

# Errata

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

17. Mai 2025

- Seite 10, Zeile −1:

$$\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n x_i y_i - \bar{y} \sum_{i=1}^n x_i - \bar{x} \sum_{i=1}^n y_i + n \bar{x} \bar{y} = \left( \sum_{i=1}^n x_i y_i \right) - n \bar{x} \bar{y}.$$

- Seite 11, Zeile −1:

$$0 = \sum_{i=1}^n (ax_i + b - y_i) = a \sum_{i=1}^n x_i + nb - \sum_{i=1}^n y_i = an\bar{x} + n\mathbf{b} - n\bar{y},$$

- Seite 13, Zeile 22: Seitenlänge

- Seite 15, Zeile −10: ... konstante Zufallsvariable  $\mathbf{av}$ .

- Seite 15, Zeile −3 und Seite 16, Zeile 2:  $\mathbf{Var}(\overline{x\mathcal{E}})^{(n)}$

- Seite 15, Zeile −2:  $\dots + \frac{2}{n^2} \sum_{i < j} x_i x_j \mathbb{E}(\mathcal{E}_i) \mathbb{E} \mathcal{E}_j)$

- Seite 16, Zeile 3: da  $\overline{x^{(n)^2}} = \mathbf{var}(x^{(n)}) + \overline{x^{(n)}}^2$  als Summe ...

- Seite 17, Zeile 8: Ist Letzteres der Fall, so ist  $\text{sign}(r_{xy}) = \text{sign}(\langle \mathbf{u}, \mathbf{v} \rangle) = \dots$

- Seite 17, Zeile 11:  $r_{xy}$

- Seite 18, Zeile −3:  $\dots f^*(x) = \langle a^*, x \rangle + b^* \dots$

- Seite 19, Zeile −6: ... berechnen wir (mit nur für den Moment  $\phi(\tilde{a}) = \langle \tilde{a}, X^T X \tilde{a} \rangle$ ):

- Seite 25, Zeile −5: Im Bild muss es die  $z$ -Achse sein.

- Page 26, Zeile 9: ... und  $(w, b) = (w_1, \dots, w_d, b)$  für ...

- Seite 26, Zeile 18:  $f(\mathbf{z}) = \text{sig}(\langle w, \cdot \rangle)$

- Seite 26, Zeile −1:  $\mathbb{P}[Y_i(\mathbf{f}) = y_i \text{ für alle } i] = \dots$

- Seite 27, Zeile 2:  $L: \{f: \mathbb{R}^d \rightarrow (0, 1) \mid f \text{ logistische Funktion}\} \rightarrow \mathbb{R}$

- Seite 26, Zeile −1:  $h'(t) = \text{sig}(t) + 1$

- Seite 30, Seite 6: ... überlappen, wenn für jedes  $w \in \mathbb{R}^{d+1} \setminus \{0\}$  ein  $k$  existiert, ...

- Seite 30, Zeile 9: falls  $\text{sig}(\langle w, \hat{x} \rangle) \geq 1/2$

- Seite 31, Zeile 12: ... und  $\langle w, \hat{w}_k \rangle < 0$  gelten.

- Seite 38, Zeile 13: ... und  $x_1, \dots, x_k \in D_1 := \{\mathbf{x} \mid (x, y) \in D\}$  gelten.

- Seite 38, Zeile 14:  $x_1 \in \underset{z \in D_1}{\text{argmin}} \rho(x, z)$  sowie  $x_j \in \underset{z \in D_1 \setminus \{x_1, \dots, x_{j-1}\}}{\text{argmin}} \rho(x, z)$  für  $j \geq 2$

