



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

Fachbereich Mathematik

Masterarbeit

**Analyse der Konvergenzgeschwindigkeit eines einfach  
berechenbaren Neuronale-Netze-Regressionsschätzers**

Adrian Gabel

8. April 2020

Betreuer: Prof. Dr. Michael Kohler



## **Erklärung zur Abschlussarbeit gemäß § 22 Abs. 7 und § 23 Abs. 7 APB TU Darmstadt**

Hiermit versichere ich, Adrian Gabel, die vorliegende Master-Thesis gemäß § 22 Abs. 7 APB der TU Darmstadt ohne Hilfe Dritter und nur mit den angegebenen Quellen und Hilfsmitteln angefertigt zu haben. Alle Stellen, die Quellen entnommen wurden, sind als solche kenntlich gemacht worden. Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Mir ist bekannt, dass im Falle eines Plagiats (§ 38 Abs. 2 APB) ein Täuschungsversuch vorliegt, der dazu führt, dass die Arbeit mit 5,0 bewertet und damit ein Prüfungsversuch verbraucht wird. Abschlussarbeiten dürfen nur einmal wiederholt werden.

Bei der abgegebenen Thesis stimmen die schriftliche und die zur Archivierung eingereichte elektronische Fassung gemäß § 23 Abs. 7 APB überein.

Darmstadt, 8. April 2020

Adrian Gabel



# Inhaltsverzeichnis

<b>Einleitung</b>	<b>6</b>
<b>1 Grundlagen für neuronale Netze</b>	<b>8</b>
<b>2 Konstruktion eines Neuronale-Netze-Schätzers</b>	<b>22</b>
2.1 Definition der Netzwerkarchitektur . . . . .	24
2.2 Bestimmung der Gewichte der Ausgabeschicht . . . . .	32
<b>3 Resultat zur Konvergenzgeschwindigkeit</b>	<b>36</b>
3.1 Approximationsresultate für Hauptsatz 3.1 . . . . .	37
3.2 Der Beweis von Hauptsatz 3.1 . . . . .	39
3.2.1 Abschätzung von $T_{1,n}$ . . . . .	40
3.2.2 Abschätzung von $T_{2,n}$ . . . . .	41
<b>4 Anwendungsbeispiel auf simulierte Daten</b>	<b>50</b>
<b>Appendix</b>	<b>54</b>
<b>Literaturverzeichnis</b>	<b>65</b>

# Einleitung

Erfinder träumen schon lange davon, Maschinen zu schaffen, die denken. Dieser Wunsch geht zumindest auf die Zeit des antiken Griechenlands zurück. Die mythischen Figuren Pygmalion, Daedalus und Hephaestus können alle als legendäre Erfinder interpretiert werden, und Galatea, Talos und Pandora können alle als künstliches Leben betrachtet werden (Ovid und Martin, 2004; Sparkes, 1996; Tandy, 1997).

Als programmierbare Computer zum ersten Mal konzipiert wurden, fragten sich die Menschen, ob solche Maschinen intelligent werden könnten, mehr als hundert Jahre bevor man sie baute (Lovelace, 1842). Heute ist die künstliche Intelligenz (KI) ein blühendes Feld mit vielen praktischen Anwendungen und aktiven Forschungsthemen. Künstliche Intelligenz ist längst in unserem Alltag präsent und dringt in immer mehr Bereiche vor. Sprachassistenten etwa sind bereits als Helfer auf dem Smartphone, im Auto oder zu Hause Normalität geworden. Fortschritte im Bereich der KI beruhen vor allem auf der Verwendung neuronaler Netze. Vergleichbar mit der Funktionsweise des menschlichen Gehirns verknüpfen sie mathematisch definierte Einheiten miteinander.

Es besteht eine große Lücke zwischen den Schätzungen, für die schönen Konvergenzergebnisse die in der Theorie nachgewiesen wurden, und den Schätzungen, die in der Praxis verwendet werden können.

