{i=1}^{n} y_i = an\overline{x} + nb - n\overline{y},$$
- Seite 13, Zeile 22: Seitenlänge
- Seite 15, Zeile -10: …konstante Zufallsvariable av.
- Seite 15, Zeile -3 und Seite 16, Zeile 2: $\mathrm{V}\cancel{\mathrm{ar}}\overline{(x\mathcal{E})}^{(n)}$
- Seite 15, Zeile -2: $\cdots + \frac{2}{n^2}\sum_{i<j} x_ix_j\,\mathrm{E}(\mathcal{E}_i)\,\mathrm{E}\,\mathcal{E}_j)$
- Seite 16, Zeile 3: da $\overline{x^{(n)2}} = \operatorname{var}(x^{(n)}) + \overline{x^{(n)}}^2$ als Summe …
- Seite 17, Zeile 8: Ist Letzteres der Fall, so ist $\operatorname{sign}(r_{xy}) = \operatorname{sign}(\langle u, v\rangle) = \cdots$
- Seite 17, Zeile 11: r_{xy}
- Seite 24, Zeile -11: $f = \langle (a_1,\dots,a_d),\,\cdot\,\rangle + a_0$
- Seite 26, Zeile 18: $f\cancel{(z)} = \operatorname{sig}(\langle w,\,\cdot\,\rangle)$
- Seite 31, Zeile 12: …und $\langle w, \widehat{w}_k\rangle < 0$ gelten.
- Seite 38, Zeile 13: …und $x_1,\dots,x_k \in D_1 := \{x \mid (x,y)\in D\}$ gelten.
- Seite 38, Zeile 14: $x_1 \in \operatorname*{argmin}_{z\in D_1} \rho(x,z)$ sowie $x_j \in \operatorname*{argmin}_{z\in D_1\setminus\{x_1,\dots,x_{j-1}\}} \rho(x,z)$ für $j \geqslant 2$
- Seite 38, Zeile 20: $f(x) \in \operatorname*{argmax}_{y\in Y} N(y)$
- Seite 39, Zeile 4: $z^\star \leftarrow \operatorname{argmin}_{z\in D_1'} \rho(x,z)$
- Seite 39, Zeile 5: $D_1' \leftarrow D_1'\setminus\{(z^\star, y^\star)\}$
- Seite 39, Zeile 10–11: Hierbei bezeichnet $\pi_2(x,y) = x$ die Projektion auf den zweiten Eintrag von $(x,y)\in D$ und $y^\star$ das Label von $z^\star$.
- Seite 39, Zeile -2: …vorgenannten grauen Punkt x mit …
- Seite 41, Zeile 6–7: … *Trainingsdaten* D_{Train} und *Testdaten* D_{Test}. Dann bestimmt man einen Klassifizierer $f\colon X\to Y$ anhand von D_{Train}, und stellt fest, welcher Anteil der Punkte aus D_{Test} durch f korrekt klassifiziert wird.
- Seite 41, Zeile 17: *Die Berechnung der k-nächsten Nachbarn von x kann so implementiert werden, dass dabei höchstens $(n\cdot d\cdot k)$-viele Mulitplikationen anfallen.*
- Seite 41, Zeile 21: In der euklidischen Metrik müssen für eine solche Distanz $(d-1)$-viele Multiplikationen ausgeführt werden, wenn wir für die Berechnung des Argmin die Wurzel weglassen. Dies führt auf
$$(d-1)\cdot(n+(n-1)+\cdots+(n-k+1)) \leqslant \cancel{C}\,d\cdot k\cdot n$$
Multiplikationen ~~mit einem geeigneten $C\in\mathbb{N}$~~.
- Seite 42, Zeile 7: …, wählen wir k-nächste Nachbarn $x_1,\dots,x_k$ von x und bezeichnen mit $y_1,\dots,y_k$ deren Label.
- Seite 42, Zeile 13: $f\colon X\to Y,\ f(x) = \dfrac{\sum_{i=1}^{k} w(x_i,x)\cdot y_i}{\sum_{i=1}^{k} w(x_i,x)}$
- Seite 43, Zeile 11: $\tilde{x}^{(i)} = \Big(a + \dfrac{(x_1^{(i)} - \min_{j=1,\dots,n} x_1^{(j)})(b-a)}{\max_{j=1,\dots,n} x_1^{(j)} - \min_{j=1,\dots,n} x_1^{(j)}}, \dots\Big)$