- Seite 38, Zeile 20:  $f(x) \in \underset{y \in Y}{\text{argmax}} N(y)$

- Page 37, Line 1–9:

```
1: function K-NN KLASSIFIZIERER( $D, k, x$ )
2:    $D' \leftarrow D, A \leftarrow \emptyset$ 
3:   for  $j \leftarrow 1$  to  $k$  do
4:      $z^* \leftarrow \underset{z \in D'}{\text{argmin}} \rho(x, \pi_1(z))$ 
5:      $A \leftarrow A \cup \{z^*\}, D' \leftarrow D' \setminus \{z^*\}$ 
6:   for  $y$  in  $Y$  do
7:      $N(y) \leftarrow \#\{a \in A \mid \pi_2(a) = y\}$ 
8:    $\ell \leftarrow \underset{y \in Y}{\text{argmax}} N(y)$ 
9:   return  $\ell$ 
```

Hierbei bezeichnet  $\pi_2(x, y) = y$  die Projektion auf den zweiten Eintrag von  $(x, y) \in D$  und  $y^*$  das Label von  $z^*$ .

- Seite 39, Zeile −2: ... vorgenannten grauen Punkt  $\mathbf{x}$  mit ...

- Seite 41, Zeile 6–7: ... Trainingsdaten  $D_{\text{Train}}$  und Testdaten  $D_{\text{Test}}$ . Dann bestimmt man einen Klassifizierer  $f: X \rightarrow Y$  anhand von  $D_{\text{Train}}$ , und stellt fest, welcher Anteil der Punkte aus  $D_{\text{Test}}$  durch  $f$  korrekt klassifiziert wird.

- Seite 41, Zeile 16: ..., sowie  $k < n$ .
- 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 Multiplikationen 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

$$(d-1) \cdot (n + (n-1) + \dots + (n-k+1)) \leq C \cdot d \cdot k \cdot n$$

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:  $f: X \rightarrow Y, f(x) = \frac{\sum_{i=1}^k w(x_i, x) \cdot y_i}{\sum_{i=1}^k w(x_i, x)}$

- Seite 43, Zeile 11:  $\tilde{x}^{(i)} = \left( a + \frac{(x_1^{(i)} - \min_{j=1, \dots, n} x_1^{(j)})(b-a)}{\max_{j=1, \dots, n} x_1^{(j)} - \min_{j=1, \dots, n} x_1^{(j)}}, \dots \right)$

- Seite 43, Zeile 13:  $\tilde{x}^{(i)} = \left( \frac{x_1^{(i)} - \overline{x_1^{(i)}}}{\sigma_1}, \dots \right)$

- 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:

```

1: function VERKNÜPFUNGSBASIERTES CLUSTERING ( $X, \rho, D, \delta$ )
2:    $k \leftarrow \#D$ 
3:   for  $i \leftarrow 1$  to  $k$  do
4:      $C_i \leftarrow \{x_i\}$ 
5:   while  $\min_{i \neq j} \rho(C_i, C_j) \leq \delta$  and  $k \geq 2$  do
6:      $m \leftarrow 0$ 
7:      $(i^*, j^*) \leftarrow \operatorname{argmin}_{i \neq j} \rho(C_i, C_j)$ 
8:     for  $\ell \leftarrow 1$  to  $k-1$  do
9:       if  $\ell = \min(i^*, j^*)$  then
10:         $C_\ell \leftarrow C_{i^*} \cup C_{j^*}$ 
11:       if  $\ell = \max(i^*, j^*)$  then
12:         $m \leftarrow 1$ 
13:         $C_\ell \leftarrow C_{\ell+m}$ 
14:       else
15:         $C_\ell \leftarrow C_{\ell+m}$ 
16:      $k \leftarrow k-1$ 
17:   return  $C_1, \dots, C_k$ 

```

- Seite 56, Zeile 4: ..., wie in Definition 4.2, ...

