Wahrscheinlichkeitstheorie und Statistik

Anwendungsbeispiele und Prüfungsvorbereitung

Aufgabe 13-1 (Batching) Betrachten Sie den Machine-Learning Algorithmus zum Erkennen handgeschriebener Ziffern in [1] und beantworten Sie folgende Fragen:

(a) Wie lautet die Verlustfunktion?

(b) Welcher Optimierer wird zum Minimieren der Verlustfunktion eingesetzt?

(c) Welche Gradienten verwendet das Optimierungsverfahren und sind diese zufällig oder deterministisch?

MC 13-2 (Universalität) Sei F eine universelle Menge von neuronalen Netzen $f : \mathbb{R}^d \to \mathbb{R}$ und $g : \mathbb{R}^d \to \mathbb{R}$ eine stetige Funktion. Welche Aussagen gelten für alle $\epsilon > 0$? (Mehrere richtige Antworten sind möglich.)

(a) $\forall n \in \mathbb{N}, \forall x_1, \dots, x_n \in \mathbb{R}^d, \exists f \in F, \forall i \in \{1, \dots, n\} : |f(x_i) - g(x_i)| < \epsilon$.

(b) $\exists f \in F, \forall n \in \mathbb{N}, \forall x_1, \dots, x_n \in \mathbb{R}^d, \forall i \in \{1, \dots, n\} : |f(x_i) - g(x_i)| < \epsilon$.

(c) $\forall n \in \mathbb{N}, \exists f \in F, \forall x \in [-n, n]^d : |f(x) - g(x)| < \epsilon$.

(d) $\exists f \in F, \forall n \in \mathbb{N}, \forall x \in [-n, n]^d : |f(x) - g(x)| < \epsilon$.

MC 13-3 (Approximationsfehler). Sei F eine universelle Menge von neuronalen Netzen $f : \mathbb{R}^d \to \mathbb{R}$. Welche Aussagen gelten für den empirischen Verlust

$$L(f) = \sum_{i=1}^{n} \big(f(X_i(\omega)) - Y_i(\omega)\big)^2$$

bezüglich eines Datensatzes von unabhängigen, identisch verteilten Beobachtungen (X_i, Y_i), $i \in \{1, \dots, n\}$? (Mehrere richtige Antworten sind möglich.)

(a) $\inf_{f \in F} L(f) = \inf_{f \in C(\mathbb{R}^d, \mathbb{R})} L(f)$

(b) $\arg\min_{f \in F} L(f) \subseteq \arg\min_{f \in C(\mathbb{R}^d, \mathbb{R})} L(f)$

(c) $\inf_{f \in F} L(f) = \inf_{f \in F} \mathbb{E}(f(X_1) - Y_1)^2)$

MC 13-4 (Machine Learning als inverses Problem). Wir betrachten den empirischen Verlust eines ReLU Netzes auf einem Datensatz von Paaren reeller Zahlen (x_i, y_i):

$$L(\theta) = \frac{1}{n} \sum_{i=1}^{n} \big(f_\theta(x_i) - y_i\big)^2, \qquad \theta = (w, a, b) \in \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^m, \qquad f_\theta(x) = \sum_{i=1}^{m} w_i \max\{0, a_i x + b_i\}.$$

Welche Aussagen sind für hinreichend grosses m korrekt? (Mehrere richtige Antworten sind möglich.)

(a) Es gibt genau ein $\theta^* \in \arg\min_\theta L(\theta)$.

(b) Es gibt genau ein $\theta^* \in \arg\min_\theta L(\theta)$ und θ^* hängt stetig von den Daten (x_i, y_i) ab.

(c) $\inf_\theta L(\theta)$ hängt stetig von den Daten (x_i, y_i) ab.

(d) $\inf_\theta L(\theta) = 0$.

MC 13-5 (Lösungsbegriff Bayesianische Optimierung). Auf einem Wahrscheinlichkeitsraum $(\Theta, \mathcal{A}, \pi_0)$ betrachten wir eine messbare Funktion $L : \Theta \to \mathbb{R}$ mit Minimierer $\theta^* \in \arg\min_\theta L(\theta)$ und Minimum $m = L(\theta^*)$. Welche der folgenden Aussagen treffen auf π_n-verteilte Zufallsvariablen θ_n zu, wobei

$$\pi_n(d\theta) \propto \exp(-nL(\theta)\pi_0(d\theta), \qquad n \in \mathbb{N}.$$

(Mehrere richtige Antworten sind möglich.)

(a) $L(\theta_n) \to m$ in Wahrscheinlichkeit.

(b) $L(\theta_n) \to 0$ in Wahrscheinlichkeit.

(c) $\theta_n \to \theta^*$ in Wahrscheinlichkeit.

(d) $\theta_n \to \theta^*$ in Wahrscheinlichkeit, falls L stetig ist.