- Seite 43, Zeile 13: $\tilde{x}^{(i)} = \left(\frac{x_1^{(i)} - \overline{x_1^{(\cdot)}}}{\sigma_1}, \dots\right)$
- Seite 44, Zeile −9: $\rho(x^{(1)}, x^{(4)}) = 3.681$
- Seite 48, Zeile 5–6: …, dass sich die Texte 1 und 2 deutlich kosinusähnlicher sind als 1 und 3 bzw. 2 und 3.
- Seite 48, Zeile 16: Einige solcher Methoden diskutieren wir in Aufgabe 3.10.
- Seite 48, Zeile 26: ~~Der Kosinusabstand wird hingegen immer kleiner, je öfter das Wort im zweiten Text auftritt.~~ Der Kosinusabstand verhält sich hier natürlicher, weil sich der Skalarprodukt vergrößert, wenn sich die Häufigkeit des fixierte Wortes im zweiten Text erhöht.
- Seite 52, Zeile −14: D
- Seite 54, Zeile 12: Für endliche Teilmengen $A, B \subseteq X$ definieren wir …
- Seite 55, Zeile 21–23: … bevor der minimale Abstand zwischen den Clustern in der nächsten Runde erstmalig über einen einzugebenden Wert $\delta > 0$ wachsen würde.
- Seite 55, Zeile 24:

```
function VERKNÜPFUNGSBASIERTES CLUSTERING (X, ρ, D, δ)
    k ← #D
    for i ← 1 to k do
        C_i ← {x_i}
    while min_{i≠j} ρ(C_i, C_j) ⩽ δ and k ⩾ 2 do
        m ← 0
        (i*, j*) ← argmin_{i≠j} ρ(C_i, C_j)
        for ℓ ← 1 to k − 1 do
            if ℓ = min(i*, j*) then
                C_ℓ ← C_{i*} ∪ C_{j*}
            if ℓ = max(i*, j*) then
                m ← 1
                C_ℓ ← C_{ℓ+m}
            else
                C_ℓ ← C_{ℓ+m}
        k ← k − 1
    return C_1, …, C_k
```

- Seite 56, Zeile 4: …, wie in Definition 4.2, …
- Seite 56, Zeile 8: Aufgabe 4.2
- Seite 56, Zeile −2: $K\colon \mathcal{C}_k \to \mathbb{R}$
- Seite 58, Zeile −16: *Der folgende Pseudocode approximiert einen Minimierer der k-means-Kostenfunktion.*
- Seite 58, Pseudocode:

```
function K-MEANS (D, k, X, ρ)
    μ_1, …, μ_k ← pairwise different points from X
    for i ← 1 to k do
        C_i ← {x ∈ D | i ∈ argmin_{j=1,…,k} ρ(x, μ_j)}
    U ← True
    while U = True do
        U ← False
        for i ← 1 to k do
            μ_i ← μ(C_i)
        for i ← 1 to k do
            C'_i ← {x ∈ D | i ∈ argmin_{j=1,…,k} ρ(x, μ_j)}
            if C'_i ≠ C_i
                C_i ← C'_i
                U ← True
    return C_1, …, C_k
```

Hierbei ist in den Zeilen 4 und 9 darauf zu achten, dass man jeweils ein i im Argmin auswählt, falls dieses nicht eindeutig ist.

- Seite 59, Zeile 8:

$$\mu(A) \in \underset{\mu\in A}{\operatorname{argmin}} \sum_{x\in A} \rho(x,\mu)^2 \quad \text{bzw.} \quad \mu(A) \in \underset{\mu\in X}{\operatorname{argmin}} \sum_{x\in A} \rho(x,\mu)$$

- Seite 59, Zeile −5: *Beweis.* Für $j \geqslant 1$ bezeichne $(C_1^{(j)}, \dots, C_k^{(j)})$ das Clustering, welches in der j-ten Runde des Algorithmus definiert wird.
- Seite 59, Zeile −2: $K(C_1^{(j)}, \dots, C_k^{(j)}) \;=\; \min_{\mu_1,\dots,\mu_k\in X} \sum_{i=1}^{k} \sum_{x\in C_i^{(j)}} \rho(x,\mu_i)^2$
- Seite 60, Zeile 5: … Zeile 11 von Algorithmus …
- Dort entspricht Bild 1b der vorletzten Zeile in der Abschätzung und Bild 2a der darüber.
- Seite 60, Zeile 13: weil wir ja gerade wegen $\rho(x_3,\mu_1) < \rho(x_3,\mu_2)$ den Punkt x_3 vom Cluster C_2 in Bild 1b zu Cluster C_1 in Bild 2a verschoben haben.
- Seite 62, Zeile 4: ~~Mittelpunkten~~ Startwerten
- Seite 62, Zeile 5: $\operatorname{argmin}_{i=1,2} \|x-\mu_i\|$
- Seite 64, Zeile 7: … sondern ~~nur~~ z.B. nur durch eine Liste …
- Seite 66, Zeile 2: … In Beispiel 5.7 ist $\lambda_2 \neq 0$ und es gibt keine Cluster (oder anders ausgedrückt nur ein einziges), in Beispiel 5.8 …
- Seite 69, Zeile 15: Für die andere Richtung sei $\{v_1,\dots,v_n\}$ eine Basis aus Eigenvektoren zu den λ_i und $U \subseteq \mathbb{R}^n$ ein Unterraum mit $\dim U = n-k+1$. Dann ist aus Dimensionsgründen $U \cap \operatorname{span}\{v_1,\dots,v_k\} \neq \{0\}$ und wir können $0 \neq x = \alpha_1 v_1 + \dots + \alpha_k v_k \in U$ wählen. Dann folgt