Maschinelle Lernverfahren können *als Lernen einer Funktion* ( $f$ ) zusammengefasst werden, die Eingangsvariablen ( $X$ ) auf Ausgangsvariablen ( $Y$ ) abbildet.

$$Y = f(x)$$

Ein Algorithmus lernt diese Zielabbildungsfunktion aus Trainingsdaten. Die Form der Funktion ist unbekannt, sodass es unsere Aufgabe als Praktiker des maschinellen Lernens ist, verschiedene Algorithmen des maschinellen Lernens zu evaluieren und zu sehen, welcher die zugrunde liegende Funktion besser annähert. Unterschiedliche Algorithmen machen unterschiedliche Annahmen oder Vorurteile über die Form der Funktion und die Art und Weise, wie sie gelernt werden kann.

Algorithmen, die keine starken Annahmen über die Form der Abbildungsfunktion treffen, werden als nichtparametrische Algorithmen des maschinellen Lernens bezeichnet. Indem

sie keine Annahmen treffen, können sie jede beliebige Funktionsform aus den Trainingsdaten lernen. Nichtparametrische Methoden sind gut, wenn Sie über viele Daten und keine Vorkenntnisse verfügen und wenn Sie sich nicht allzu sehr um die Auswahl der richtigen Funktionen kümmern wollen. (Referenz - Artificial Intelligence: A Modern Approach, Seite 757)

Ziel dieser Arbeit ist es, die folgende Frage genauer zu betrachten: Wenn wir eine Regressionsschätzung des neuronalen Netzes theoretisch genau so definieren, wie sie in der Praxis umgesetzt wird, welches Konvergenzergebnis können wir dann für diese Schätzung vorweisen?

Als Erstes werden wir in Kapitel 1 grundlegende Definition und Hilfsresultate für den weiteren Verlauf der Arbeit sammeln. Anschließend definieren wir in Kapitel 2 eine neue Regressionsschätzung für neuronale Netze, bei der die meisten Gewichte unabhängig von den Daten gewählt werden, die durch einige neuere Approximationsergebnisse für neuronale Netze motiviert sind, und die daher leicht zu implementieren ist. In Kapitel 3 zeigen wir dann unser Hauptresultat, dass wir für diese Schätzung Konvergenzraten ableiten können, falls die Regressionsfunktion glatt ist.

Mit diesem Hauptergebnis ist es nun möglich, einfach zu implementierende Regressionsverfahren für neuronale Netze zu definieren, die die gleiche Konvergenzrate wie lineare Regressionsschätzungen (z.B. Kernel- oder Spline-Schätzungen) erreichen, d.h. sie erreichen (bis zu einem logarithmischen Faktor) die optimale Minimax-Konvergenzrate  $n - 2p/(2p + d)$  im Falle einer  $(p, C)$ -glatten Regressionsfunktion, für jedes  $p > 0$ . Abschließend wird die Leistung unseres neu vorgeschlagenen Schätzers für simulierte Daten in Kapitel 4 veranschaulicht. Diese Arbeit orientiert sich an [AB19].

# Kapitel 1

## Grundlagen für neuronale Netze

Der Zweck dieses Kapitels ist es auf die in dieser Arbeit verwendeten neuronalen Netze einzugehen, die als Bausteine für unseren einfach berechenbaren Neuronale-Netze-Regressionsschätzer verwendet werden. Weiterhin werden wir Approximationsresultate darstellen und beweisen, welche wir für das Resultat der Konvergenzgeschwindigkeit des einfach berechenbaren Neuronale-Netze-Regressionsschätzers benötigen werden. Wenn wir in dieser Arbeit von *unserem Neuronale-Netze-Regressionsschätzer* sprechen, beziehen wir uns immer auf den *einfach berechenbaren Neuronale-Netze-Regressionsschätzer* aus [AB19]. In dieser Arbeit bezeichnen wir mit  $d \geq 1$  immer eine natürliche Zahl.

