

Wahrscheinlichkeitstheorie und Statistik

Serie 12

MC 12-1 (Neuronale Netze). Unter welchen Operationen ist die Menge der neuronalen Netze mit fixer Aktivierungsfunktion abgeschlossen? (Mehrere korrekte Antworten sind möglich.)

- (a) Linearkombinationen neuronaler Netze sind neuronale Netze.
- (b) Das Produkt neuronaler Netze ist ein neuronales Netz.
- (c) Die Komposition neuronaler Netze ist ein neuronales Netz.
- (d) Das Maximum neuronaler Netze ist ein neuronales Netz.

Aufgabe 12-2 (Universelle Approximation durch ReLU Netze).

- (a) Stellen Sie folgende Dreiecksfunktion als neuronales Netzwerk mit ReLU Aktivierungsfunktion dar:

$$h : \mathbb{R} \rightarrow \mathbb{R}, \quad h(x) = \begin{cases} x + 1, & x \in [-1, 0], \\ 1 - x, & x \in [0, 1], \\ 0, & \text{sonst.} \end{cases}$$

- (b) Zeigen Sie mit Hilfe von (a), dass jede stetige Funktion $f : [0, 1] \rightarrow \mathbb{R}$ gleichmässig durch neuronale Netze mit ReLU Aktivierungsfunktion approximiert werden kann.

Hinweis: Für jedes $n \in \mathbb{N}$ sei die Faber–Schauder Approximation f_n definiert als die stückweise lineare Funktion, die an Vielfachen von 2^{-n} mit f übereinstimmt. Da f gleichmässig stetig ist, konvergieren f_n gleichmässig gegen f .

Definition (Empirische Risikominimierung). Seien $X_1, \dots, X_n$ unabhängige, identisch verteilte Zufallsvariablen. Empirische Risikominimierung ist ein Optimierungsproblem der Form

$$\inf_{\theta} \frac{1}{n} \sum_{i=1}^n L(\theta, X_i),$$

für eine Verlustfunktion L .

Aufgabe 12-3 (Maximum-Likelihood Schätzung als empirische Risikominimierung). Beschreiben Sie in folgenden statistischen Modellen Maximum-Likelihood Schätzung als empirische Risikominimierung und identifizieren Sie die Verlustfunktion L :

- (a) Der unbekannte Parameter ist eine messbare Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$ und die Beobachtungen sind unabhängige identisch verteilte Zufallsvariablen $(X_1, Y_1), \dots, (X_n, Y_n)$, wobei Y_i gegeben X_i normalverteilt mit Mittelwert $f(X_i)$ und Standardabweichung 1 ist.
- (b) Der unbekannte Parameter ist eine messbare Funktion $p : \mathbb{R} \rightarrow \mathbb{R}$ und die Beobachtungen sind unabhängige identisch verteilte Zufallsvariablen $(X_1, Y_1), \dots, (X_n, Y_n)$, wobei Y_i gegeben X_i Bernoulli-verteilt ist mit Erfolgswahrscheinlichkeit $p(X_i)$.

MC 12-4 (Verrauschte Gradienten). Im Kontext von empirischer Risikominimierung definieren wir die Gradienten

$$g_n(\theta) = \frac{1}{n} \sum_{i=1}^n \nabla L(\theta, X_i), \quad g(\theta) = \mathbb{E}(\nabla L(\theta, X_1)),$$

wobei wir annehmen, dass $\nabla L(\theta, X_1)$ quadrat-integrierbar ist. Welche Aussagen sind korrekt? (Mehrere richtige Antworten sind möglich.)

- (a) $g_n(\theta)$ ist in Erwartung gleich $g(\theta)$.
- (b) $g_n(\theta)$ konvergiert in Wahrscheinlichkeit gegen $g(\theta)$.
- (c) $g_n(\theta)$ hat Varianz von Ordnung $O(1/n)$.
- (d) $\sqrt{n}(g_n(\theta) - g(\theta))$ konvergiert in Verteilung gegen eine Normalverteilung.

Aufgabe 12-5 (Implementierung). Verwenden Sie die Software-Bibliothek Ihrer Wahl, um die Funktion f im statistischen Modell aus Aufgabe 12-3 (a) für folgende Daten (X_i, Y_i) , $i \in \{1, \dots, 100\}$, zu lernen:

```
import numpy as np
x = np.random.uniform(-2, 2, 100)
y = np.random.normal(10*np.sin(x**2), 1)
```