$$\frac{\langle x, Mx\rangle}{\langle x,x\rangle} = \frac{\sum_{i=1}^{k} \lambda_i \alpha_i^2}{\sum_{i=1}^{k} \alpha_i^2} \leqslant \frac{\lambda_k \sum_{i=1}^{k} \alpha_i^2}{\sum_{i=1}^{k} \alpha_i^2} = \lambda_k,$$

weil die λ_i wachsend sind. Damit erhalten wir $\min_{0\neq x\in U} \frac{\langle x,Mx\rangle}{\langle x,x\rangle} \leqslant \lambda_k$ und es folgt

$$\max_{\substack{U\subseteq\mathbb{R}^n \\ \dim U = n-k+1}} \; \min_{\substack{x\in U\\ x\neq 0}} \frac{\langle x, Mx\rangle}{\langle x,x\rangle} \leqslant \lambda_k.$$

- Seite 70, Zeile 3–4: Wir wählen $v_1 = \mathbb{1}$ als Eigenvektor zum Eigenwert λ_1. Dann gelten für Eigenwert und Eigenvektor λ_2 bzw. v_2 …
- Seite 70, Zeile 7: $v_2 \in \cdots$
- Seite 70, Zeile −3: Sei $G = (V,E)$ ein Graph mit $\deg(v) > 0$ für alle $v \in V$.
- Seite 71, Zeile 5: $\emptyset \neq S \subset V$
- Seite 72, Zeilen 6 und 8: $\emptyset \neq S \subset V$
- Seite 72, Zeile −4:

$$\lambda_2(\mathcal{L}) = \min_{\substack{x\neq 0\\ \langle Dx,\mathbb{1}\rangle = 0}} \frac{\sum_{\{i,j\}\in E} (x_i - x_j)^2}{\sum_{i=1}^{n} x_i^2 d_i}.$$

- Seite 73, Zeile 2:

$$\mathcal{L} D^{1/2}\mathbb{1} = D^{-1/2} L D^{-1/2} D^{1/2}\mathbb{1} = D^{-1/2} L \mathbb{1} \underset{\substack{\uparrow\\ \text{Proposition}\\ 5.6}}{=} D^{-1/2} 0 \mathbb{1} = 0 D^{1/2}\mathbb{1}.$$

- Seite 73, Zeile 4:

$$\lambda_2(\mathcal{L}) \underset{\substack{\uparrow \\ \text{Satz} \\ 5.13}}{=} \min_{\substack{x\neq 0 \\ \langle x, D^{1/2}\mathbb{1}\rangle = 0}} \frac{\langle x, \mathcal{L}x\rangle}{\langle x, x\rangle} \overset{(*)}{=} \min_{\substack{y\neq 0 \\ \langle D^{1/2}y, D^{1/2}\mathbb{1}\rangle = 0}} \frac{\langle D^{1/2}y, \mathcal{L}D^{1/2}y\rangle}{\langle D^{1/2}y, D^{1/2}y\rangle}$$

$$= \min_{\substack{y\neq 0 \\ \langle y, D\mathbb{1}\rangle = 0}} \frac{\langle y, Ly\rangle}{\langle y, Dy\rangle} = \min_{\substack{y\neq 0 \\ \langle Dy, \mathbb{1}\rangle = 0}} \frac{\sum_{\{i,j\}\in E}(y_i - y_j)^2}{\sum_{i=1}^n y_i^2 d_i}$$

- Seite 73, Zeile 10: Ist G d-regulär (d.h., es gilt $\deg(v) = d$ für alle Vertices $v \in V$ mit einem $d \in \mathbb{N}$), so gilt ...
- Seite 73, Zeile −9: *(ii) Für* $r \leqslant k \leqslant n-1$ *gelten* ...
- Seite 73, Zeile −2:

$$\operatorname{vol} S_k^{\mathrm{c}} - \operatorname{vol} S_{k+1}^{\mathrm{c}} = \sum_{i=k+1}^{n} d_i - \sum_{i=k+2}^{n} d_i = d_{k+1}$$

