Errata

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

8. Juni 2024

- Seite 15, Zeile −3 und Seite 16, Zeile 2: $\mathrm{V}\cancel{\mathrm{ar}}\overline{(x\mathcal{E})}^{(n)}$
- 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: $f\cancel{(z)} = \mathrm{sig}(\langle w, \,\cdot\,\rangle)$
- 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 $x_j \in \operatorname*{argmin}_{z\in D_1\backslash\{x_1,\dots,x_{j-1}\}} \rho(x, z)$ 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: $D_1' \leftarrow D_1'\backslash\{(z^\star, y^\star)\}$
- 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 48, Zeile 5–6: …, dass sich die Texte 1 und 2 deutlich kosinusähnlicher sind als 1 und 3 bzw. 2 und 3.
- Seite 52, Zeile −14: Text D
- 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 −15: D
- Zeile −12: **while** $\min_{i\neq j} \rho(C_i, C_j) \leqslant \delta$ **do**
- Seite 56, Zeile 4: …, wie in Definition 4.2, …
- Seite 56, Zeile 8: Aufgabe 4.2
- Seite 62, Zeile 5: $\operatorname{argmin}_{i=1,2} \|x - \mu_i\|$
- 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 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 74, Zeile 15: $\emptyset \neq S \subset V$
- Seite 74, Zeile 16: $\mathrm{Cheeg}(G) = \frac{\#\partial S}{\operatorname{vol} S}$
- Seite 87, Zeilen 11 und 14: $\cancel{A}X$
- Seite 91, Zeile 12: …Singulärwert von A, …
- Seite 95, Zeile 6: Ist $1 \leqslant k < p$ und haben wir $\sigma_{k+1} = \cdots = \sigma_p = 0$, …
- Seite 102, Zeile 17: $r = \mathrm{rk}(A)$
- Seite 106, Zeile 3: $\forall\, i, j \in \{1, \dots, n\}\colon \big|\|T_{V_k}a_i - T_{V_k}a_j\| - \|a_i - a_j\|\big| \leqslant 2\Big(\sum_{\ell=k+1}^{r} \sigma_\ell\Big)^{1/2}$

- Seite 110, Zeile 7: $\mathbb{R}^5_{\mathcal{F}}$
- Seite 112, Zeile 5: (Film)$_{\mathcal{F}}$ $\quad \mathbb{R}_{\mathcal{F}} \xrightarrow{A} \mathbb{R}_{\mathcal{B}} \quad$ (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: $\dots = \big\langle w', x_0 - (\langle w', x_0\rangle + b')w'\big\rangle + b'$
- Seite 193, Zeile 12: $\geqslant \|x_0 - x_1\|^2 + 2\big\langle (\langle w', x_0\rangle + b')w', x_1 - x\big\rangle$
- 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 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: $\cdots = \displaystyle\int_0^1 \langle \nabla f(y + t(x-y)), x - y\rangle \,\mathrm{d}t$
- Seite 270, Zeile 18: $\cdots = \displaystyle\int_0^1 \langle \nabla f(y + t(x-y)), x - y\rangle - \langle f(y), x - y\rangle \,\mathrm{d}t$
- 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 Propsition 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 296, Zeile −14: S. Shalev-Shwartz und S. Ben-David, Understanding