Da wir uns in dieser Arbeit mit neuronalen Netzen beschäftigen, ist es hilfreich zu wissen, was man darunter versteht. Ein neuronales Netz ist nichts anderes als eine Ansammlung von Neuronen, welche als Informationsverarbeitungseinheiten dienen, die schichtweise in einer Architektur angeordnet sind. Beginnend mit der Eingabeschicht (*Input Layer*) fließen Informationen über eine oder mehrere verborgene Schichten (*Hidden Layer*) bis hin zur Ausgabeschicht (*Output Layer*). Die Informationsweitergabe der Neuronen verläuft so, dass sie die Eingaben  $x_1, \dots, x_n$ , die einerseits aus dem beobachteten Prozess resultieren können, dessen Werte dem Neuron übergeben werden, oder andererseits aus den Ausgaben anderer Neuronen stammen und verarbeiten. Dazu werden für ein Neuron  $j$  die Eingaben mit  $w_{1j}, \dots, w_{nj}$  gewichtet an eine Aktivierungsfunktion  $\sigma$  übergeben, welche die Neuronenaktivierung berechnet. Der Endpunkt des Informationsflusses in einem neuronalen Netz ist die Ausgabeschicht, die hinter den verborgenen Schichten liegt. Sie bildet damit die letzte Schicht in einem neuronalen Netz. Die Ausgabeschicht enthält somit das Ergebnis der Informationsverarbeitung durch das Netz. Zwei wichtige Charakteristika, die neuronale Netze aufweisen können, sind:

- Wenn in einem neuronalen Netz die Information von der Eingabeschicht über die verborgenen Schichten bis hin zur Ausgabeschicht in eine Richtung („vorwärts“)



weitergereicht wird, spricht man von einem *feedforward* neuronalen Netz.

- Ein neuronales Netz wird als ein *fully connected* („vollständig verbundenes“) neuronales Netz bezeichnet, wenn sämtliche Neuronen einer Schicht mit allen der darauffolgenden verbunden sind. Da man bei fully connected neuronalen Netzen die Gewichte der Verbindungen zwischen zwei Neuronen auf Null setzen kann, unterscheiden wir hier nicht mehr zwischen neuronalen Netzen die fully connected sind oder nicht.

Als Nächstes kommen wir zur mathematischen Definition eines neuronalen Netzes, welche wir in dieser Form im weiteren Verlauf dieser Arbeit benötigen werden.

**Definition 1.1.** Sei  $L \in \mathbb{N}$ ,  $\mathbf{k} \in \mathbb{N}^L$  und  $\sigma: \mathbb{R} \rightarrow \mathbb{R}$ . Ein *mehrschichtiges feedforward neuronales Netz mit Architektur  $(L, \mathbf{k})$  und Aktivierungsfunktion  $\sigma$* , ist eine reellwertige Funktion  $f: \mathbb{R}^d \rightarrow \mathbb{R}$  definiert durch:

$$f(x) = \sum_{i=1}^{k_L} c_i^{(L)} \cdot f_i^{(L)}(x) + c_0^{(L)}$$

mit Gewichten  $c_0^{(L)}, \dots, c_{k_L}^{(L)} \in \mathbb{R}$  und für  $f_i^{(L)}$  mit  $i = 1, \dots, k_L$  rekursiv definiert durch:

$$f_i^{(r)}(x) = \sigma \left( \sum_{j=1}^{k_{r-1}} c_{i,j}^{(r-1)} \cdot f_j^{(r-1)}(x) + c_{i,0}^{(r-1)} \right) \quad (1.1)$$

mit Gewichten  $c_{i,0}^{(r-1)}, \dots, c_{i,k_{r-1}}^{(r-1)} \in \mathbb{R}$  mit  $r = 2, \dots, L$  und:

$$f_i^{(1)}(x) = \sigma \left( \sum_{j=1}^d c_{i,j}^{(0)} \cdot x^{(j)} + c_{i,0}^{(0)} \right)$$

für die Gewichte  $c_{i,0}^{(0)}, \dots, c_{i,d}^{(0)} \in \mathbb{R}$ , weiterhin sei  $\mathfrak{N}(L, \mathbf{k}, \sigma)$  die Klasse aller mehrschichtigen feedforward neuronalen Netze mit Architektur  $(L, \mathbf{k})$  und Aktivierungsfunktion  $\sigma$ .