- Seite 74, Zeile 15: $\emptyset \neq S \subset V$
- Seite 74, Zeile 16: $\operatorname{Cheeg}(G) = \frac{\#\partial S}{\operatorname{vol} S}$
- Seite 74, Zeile 20:

$$\langle Dx, \mathbb{1}\rangle = \sum_{i=1}^{n} d_i x_i = \dots$$

- Seite 75, Zeile 7–8:

$$\begin{aligned} \dots &= \operatorname{vol} S - \operatorname{vol} S \cdot \frac{\operatorname{vol} S}{\operatorname{vol} S + \operatorname{vol} S^{\mathrm{c}}} \\ &\geqslant \operatorname{vol} S - \operatorname{vol} S \cdot \frac{\operatorname{vol} S}{2 \operatorname{vol} S}, \end{aligned}$$

- Seite 75, Zeile −6 (Gleichung (5.2)):

$$\cdots \text{ und } \langle Dx, \mathbb{1}\rangle = \sum_{i=1}^{n} d_i x_i = 0$$

- Seite 75, Zeile −5: ..., zu zeigen, dass $\lambda_2 \geqslant \alpha^2/2$ ist, ...
- Seite 76, Zeile 4:

$$\begin{bmatrix} x_1 - x_r \\ \vdots \\ x_{r-1} - x_r \\ 0 \\ x_{r+1} - x_r \\ \vdots \\ x_n - x_r \end{bmatrix} = \begin{bmatrix} x_1 - x_r \\ \vdots \\ x_{r-1} - x_r \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 0 \\ x_r - x_{r+1} \\ \vdots \\ x_r - x_n \end{bmatrix} =: p - n.$$

- Seite 77, Zeile 4:

$$\begin{aligned} \lambda_2 &= \frac{\sum_{\{i,j\}\in E}(x_i - x_j)^2}{\sum_{i=1}^n x_i^2 d_i} \\ &\underset{\substack{\uparrow \\ (5.3) \\ (5.4)}}{\geqslant} \frac{\sum_{\{i,j\}\in E}\big((p_i - p_j)^2 + (n_i - n_j)^2\big)}{\sum_{i=1}^n (p_i^2 + n_i^2) d_i} \\ &= \frac{\sum_{\{i,j\}\in E}(p_i - p_j)^2 + \sum_{\{i,j\}\in E}(n_i - n_j)^2}{\sum_{i=1}^n p_i^2 d_i + \sum_{i=1}^n n_i^2 d_i} \\ &\underset{\substack{\uparrow \\ (5.5)}}{\geqslant} \min\left(\frac{\sum_{\{i,j\}\in E}(p_i - p_j)^2}{\sum_{i=1}^n p_i^2 d_i}, \frac{\sum_{\{i,j\}\in E}(n_i - n_j)^2}{\sum_{i=1}^n n_i^2 d_i}\right) \\ &= \min\left(\frac{\sum_{\{i,j\}\in E}(p_i - p_j)^2}{\sum_{i=1}^n p_i^2 d_i} \cdot \frac{\sum_{\{i,j\}\in E}(p_i + p_j)^2}{\sum_{\{i,j\}\in E}(p_i + p_j)^2}, \dots\right) \\ &=: \min\left(\frac{Z}{N}, \dots\right). \end{aligned}$$

- Seite 77, Zeile 14:

$$N = \sum_{i=1}^{n} p_i^2 d_i \cdot \sum_{\{i,j\}\in E} (p_i + p_j)^2 \underset{\substack{\uparrow \\ (5.5)}}{\geqslant} \sum_{i=1}^{n} p_i^{\color{red}2} d_i \cdot \sum_{\{i,j\}\in E} 2(p_i^2 + p_j^2)$$

- Seite 78, Zeile 10:

$$\cdots \underset{\substack{\uparrow \\ \text{Teleskop-} \\ \text{summe}}}{=} \Big(\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathbb{1}_E(i,j) \sum_{k=i}^{j-1} {\color{red}(} p_k^2 - p_{k+1}^2 {\color{red})} \Big)^2$$

- Seite 79, Zeile 13: Als Letztes wollen wir noch notieren, dass Satz 5.21 zusammen mit Bemerkung 5.17(i) eine obere Schranke für den Eigenwert $\lambda_2(\mathcal{L})$ liefert.

  Korollar 5.22. Sei $G = (V, E)$ ein Graph mit $\deg(i) > 0$ für alle $i \in V$. Dann gilt für den zweitkleinsten Eigenwert λ_2 der normalisierten Laplacematrix von G die Abschätzung $\lambda_2 \leqslant {\color{red}2}$. □