- Seite 56, Zeile 8: Aufgabe 4.2

- Seite 56, Zeile -2:  $K: \mathbb{C}_k \rightarrow \mathbb{R}$

- Seite 58, Zeile –16: Der folgende Pseudocode **approximiert** einen Minimierer der  $k$ -means-Kostenfunktion.
- Seite 58, Pseudocode:

```

1: function K-MEANS ( $D, k, X, \rho$ )
2:    $\mu_1, \dots, \mu_k \leftarrow$  pairwise different points from  $X$ 
3:   for  $i \leftarrow 1$  to  $k$  do
4:      $C_i \leftarrow \{x \in D \mid i \in \operatorname{argmin}_{j=1, \dots, k} \rho(x, \mu_j)\}$ 
5:    $U \leftarrow \text{True}$ 
6:   while  $U = \text{True}$  do
7:      $U \leftarrow \text{False}$ 
8:     for  $i \leftarrow 1$  to  $k$  do
9:        $\mu_i \leftarrow \mu(C_i)$ 
10:    for  $i \leftarrow 1$  to  $k$  do
11:       $C'_i \leftarrow \{x \in D \mid i \in \operatorname{argmin}_{j=1, \dots, k} \rho(x, \mu_j)\}$ 
12:      if  $C'_i \neq C_i$ 
13:         $C_i \leftarrow C'_i$ 
14:       $U \leftarrow \text{True}$ 
15:  return  $C_1, \dots, C_k$ 

```

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:

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

- Seite 59, Zeile –5: *Beweis.* Für  $j \geq 1$  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 64, Zeile 10:  $A = (a_{ij})_{i,j=1, \dots, n}$
- Seite 64, Zeile 12:  $L = (\ell_{ij})_{i,j=1, \dots, n}$
- Seite 65, Zeile 11: Wir beginnen mit zwei Beispielen.
- 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  $\lambda_i$  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

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

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

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

- Seite 70, Zeile 3–4: Wir wählen  $v_1 = \mathbf{1}$  als Eigenvektor zum Eigenwert  $\lambda_1$ . Dann gelten für Eigenwert und Eigenvektor  $\lambda_2$  bzw.  $v_2$  mit  $\|v_2\| = 1 \dots$
- Seite 70, Zeile 7:  $v_2 \in \dots$
- 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:

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

- Seite 73, Zeile 2:

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

- Seite 73, Zeile 4:

$$\begin{aligned} \lambda_2(\mathcal{L}) &\underset{\substack{\uparrow \\ \text{Satz} \\ 5.13}}{=} \min_{\substack{x \neq 0 \\ \langle x, D^{1/2}\mathbf{1} \rangle = 0}} \frac{\langle x, \mathcal{L}x \rangle}{\langle x, x \rangle} \stackrel{(*)}{=} \min_{\substack{y \neq 0 \\ \langle D^{1/2}y, D^{1/2}\mathbf{1} \rangle = 0}} \frac{\langle D^{1/2}y, \mathcal{L}D^{1/2}y \rangle}{\langle D^{1/2}y, D^{1/2}y \rangle} \\ &= \min_{\substack{y \neq 0 \\ \langle y, D\mathbf{1} \rangle = 0}} \frac{\langle y, Ly \rangle}{\langle y, Dy \rangle} = \min_{\substack{y \neq 0 \\ \langle Dy, \mathbf{1} \rangle = 0}} \frac{\sum_{\{i,j\} \in E} (y_i - y_j)^2}{\sum_{i=1}^n y_i^2 d_i} \end{aligned}$$

- 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  $r \leq k \leq n-1$  gelten ...
- Seite 73, Zeile –2:

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

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

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

- Seite 75, Zeile 7–8:

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

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

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

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

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

- Seite 77, Zeile 4:

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

- Seite 77, Zeile 14:

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

- Seite 78, Zeile 10:

$$\dots \stackrel{\substack{\uparrow \\ \text{Teleskop-} \\ \text{summe}}}{=} \left( \sum_{i=1}^{n-1} \sum_{j=i+1}^n \mathbb{1}_E(i, j) \sum_{k=i}^{j-1} (p_k^2 - p_{k+1}^2) \right)^2$$

- 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  $\lambda_2$  der normalisierten Laplacematrix von  $G$  die Abschätzung  $\lambda_2 \leq 2$ .  $\square$