Wir werden ab jetzt der Einfachheit halber nur noch von *neuronalen Netzen mit Architektur  $(L, \mathbf{k})$*  reden und beziehen uns damit immer auf Definition 1.1. Wenn aus dem Kontext die *Architektur  $(L, \mathbf{k})$*  bereits bekannt ist, sprechen wir nur noch von einem *neuronalen Netz*, beziehen uns damit aber dennoch auf Definition 1.1.

*Bemerkung 1.2.* In obiger Definition versteht man unter:

- $L \in \mathbb{N}$  die Anzahl an verborgenen Schichten von  $f$ .

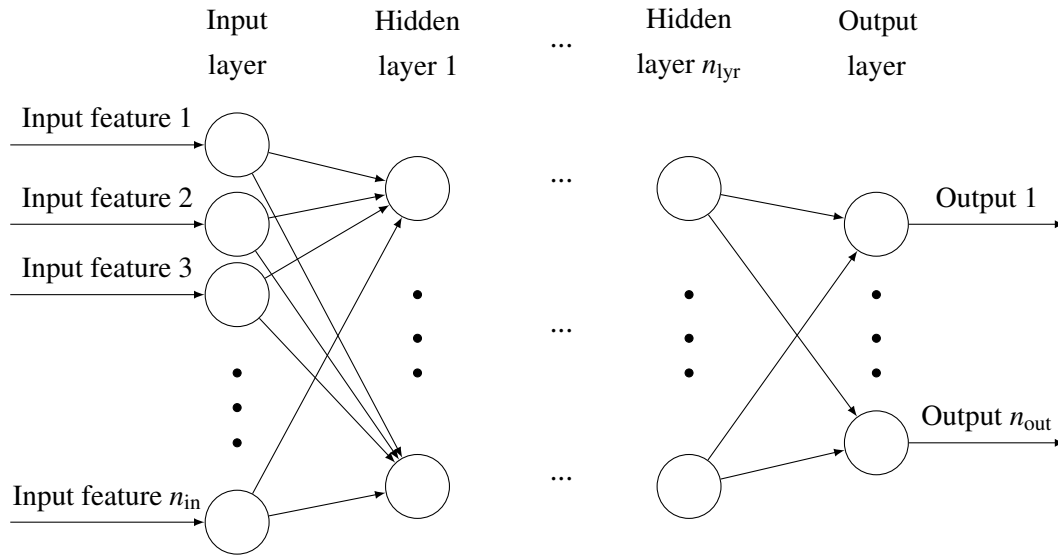


Abbildung 1.1: Mehrschichtiges feedforward neuronales Netz mit einer Eingabeschicht mit  $n_{in}$  Neuronen,  $n_{lyr}$  vielen verborgenen Schichten, deren Anzahl an Neuronen variieren kann und einer Ausgabeschicht bestehend aus  $n_{out}$  Neuronen.

- $\mathbf{k} = (k_1, \dots, k_L) \in \mathbb{N}^L$  einen Vektor, der die Anzahl an Neuronen in jeder verborgenen Schicht angibt.

Da wir bei Neuronale-Netze-Regressionsschätzern Funktionswerte schätzen möchten, haben wir in der Ausgabeschicht nur ein Neuron. Damit erreichen wir einen eindimensionalen Output. Wie wir an der Konstruktion des neuronalen Netzes erkennen können, nehmen wir in der Ausgabeschicht die Identität als Aktivierungsfunktion.