- Seite 87, Zeilen 11 und 14: $\cancel{A}X$
- Seite 91, Zeile 12: …Singulärwert von ${\color{red}A}$, …
- Seite 95, Zeile 6: Ist $1 \leqslant k {\color{red}<} p$ und haben wir $\sigma_{\color{red}k+1} = \cdots = \sigma_p = 0$, …
- Seite 101, Zeile -8: ~~Multiplikation mit V^{T} von links in~~ Proposition 7.14 liefert also ~~genau~~ eine …
- Seite 102, Zeile 17: $r = {\color{red}\mathrm{rk}}(A)$
- Seite 106, Zeile 3: $\forall\, i, j \in \{1, \dots, n\}$: ${\color{red}\big|}\|T_{V_k} a_i - T_{V_k} a_j\| {\color{red}- \|a_i - a_j\|\big|} \leqslant 2\Big(\sum_{\ell=k+1}^{r} \sigma_\ell \Big)^{1/2}$
- Seite 107, Zeile -1 bis Seite 108, Zeile 1:

$$A = \underbrace{\begin{bmatrix} 0.07 & 0.29 & 0.32 & 0.51 & 0.66 & 0.18 & -0.23 \\ 0.13 & -0.02 & -0.01 & -0.79 & 0.59 & -0.02 & -0.06 \\ 0.68 & -0.11 & -0.05 & -0.05 & -0.24 & 0.56 & -0.35 \\ 0.15 & 0.59 & 0.65 & -0.25 & -0.33 & -0.09 & 0.11 \\ 0.41 & -0.07 & -0.03 & 0.10 & -0.02 & -0.78 & -0.43 \\ 0.07 & 0.73 & -0.67 & 0.00 & -0.00 & 0.00 & 0.00 \\ 0.55 & {\color{red}-0.09} & -0.04 & 0.17 & 0.17 & -0.11 & 0.78 \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} 12.4 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 9.5 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 1.3 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} 0.56 & 0.09 & 0.56 & 0.09 & 0.59 \\ -0.12 & 0.69 & -0.12 & 0.69 & 0.02 \\ -0.40 & -0.09 & -0.40 & -0.09 & 0.80 \\ \cancel{0.41} & \cancel{0.09} & \cancel{0.40} & \cancel{0.09} & \cancel{-0.80} \\ 0.51 & 0.48 & -0.51 & -0.48 & -0.00 \\ 0.48 & -0.51 & -0.48 & 0.51 & -0.00 \end{bmatrix}}_{V^{\mathrm{T}}}$$

$$= \begin{bmatrix} 0.07 & 0.29 & 0.32 \\ 0.13 & -0.02 & -0.01 \\ 0.68 & -0.11 & -0.05 \\ 0.15 & 0.59 & 0.65 \\ 0.41 & -0.07 & -0.03 \\ 0.07 & 0.73 & -0.67 \\ 0.55 & {\color{red}-0.09} & -0.04 \end{bmatrix} \begin{bmatrix} 12.4 & & \\ & 9.5 & \\ & & 1.3 \end{bmatrix} \begin{bmatrix} 0.56 & 0.09 & 0.56 & 0.09 & 0.59 \\ -0.12 & 0.69 & -0.12 & 0.69 & 0.02 \\ -0.40 & -0.09 & -0.40 & -0.09 & 0.80 \end{bmatrix}.$$

- Seite 108, Zeile 5:

$$\breve{A} = \begin{bmatrix} 0.15 & 1.97 & 0.15 & 1.97 & 0.56 \\ 0.92 & 0.01 & 0.92 & 0.01 & {\color{red}0.94} \\ 4.84 & 0.03 & 4.84 & 0.03 & 4.95 \\ 0.36 & 4.03 & 0.36 & 4.03 & 1.20 \\ 2.92 & {\color{red}-0.00} & 2.92 & {\color{red}-0.00} & 2.98 \\ -0.34 & 4.86 & -0.34 & 4.86 & 0.65 \\ {\color{red}3.92} & {\color{red}0.02} & {\color{red}3.92} & {\color{red}0.02} & {\color{red}4.00} \end{bmatrix} = \begin{bmatrix} 0.07 & 0.29 \\ 0.13 & -0.02 \\ 0.68 & -0.11 \\ 0.15 & 0.59 \\ 0.41 & -0.07 \\ 0.07 & 0.73 \\ 0.55 & {\color{red}-0.09} \end{bmatrix} \begin{bmatrix} 12.4 & \\ & 9.5 \end{bmatrix} \begin{bmatrix} 0.56 & 0.09 & 0.56 & 0.09 & 0.59 \\ -0.12 & 0.69 & -0.12 & 0.69 & 0.02 \end{bmatrix}$$

- Seite 109, Zeile 16:

$$= \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} 0.07 & 0.29 & \cdots & -0.23 \\ 0.13 & -0.02 & & -0.06 \\ 0.68 & -0.11 & & -0.35 \\ 0.15 & 0.59 & & 0.11 \\ 0.41 & -0.07 & & -0.43 \\ 0.07 & 0.73 & & 0.00 \\ 0.55 & {\color{red}-0.09} & \cdots & 0.78 \end{bmatrix} \begin{bmatrix} 12.4 & & & \\ & 9.5 & & \\ & & 1.3 & \\ & & & \ddots \end{bmatrix} \begin{bmatrix} 0.56 & 0.09 & 0.56 & 0.09 & 0.59 \\ -0.12 & 0.69 & -0.12 & 0.69 & 0.02 \\ \vdots & & & & \vdots \\ 0.48 & -0.51 & -0.48 & 0.51 & 0.00 \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