- Seite 84, Zeile -3: Sei eine Datenmenge  $D = \{x_1, \dots, x_n\} \subseteq \mathbb{R}^d$  optimieren, ...
- Seite 85, Line 5:

$$\sum_{j=1}^k \|X v_j\|^2 \geq \sum_{j=1}^k \|X w_j\|^2$$

- Seite 85, Zeile -5:

$$\{v_1, \dots, v_k\} \in \underset{\substack{\{\tilde{v}_1, \dots, \tilde{v}_k\} \subseteq \mathbb{R}^d \\ \text{Orthonormal-} \\ \text{system}}}{\text{argmax}} \sum_{j=1}^n \|X \tilde{v}_j\|^2.$$

- Seite 86, Zeile 16: Sei  $D = \{x_1, \dots, x_n\} \subseteq \mathbb{R}^d$  gegeben und ...
- Seite 86, Zeile 19: sukzessive für  $j = 1, \dots, k$  wählen, wobei ...
- Seite 87, Zeile 5:  $\langle w_{k+1}, v_i \rangle = \langle w_{k+1}, \underbrace{v_i - \pi_W(v_i)}_{\perp W} \rangle + \langle w_{k+1}, \underbrace{\pi_W(v_i)}_{\in U} \rangle = 0 + 0 = 0$
- Seite 87, Zeilen 11 und 14:  $\mathcal{A}X$
- Seite 88, Line -11: Sei  $D = \{x_1, \dots, x_n\} \subseteq \mathbb{R}^d$  gegeben und ...
- Seite 88, Zeile -7: ..., falls  $r = \dim \text{span } D > 0$ . Falls  $r = 0$ , so sind alle  $\sigma_k$  Null.
- Seite 88, Zeile -5: Für  $k \geq r + 1$  wenden wir die Gleichung aus dem Beweis von Lemma 6.2 zweimal an. Einerseits auf  $V_k$  mit ONB  $\{v_1, \dots, v_k\}$ , und andererseits auf  $V_{k-1}$  mit ONB  $\{v_1, \dots, v_{k-1}\}$ . Durch Umstellen erhalten wir

$$\sigma_k^2 = \|X v_k\|^2 = \sum_{i=1}^n \text{dist}(x_i, V_{k-1})^2 - \sum_{i=1}^n \text{dist}(x_i, V_k)^2.$$

Weil  $V_r \subseteq V_{k-1} \subseteq V_k$  gilt, impliziert Lemma 6.4 dass beide Terme auf der rechten Seite verschwinden, also  $\sigma_{r+1}, \dots, \sigma_d = 0$  gilt.

- Seite 89, Zeile -9: Bestimmen Sie, durch Anwendung von Theorem 6.3 in Kombination mit einem Python Paket zur Maximierung unter nichtlinearen Nebenbedingungen, einen 1-bestpassenden Unterraum zu ...
- Seite 91, Zeile 12: ... Singulärwert von  $A$ , ...
- Seite 92, Zeile 8:  $\|u\|^2 = \|Av\|^2 = \langle Av, Av \rangle = \dots$
- Seite 95, Zeile 6: Ist  $1 \leq k < p$  und haben wir  $\sigma_{k+1} = \dots = \sigma_p = 0$ , ...
- Seite 96, Zeile -9:  $V \in \mathbb{R}^{d \times d}$
- Seite 96, Zeile -8: Dann sind genau  $r$ -viele der  $\sigma_i$  ungleich Null. Dies sind die Singulärwerte von  $A$  ...
- Seite 98, Zeile 8: ... und  $\{v_1, \dots, v_r\}$  irgendein Orthonormalsystem ...
- Seite 98, Zeile 11: Per Konstruktion sind die im Satz angegebenen  $v_1, \dots, v_r$  ein Orthonormalsystem.
- Seite 98, Zeile 20:

$$\|Av\|^2 = \langle v, A^T Av \rangle = \langle v, B \operatorname{diag}(\dots) B^T v \rangle = \langle B^T v, \operatorname{diag}(\dots) B^T v \rangle = \sum_{j=1}^r \lambda_j^2 (B^T v)_j^2$$

Im gesamten restlichen Beweis müssen die Quadrate an den Eigenwerten  $\lambda_j$  ebenfalls gestrichen werden!