Abbildung 1.1 zeigt schematisch ein mehrschichtiges feedforward neuronales Netz, welches aus einer Eingabeschicht (*Input feature 1 - Input feature  $n_{in}$* ),  $n_{lyr}$  verborgenen Schichten und einer Ausgabeschicht (*Output 1 - Output  $n_{out}$* ) besteht.

Wie zuvor erwähnt ist einer der Ausgangspunkte für die Definition eines neuronalen Netzes die Wahl einer Aktivierungsfunktion  $\sigma: \mathbb{R} \rightarrow \mathbb{R}$ . Wir haben uns in dieser Arbeit für eine sogenannte *Squashing Function* entschieden, welche eine monoton wachsende Funktion ist, für die  $\lim_{x \rightarrow -\infty} \sigma(x) = 0$  und  $\lim_{x \rightarrow \infty} \sigma(x) = 1$  gilt. In dieser Arbeit verwenden wir als Squashing Function immer den sogenannten *sigmoidal* bzw. *logistischen Squasher*

$$\sigma(x) = \frac{1}{1 + \exp(-x)} \quad (x \in \mathbb{R}). \quad (1.2)$$

Wir kommen nun zu einer Eigenschaft, die wir für den logistischen Squasher benötigen und in Lemma 1.4 für  $N \in \mathbb{N}_0$  zeigen werden.

**Definition 1.3.** Sei  $N \in \mathbb{N}_0$ . Eine Funktion  $\sigma: \mathbb{R} \rightarrow [0, 1]$  wird *N-zulässig* genannt, wenn sie monoton wachsend und lipschitzstetig ist und wenn zusätzlich die folgenden drei Bedingungen erfüllt sind:

- (i) Die Funktion  $\sigma$  ist  $(N + 1)$ -mal stetig differenzierbar mit beschränkten Ableitungen.
- (ii) Es existiert ein Punkt  $t_\sigma \in \mathbb{R}$  mit

$$\frac{\partial^i}{\partial x^i} \sigma(t_\sigma) \neq 0 \text{ für } i = 0, \dots, N.$$

- (iii) Wenn  $y > 0$  ist, gilt  $|\sigma(y) - 1| \leq \frac{1}{y}$ . Wenn  $y < 0$  ist, gilt  $|\sigma(y)| \leq \frac{1}{|y|}$ .

Als Nächstes stellen wir ein Resultat vor, welches eine wichtige Eigenschaft für die Aktivierungsfunktion der hier betrachteten neuronalen Netze darstellt.

**Lemma 1.4.** Sei  $N \in \mathbb{N}_0$  beliebig, dann ist der logistische Squasher

$$\sigma: \mathbb{R} \rightarrow [0, 1] \text{ mit } \sigma(x) = \frac{1}{1 + \exp(-x)}$$

*N-zulässig.*

*Beweis.* Sei  $N \in \mathbb{N}_0$  beliebig. Dadurch, dass  $0 \leq \frac{1}{1 + \exp(-x)} \leq 1$  für alle  $x \in \mathbb{R}$  gilt, erhalten wir schließlich auch  $\sigma: \mathbb{R} \rightarrow [0, 1]$ . Wir wissen, dass  $\sigma$  monoton wachsend ist, da für beliebige  $s, t \in \mathbb{R}$  mit  $s \leq t$  gilt:

$$\sigma(s) = \frac{1}{1 + \exp(-s)} \leq \frac{1}{1 + \exp(-t)} = \sigma(t),$$

wobei wir verwendet haben, dass aufgrund der Monotonie der Exponentialfunktion für  $s \leq t$  auch  $\exp(-s) \geq \exp(-t)$  gilt. Zudem ist  $\sigma$  als Komposition glatter Funktionen insbesondere  $(N + 1)$ -mal stetig differenzierbar. Die Funktion  $\sigma$  erfüllt die gewöhnliche Differentialgleichung  $\frac{\partial \sigma}{\partial x} = (1 - \sigma) \cdot \sigma$ , da:

$$\begin{aligned} \frac{\partial \sigma}{\partial x}(x) &= -\frac{1}{(1 + \exp(-x))^2} \cdot (-\exp(-x)) \\ &= \frac{\exp(-x)}{1 + \exp(-x)} \cdot \frac{1}{1 + \exp(-x)} \\ &= \left(1 - \frac{1}{1 + \exp(-x)}\right) \cdot \frac{1}{1 + \exp(-x)} \\ &= (1 - \sigma(x)) \cdot \sigma(x). \end{aligned}$$