- Seite 109, Zeile -2:

$$u_2 = \quad 0.29 \cdot \text{Antje} - 0.02 \cdot \text{Birgit} + \cdots {\color{red}-0.09} \cdot \text{Gül},$$

- Seite 111, Zeile 12:

$$\breve{A} = \begin{array}{r} \text{Antje} \rightarrow \\ \text{Birgit} \rightarrow \\ \text{Constanze} \rightarrow \\ \text{Dorothee} \rightarrow \\ \\ \text{Eleonore} \rightarrow \\ \text{Fatema} \rightarrow \\ \text{Gül} \rightarrow \end{array} \begin{bmatrix} \boxed{0.07 \quad 0.29} \\ \boxed{0.13 \quad -0.02} \\ \boxed{0.68 \quad -0.11} \\ \boxed{0.15 \quad 0.59} \\ \boxed{0.41 \quad -0.07} \\ \boxed{0.07 \quad 0.73} \\ \boxed{0.55 \quad {\color{red}-0.09}} \end{bmatrix} \begin{bmatrix} 12.4 & \\ & 9.5 \end{bmatrix} \begin{array}{ccccc} \text{Alien} & \text{Casablanca} & \text{Star Wars} & \text{Titanic} & \text{Matrix} \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \end{array} \begin{bmatrix} \boxed{\begin{smallmatrix} 0.56 \\ -0.12 \end{smallmatrix}} & \boxed{\begin{smallmatrix} 0.09 \\ 0.69 \end{smallmatrix}} & \boxed{\begin{smallmatrix} 0.56 \\ -0.12 \end{smallmatrix}} & \boxed{\begin{smallmatrix} 0.09 \\ 0.69 \end{smallmatrix}} & \boxed{\begin{smallmatrix} 0.59 \\ 0.02 \end{smallmatrix}} \end{bmatrix}$$

- Seite 110, Zeile 7: $\mathbb{R}^5_{\mathcal{F}}$
- Seite 112, Zeile 5: $(\text{Film})_{\mathcal{F}} \quad \mathbb{R}_{\mathcal{F}} \xrightarrow{A} \mathbb{R}_{\mathcal{B}} \quad (\text{Bewerterin})_{\mathcal{B}}$
- Seite 111, Zeile 4: R (statt R)
- Seite 112, Zeilen 14 und 25: R (statt R)
- Seite 112, Zeile −3: ... die Daten ~~erst~~ einem geeigneten Pre- oder Postprocessing zu unterziehen ...
- Seite 125, Zeile −3: $\mathrm{H}_d\cancel{(1)}$
- Seite 129, Zeile −1: $0 < \varepsilon < 1$
- Seite 131, Zeile 9: $H_{\delta,x_1} = \big\{(x_2,\dots,x_d \in \mathbb{R}^{d-1} \,\big|\, (x_1,x_2,\dots,x_d) \in H_{\delta}\big\}$
- Seite 134, Zeile 9: Für solche Zufallsvektoren erhält man ...
- Seite 139, Zeile −1: $\|X^{(i)}\| \approx 1$
- Seite 146, Zeile 5: $\mathrm{P}\big[\,\big|\|X\| - \sqrt{d}\big| \geqslant \varepsilon\big]$
- Seite 164, Zeile −17: ... der Abstand der Mittelpunkte, ...
- Seite 167, Zeile 5–7: ~~Im Fall der Varianz sind diese eher technisch, und wir formulieren daher im Satz für die Varianz nur die sich ergebende qualitative asymptotische Aussage.~~
- Seite 183, Zeile −1: Wenn D linear trennbar ist, ...
- Seite 186, Zeile 8: ␣ **return** $w^{(j)}$
- Seite 189, Zeile −5: $(x,y) \in D$
- Seite 190, Zeile 14–15: ... die Worte „Bonus“, „Vertrag“, „das“ und „Mensa“ ...
- Seite 193, Zeile 6: $\cdots = \big\langle w', x_0 - (\langle w', x_0\rangle + b')w'\big\rangle + b'$
- Seite 193, Zeile 12: $\geqslant \|x_0 - x_1\|^2 + 2\big\langle(\langle w', x_0\rangle + b'\big)w', x_1 - x\big\rangle$
- Seite 193, Zeile 13: $= \|x_0 - x_1\|^2 + 2(\langle w', x_0\rangle + b')\langle w', x_1 - x\rangle$
- Seite 194, Zeilen −2 und −1: $\mathcal{R}(D)$, $\mathcal{K}(D)$
- Seite 194, Zeile −1: $h\colon \mathbb{R}^d \to \mathbb{R}^{\cancel{d}}$
- Seite 195, Zeilen 2, 4, 5, 9, 10, 12, 15 und 20: $\mathcal{R}(D)$, $\mathcal{K}(D)$
- Seite 196, Zeilen 6–8, 12 und 13: $\mathcal{R}(D)$, $\mathcal{K}(D)$
- Seite 198, Zeile 6: $(w^*, b^*) \in M$
- Seite 198, Zeile 11–16: Es folgt also nach Proposition 17.14, dass $w_1^* = w_2^* =: w^*$ ist. Gelte ohne Einschränkung $b_1^* < b_2^*$. Dann wählen wir für ~~(w^*, b_1^*) und~~ (w^*, b_2^*) ~~jeweils i_1 und~~ einen Index i wie in Teil 2 und erhalten