- Seite 101, Zeile -8: ~~Multiplikation mit  $V^T$  von links in~~ Proposition 7.14 liefert also genau eine ...
- Seite 102, Zeile 17:  $r = \operatorname{rk}(A)$
- Seite 106, Zeile 3:  $\forall i, j \in \{1, \dots, n\}$ :  $\|T_{V_k} a_i - T_{V_k} a_j\| - \|a_i - a_j\| \leq 2 \left( \sum_{\ell=k+1}^r \sigma_\ell \right)^{1/2}$
- Seite 107, Zeile -1 bis Seite 108, Zeile 1:

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

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

- Seite 108, Zeile 5:

$$\tilde{A} = \begin{bmatrix} 0.15 & 1.97 & 0.15 & 1.97 & 0.56 \\ 0.92 & 0.01 & 0.92 & 0.01 & 0.94 \\ 4.84 & 0.03 & 4.84 & 0.03 & 4.95 \\ 0.36 & 4.03 & 0.36 & 4.03 & 1.20 \\ 2.92 & -0.00 & 2.92 & -0.00 & 2.98 \\ -0.34 & 4.86 & -0.34 & 4.86 & 0.65 \\ 3.92 & 0.02 & 3.92 & 0.02 & 4.00 \end{bmatrix} = \begin{bmatrix} 0.07 & 0.29 \\ 0.13 & -0.02 \\ 0.68 & -0.11 \\ 0.15 & 0.59 \\ 0.41 & -0.07 \\ 0.07 & 0.73 \\ 0.55 & -0.09 \end{bmatrix} \begin{bmatrix} 12.4 & & \\ & 9.5 & \\ & & 1.3 \end{bmatrix} \begin{bmatrix} 0.56 & 0.09 & 0.56 & 0.09 & 0.59 \\ -0.12 & 0.69 & -0.12 & 0.69 & 0.02 \end{bmatrix}$$

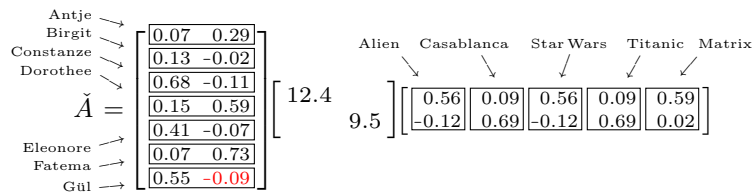
- Seite 109, Zeile 16:

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

- Seite 109, Zeile -2:

$$u_2 = 0.29 \cdot \text{Antje} - 0.02 \cdot \text{Birgit} + \dots - 0.09 \cdot \text{Gül},$$

- Seite 111, Zeile 12:



- Seite 110, Zeile 7:  $\mathbb{R}_{\mathcal{F}}^5$
- Seite 112, Zeile 1: ... und  $\check{V} = \{v_1, v_2\} \dots$
- Seite 112, Zeile 5: (Film) $_{\mathcal{F}}$   $\mathbb{R}_{\mathcal{F}} \xrightarrow{A} \mathbb{R}_{\mathcal{B}}$  (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:  $H_d(\mathbf{1})$
- Seite 129, Zeile -1:  $0 < \varepsilon < 1$
- Seite 131, Zeile 9:  $H_{\delta, x_1} = \{(x_2, \dots, x_d \in \mathbb{R}^{d-1} \mid (x_1, x_2, \dots, x_d) \in H_{\delta})\}$
- Seite 134, Zeile 9: Für solche Zufallsvektoren erhält man ...
- Seite 139, Zeile -1:  $\|X^{(i)}\| \approx 1$
- Seite 146, Zeile 5:  $P[\|X\| - \sqrt{d} \geq \varepsilon]$
- 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:  $\dots = \langle w', x_0 - (\langle w', x_0 \rangle + b')w' \rangle + b'$
- Seite 193, Zeile 12:  $\geq \|x_0 - x_1\|^2 + 2\langle (\langle w', x_0 \rangle + b')w', x_1 - x \rangle$
- 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:  $h: \mathbb{R}^d \rightarrow \mathbb{R}^d$
- 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 199, Zeile 13: ... über die im Satz angegebene Menge zu **minimieren**, ...
- Seite 200, Zeile 7:  $L(x, \theta, \mu) := f(x) - \sum_{i=1}^q \theta_i g_i(x) + \# \sum_{j=r}^q \mu_i h_j(x)$
- 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:  $\lambda_i^*$
- 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  $y_i(\langle w, x_i \rangle + b) \geq 1$  für ....
- Seite 220, Zeile 16: ... mit  $\lambda_{i_0}^* \neq 0$  ist ...
- Seite 220, Zeile 18: ... Wahl von  $\lambda^*$  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:  $\cdots = \int_0^1 \langle \nabla f(y + t(x - y)), x - y \rangle dt$
- Seite 270, Zeile 18:  $\cdots = \int_0^1 \langle \nabla f(y + t(x - y)), x - y \rangle - \langle f(y), x - y \rangle dt$
- Seite 270, Zeile -1: ... mithilfe von **Proposition 17.4** ...
- Seite 271, Zeile 14: **Folgerung 17.10.** Sei  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  differenzierbar **und konvex**.
- Seite 271, Zeile 18-19: (i) $\implies$ (ii): Aus der Konvexität folg die erste Ungleichung mit **Proposition 17.4** und ...
- Seite 271, Zeile 20: (ii) $\implies$ (i): Die erste Ungleichung impliziert Konvexität mit **Proposition 17.4** und ...
- Seite 272, Zeile 16: Aus Lemma 17.9 und **Proposition 17.4** folgern wir
- Seite 272, Zeile 20:  $\geq$   
 $\uparrow$   
**Prop. 17.4**
- Seite 273, Zeile 10:  $\leq$   
 $\uparrow$   
**Prop. 17.4**



- 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:  $\geq$   
 $\uparrow$   
**Lem. 17.18 u.**  
**Prop. 17.4**

- 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:

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

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

$$\begin{aligned} (\rho_1 * \rho_2)(s) &= \frac{1}{2\pi\sqrt{ab}} \int_{\mathbb{R}} \exp\left(-\frac{(s-t)^2}{2a}\right) \exp\left(-\frac{t^2}{2b}\right) dt \\ &= \frac{1}{2\pi\sqrt{ab}} \int_{\mathbb{R}} \exp\left(-\frac{b(s^2 - 2st + t^2) + at^2}{2ab}\right) dt \\ &= \frac{1}{2\pi\sqrt{ab}} \int_{\mathbb{R}} \exp\left(-\frac{t^2(b+a) - 2stb + bs^2}{2ab}\right) dt \\ &= \frac{1}{2\pi\sqrt{ab}} \int_{\mathbb{R}} \exp\left(-\frac{t^2(b+a)/c - 2stb/c + bs^2/c}{2ab/c}\right) dt \\ &= \frac{1}{\sqrt{2\pi c} \sqrt{2\pi(ab/c)}} \int_{\mathbb{R}} \exp\left(-\frac{(t - (bs)/c)^2 - (sb/c)^2 + s^2(b/c)}{2ab/c}\right) dt \\ &= \frac{1}{\sqrt{2\pi c}} \exp\left(+\frac{(sb/c)^2 - s^2(b/c)}{2ab/c}\right) \frac{1}{\sqrt{2\pi(ab/c)}} \int_{\mathbb{R}} \exp\left(-\frac{(t - (bs)/c)^2}{2ab/c}\right) dt \\ &= \frac{1}{\sqrt{2\pi c}} \exp\left(+\frac{(sb/c)^2 c^2 - s^2(b/c)c^2}{2abc}\right) \\ &= \frac{1}{\sqrt{2\pi c}} \exp\left(+\frac{s^2(b^2 - bc)}{2abc}\right) \\ &= \frac{1}{\sqrt{2\pi c}} \exp\left(-\frac{s^2}{2c}\right), \end{aligned}$$

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