Mit der Leibnizregel aus der Differentialrechnung erhalten wir:

$$\frac{\partial^n \sigma}{\partial x^n} = \frac{\partial^{n-1}}{\partial x^{n-1}} \left( (1 - \sigma(x)) \cdot \sigma(x) \right) = \sum_{k=0}^{n-1} \binom{n-1}{k} (1 - \sigma(x))^{(n-1-k)} \sigma(x)^{(k)} \quad (1.3)$$

für  $n = 1, \dots, N$ . Damit erkennen wir, dass alle Ableitungen von  $\sigma$ , Polynome in  $\sigma$  sind. Dadurch folgt Bedingung (i) aus Definition 1.3, da  $\sigma$  nach Voraussetzung beschränkt ist, und die Ableitungen von  $\sigma$  als endliche Summe von Produkten beschränkter Faktoren, ebenfalls beschränkt ist. Da hiermit auch insbesondere die erste Ableitung von  $\sigma$  beschränkt ist, wissen wir, dass  $\sigma$  lipschitzstetig ist.

Nun kommen wir zum Beweis von Bedingung (ii). Da  $\sigma(x) > 0$  für alle  $x \in \mathbb{R}$  gilt, bleibt dies auch bei Potenzen und positiven Linearkombinationen erhalten. Daher wissen wir durch Gleichung (1.3), dass ein  $t_\sigma \in \mathbb{R}$  existiert, mit

$$\frac{\partial^i}{\partial x^i} \sigma(t_\sigma) \neq 0 \text{ für } i = 0, \dots, N.$$

Damit ist Bedingung (ii) ebenfalls erfüllt.

Nun zu Bedingung (iii). Wir wissen, dass für  $x \in \mathbb{R}$  und damit insbesondere für ein beliebiges  $x > 0$  die Ungleichung

$$x \leq \exp(x) + 1$$

gilt. Weiterhin erhalten wir

$$|\sigma(x) - 1| = 1 - \frac{1}{1 + \exp(-x)} = \frac{\exp(-x)}{1 + \exp(-x)}. \quad (1.4)$$

Zudem folgt aus  $x \leq \exp(x) + 1$  die Ungleichung  $x \cdot \exp(-x) \leq 1 + \exp(-x)$  und damit  $\frac{\exp(-x)}{1 + \exp(-x)} \leq \frac{1}{x}$ . Mithilfe von Gleichung (1.4) erhalten wir nun

$$|\sigma(x) - 1| \leq \frac{1}{x}.$$

Dies zeigt den ersten Teil von Definition 1.3 (iii).

Der zweite Teil folgt durch die gleiche Art und Weise, da wir durch

$$\frac{1}{1 + \exp(x)} - \frac{1}{2} = \sigma(0 - x) - \frac{1}{2} = -\sigma(0 + x) + \frac{1}{2} = -\frac{1}{1 + \exp(-x)} + \frac{1}{2} \quad (1.5)$$

wissen, dass  $\sigma$  punktsymmetrisch in  $(0, \frac{1}{2})$  ist. Gleichung (1.5) folgt aus  $\frac{1}{1 + \exp(x)} + \frac{1}{1 + \exp(-x)} = 1$ . Dies erhalten wir durch  $\frac{1}{1 + \exp(x)} - \frac{1}{2} + \frac{1}{1 + \exp(-x)} - \frac{1}{2} = 0$ , indem wir die Brüche auf einen gemeinsamen Nenner bringen und erhalten dadurch  $\frac{2 + \exp(-x) + \exp(x)}{2 + \exp(-x) + \exp(x)} - 1 = 0$ , woraus Gleichung (1.5) folgt. Aus der Punktsymmetrie folgt mit

$$\sigma(-x) - 1 = \frac{1}{1 + \exp(x)} - 1 = -\frac{1}{1 + \exp(-x)} = -\sigma(x)$$

für  $x < 0$  aus dem ersten Teil von Definition 1.3 (iii), da  $-x > 0$  ist:

$$|\sigma(x)| = |-\sigma(x)| = |\sigma(-x) - 1| \leq \frac{1}{-x} = \frac{1}{|x|}.$$