MC 13-6 (Invarianz unter der Langevin-Dynamik). Sei θ_0 eine $\mathbb{R}^d$-wertige Zufallsvariable mit Dichtefunktion proportional zu $\exp(-L(\theta)/\epsilon)$, für eine messbare Funktion $L : \mathbb{R}^d \to \mathbb{R}$ und $\epsilon > 0$. Ferner sei W unabhängig von θ_0 und standardnormalverteilt. Unter welcher Dynamik ist π infinitesimal invariant im Sinn von

$$\frac{d}{dt}\Big|_{t=0} \mathbb{E}\Big(f\big(\theta(t)\big)\Big) = 0, \qquad \text{für alle Testfunktionen } f.$$

(Mehrere richtige Antworten sind möglich.)

(a) $\dot\theta(t) = -\nabla L(\theta(t))$, $\theta(0) = \theta_0$

(b) $\dot\theta(t) = -\nabla L(\theta(t)) + \sqrt{2t/\epsilon}W$, $\theta(0) = \theta_0$

(c) $\theta(t) = \theta_0 - t\nabla L(\theta_0)$

(d) $\theta(t) = \theta_0 - t\nabla L(\theta(0)) + \sqrt{2t/\epsilon}W$

MC 13-7 (Gradientenfluss). Sei $L : \mathbb{R}^d \to \mathbb{R}$ eine glatte Funktion mit einem eindeutigen globalen Minimum θ^* und sei $\theta_0 \in \mathbb{R}^d$. Welche Aussagen sind im Allgemeinen korrekt? (Mehrere richtige Antworten sind möglich.)

(a) Wenn L ein quadratisches Polynom ist, konvergiert der Gradientenfluss gegen θ^*.

(b) Wenn L kein quadratisches Polynom ist, kann der Gradientenfluss divergieren.

(c) Wenn L kein quadratisches Polynom ist, kann der Gradientenfluss gegen $\theta \neq \theta^*$ konvergieren.

(d) Der empirische Verlust eines neuronalen Netzes ist im Allgemeinen kein quadratisches Polynom.

MC 13-8 (Aktivierungsfunktion). Welche der folgenden Funktionen Aktivierungsfunktionen liefern universale neuronale Netze? (Mehrere richtige Antworten sind möglich.)

(a) $f(x) = \max(0, x)$

(b) $f(x) = -\max(0, x)$

(c) $f(x) = \max(0, x^2)$

(d) $f(x) = \max(x/100, x)$

(e) $f(x) = x$

(f) $f(x) = x^2$

Aufgabe 13-9 (Tests für Mittelwert und Standardabweichung). Zwei verschiedene Methoden wurden verwendet, um die Schmelzenthalpie von Wasser zu messen. Methode A lieferte bei 10 Messungen eine durchschnittliche Schmelzenthalpie von 6.0 kJ/mol bei einer Standardabweichung von 0.2 kJ/mol. Methode B lieferte bei 5 Messungen eine durchschnittliche Schmelzenthalpie von 5.6 kJ/mol bei einer Standardabweichung von 0.1 kJ/mol. Die Standardabweichungen wurden ohne Besselkorrektur berechnet.

(a) Testen Sie auf einem Signifikanzniveau von 5 %, ob die beiden Methoden dieselbe Standardabweichung haben.

Hinweis: Der Quotient von zwei unabhängigen χ^2-verteilten Zufallsvariablen ist F-verteilt, und die Quantile der F-Verteilung sind in Tabellen oder Statistik-Software verfügbar.

(b) Testen Sie auf demselben Signifikanzniveau, ob sich die Mittelwerte der beiden Methoden signifikant unterscheiden.

Hinweis: Verwenden Sie den t-Test von Welch [2].

Aufgabe 13-10 (Parameterschätzung für Exponentialverteilung). Die Lebensdauer T (in Stunden) einer Glühbirne unter extremen Bedingungen wird als exponentialverteilt mit unbekanntem Parameter $\lambda > 0$ angenommen. Eine Stichprobe von 5 Glühbirnen wurde bis zum Ausfall beobachtet. Die gemessenen Lebensdauern lauten $T_1 = 4, \quad T_2 = 4, \quad T_3 = 5, \quad T_4 = 3$. Beantworten Sie die folgenden Fragen:

(a) Schätzen Sie den Parameter λ und die mittlere Lebensdauer.

(b) Schätzen Sie die Überlebenswahrscheinlichkeit $S(3) = \mathbb{P}(T > 3)$.

(c) Wenn eine Glühbirne bereits 2 Stunden funktioniert hat, wie viele Stunden wird sie voraussichtlich noch halten?

(d) Wenn zwei Glühbirnen unabhängig voneinander betrieben werden, wie groß ist die Wahrscheinlichkeit, dass beide länger als 3 Stunden halten?

References

[1] Josef Teichmann. *Introduction to Mathematics of New Technologies in Banking and Finance*. Verfügbar unter: `https://gist.github.com/jteichma/ea9452fc6cc307afcc9309741aeed9b7`. Mai 2025.

[2] Wikipedia. *Welch's t-Test*. Verfügbar unter: `https://en.wikipedia.org/wiki/Welch's_t-test`. Mai 2025.