$$y_i(\langle w^*, x_i\rangle + b_1^*) = \langle w^*, x_i\rangle + b_1^* < \langle w^*, x_i\rangle + b_2^* = 1$$

  im Widerspruch dazu, dass (w^*, b_1^*) in M liegt. ~~Ist $b_1^* > b_2^*$, so vertauschen wir die Rollen von i_1 und i_2.~~
- Seite 198, Zeile −11, −8 und −4: $\mathcal{R}(D)$, $\mathcal{K}(D)$
- Seite 200, Zeile 7: $L(x,\theta,\mu) := f(x) - \sum_{i=1}^{q} \theta_i g_i(x) + \cancel{\mu} \sum_{j=r}^{q} \mu_i h_j(x)$
- Seite 200, Zeile 18: ..., ist x^* ein Minimierer des oben angegebenen Optimierungsproblems ...
- Seite 203, Zeile 17: ... ist es möglich, i_0 mit $\lambda_{i_0} \neq 0$ ...
- Seite 206, Zeile 7: $D_1 = \{(x_i, y_i) \mid i = 1, 6, 7\}$ und $D_2 = \{(x_i, y_i) \mid i = 1, 6, 11\}$
- Seite 206, Zeile 8: Durch Streichen von Nullen in Beispiel 14.15(i) und Lösung des Optimierungsproblems für D_2 erhalten wir ...
- Seite 206, Zeile 12: ~~0.349~~ 0.439
- Seite 207, Zeile 4 (in den Bildern): $D_1 \cdots D_2$
- Seite 207, Zeilen 12 und 13: $\lambda_i^{\star}$
- Seite 208, Zeile 13: ... und Satz 14.12(ii) ist [GK02, Satz 2.46].
- Seite 209, Zeile −6: ..., wenn man einen der Punkte x_6^*, x_7^*, oder x_{11}^* weglässt?

- Seite 216, Zeile −14: . . . , sodass $y_i(\langle w, x_i\rangle + b) \geqslant 1$ für . . . .
- Seite 220, Zeile 16: . . . mit $\lambda^*_{i_0} \neq 0$ ist . . .
- Seite 220, Zeile 18: . . . Wahl von λ^* und . . .
- Seite 224, Zeile 6: Haben wir also $\langle f, f\rangle = 0$ ~~für alle $x \in X$~~, so ist . . .
- Seite 224, Zeile −10: . . . positive Semidefinitheit der Gram-Matrix . . .
- Seite 226, Zeile 12: Geben Sie an, welche $x \in \mathbb{R}$ vom zurückgezogenen Klassifizierer . . .
- Seite 230, Zeile 15: die *Rectified Linear Unit*
- Seite 231, Zeile −4: . . . Neuronen aus Proposition 16.2 zusammensetzen
- Seite 232, Zeile 3: . . . sonst,
- Seite 232, Zeile $6 - -7$: . . . ist es manchmal von Vorteil . . .
- Seite 233, Zeile 14: Alternativ erhält man mit Sigmoid-Aktivierung, . . .
- Seite 236, Zeile 9: B ist die Matrix ohne das $(\cdot)^{-1}$.
- Seite 239, Zeile 15: $\mathcal{F}(f_1 * f_2) = \mathcal{F}f_1 \cdot \mathcal{F}f_2$
- Seite 262, Zeile 1–4: Den Rieszschen Darstellungssatz in der Version 16.19 kann man in [Wer18, Theorem II.2.5] nachlesen, die andere Version 16.16, oft auch Riesz-Markov-Theorem genannt, findet man in [Rud87, Theorem 6.19].
- Seite 270, Zeile 14: Dann gilt $h'(t) = \langle \nabla f(y + t(x - y)), x - y\rangle$ . . .
- Seite 270, Zeile 16: $\cdots = \int_0^1 \langle \nabla f(y + t(x - y)), x - y\rangle \,\mathrm{d}t$
- Seite 270, Zeile 18: $\cdots = \int_0^1 \langle \nabla f(y + t(x - y)), x - y\rangle - \langle f(y), x - y\rangle \,\mathrm{d}t$
- Seite 270, Zeile −1: . . . mithilfe von Proposition 17.4 . . .
- Seite 271, Zeile 14: **Folgerung 17.10.** Sei $f\colon \mathbb{R}^n \to \mathbb{R}$ differenzierbar ~~und konvex~~.
- Seite 271, Zeile 18–19: (i)$\Longrightarrow$(ii): Aus der Konvexität folg die erste Ungleichung mit Proposition 17.4 und . . .
- Seite 271, Zeile 20: (ii)$\Longrightarrow$(i): Die erste Ungleichung impliziert Konvexität mit Propsition 17.4 und . . .
- Seite 272, Zeile 16: Aus Lemma 17.9 und Proposition 17.4 folgern wir
- Seite 272, Zeile 20: $\underset{\substack{\uparrow\\ \text{Prop. } 17.4}}{\geqslant}$
- Seite 273, Zeile 10: $\underset{\substack{\uparrow\\ \text{Prop. } 17.4}}{\leqslant}$
- Seite 275, Zeile −2: Dann gilt ~~wegen~~
- Seite 277, Zeile −10: mit Proposition 17.4
- Seite 278, Zeile 8: Dies ist nach Proposition 17.4 äquivalent zu
- Seite 278, Zeile 18: Ein Vergleich der Abschätzung in Proposition 17.4 zeigt nochmal . . .
- Seite 279, Zeile 13: $\underset{\substack{\uparrow\\ \text{Lem. } 17.18 \text{ u.}\\ \text{Prop. } 17.4}}{\geqslant}$
- Seite 283, Zeile −9: 4: $x^{(k+1)} \leftarrow x^{(k)} - \gamma_k \nabla f(x^{(k)})$.
- Seite 286, Zeile 6: (ii) Für $A, B \in \Sigma$ mit $\mathcal{P}(B) \neq 0$ heißt $\mathcal{P}(A|B) := \frac{\mathcal{P}(A \cap B)}{\mathcal{P}(B)}$
- Seite 287, Zeile 19:

$$\rho(x) = \tfrac{1}{(2\pi\sigma^2)^{d/2}}\, \mathrm{e}^{-\frac{\|x-\mu\|^2}{2\sigma^2}} \quad \text{bzw.} \quad \rho(x) = \tfrac{1}{\lambda^d(B)} \cdot \mathbb{1}_B(x),$$

- Seite 290, Zeile −5: Für $A = A_1 \times \cdots \times A_d \subseteq \mathbb{R}^d$ mit $A_i \in \mathcal{B}^{\cancel{d}}$ berechnen wir
- Seite 292, Zeile 2:

$$(\rho_1 * \rho_2)(s) = \frac{1}{2\pi\sqrt{ab}} \int_{\mathbb{R}} \exp\bigl(-\frac{(s-t)^2}{2a}\bigr) \exp\bigl(-\frac{t^2}{2b}\bigr)\,\mathrm{d}t$$

$$
\begin{aligned}
&= \frac{1}{2\pi\sqrt{ab}} \int_{\mathbb{R}} \exp\bigl(-\frac{b(s^2-2st+t^2)+at^2}{2ab}\bigr)\,\mathrm{d}t\\
&= \frac{1}{2\pi\sqrt{ab}} \int_{\mathbb{R}} \exp\bigl(-\frac{t^2(b+a)-2stb+bs^2}{2ab}\bigr)\,\mathrm{d}t\\
&= \frac{1}{2\pi\sqrt{ab}} \int_{\mathbb{R}} \exp\bigl(-\frac{t^2(b+a)/c-2stb/c+bs^2/c}{2ab/c}\bigr)\,\mathrm{d}t\\
&= \frac{1}{\sqrt{2\pi c}\sqrt{2\pi(ab/c)}} \int_{\mathbb{R}} \exp\bigl(-\frac{(t-(bs)/c)^2-(sb/c)^2+s^2(b/c)}{2ab/c}\bigr)\,\mathrm{d}t\\
&= \frac{1}{\sqrt{2\pi c}} \exp(+\frac{(sb/c)^2-s^2(b/c)}{2ab/c})\frac{1}{\sqrt{2\pi(ab/c)}} \int_{\mathbb{R}} \exp\bigl(-\frac{(t-(bs)/c)^2}{2ab/c}\bigr)\,\mathrm{d}t\\
&= \frac{1}{\sqrt{2\pi c}} \exp(+\frac{(sb/c)^2c^2-s^2(b/c)c^2}{2abc})\\
&= \frac{1}{\sqrt{2\pi c}} \exp(+\frac{s^2(b^2-bc)}{2abc})\\
&= \frac{1}{\sqrt{2\pi c}} \exp(-\frac{s^2}{2c}),
\end{aligned}
$$

- Seite 296, Zeile −14: S. Shalev-Shwartz und S. Ben-David, Understanding