Damit haben wir alle drei Bedingungen aus Definition 1.3 gezeigt und unsere Aussage bewiesen.  $\square$

Als Nächstes stellen wir ein Resultat zur Taylorformel mit Rest vor, welches wir im weiteren Verlauf der Arbeit benötigen werden.

**Lemma 1.5** (Lagrangesche Form des Restglieds, [Kö04, Kapitel 2.4 Seite 67]).

**i)** Sei  $U \subseteq \mathbb{R}^d$  und  $f: U \rightarrow \mathbb{R}$  eine  $(N+1)$ -mal stetig differenzierbare Funktion und  $u, x \in U$  Punkte, deren Verbindungsstrecke in  $U$  liegt, so gilt:

$$f(x) = T_N f(x; u) + R_{N+1}(x; u),$$

wobei

$$T_N(f; u) := \sum_{k=0}^N \frac{f^{(k)}(u)}{k!} (x-u)^k,$$

das Taylorpolynom der Ordnung  $N$  von  $f$  in  $u$  ist und das Restglied mit einem geeigneten Punkt  $\xi$  auf der Verbindungsstrecke zwischen  $u$  und  $x$ , in der Form

$$R_{N+1}(x; u) = \frac{f^{(N+1)}(\xi)}{(N+1)!} (x-u)^{N+1}$$

dargestellt werden kann.

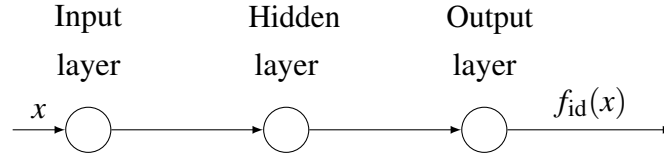
**ii)** Sei  $I \subseteq \mathbb{R}$  ein Intervall und  $f: I \rightarrow \mathbb{R}$  eine  $(N+1)$ -mal stetig differenzierbare Funktion und  $u, x \in I$ . Dann existiert ein  $\xi$  zwischen  $u$  und  $x$ , so dass

$$f(x) = \sum_{k=0}^N \frac{f^{(k)}(u)}{k!} (x-u)^k + \frac{f^{(N+1)}(\xi)}{(N+1)!} (x-u)^{N+1}.$$

Es folgen nun drei Approximationsresultate, welche wir in Kapitel 2 für die Konstruktion unseres Neuronale-Netze-Regressionschätzers benötigen werden.

Wir betrachten als Erstes eine Approximation der Identität durch das neuronale Netz  $f_{\text{id}}: \mathbb{R} \rightarrow \mathbb{R}$  mit Architektur  $(1, (1))$ , welches in Abbildung 1.2 veranschaulicht wird und anschließend die Approximation eines Quadrats auf einem Intervall durch das neuronale Netz  $f_{\text{sq}}: \mathbb{R} \rightarrow \mathbb{R}$  mit Architektur  $(1, (2))$ .

**Lemma 1.6.** Sei  $\sigma: \mathbb{R} \rightarrow \mathbb{R}$  eine Funktion,  $R, a > 0$  und  $\|\cdot\|_\infty$  die Supremumsnorm auf  $[-a, a]$ .

Abbildung 1.2: Neuronales Netz  $f_{\text{id}}(x)$  mit Architektur  $(1, (1))$ .

- a) Angenommen  $\sigma$  ist zweimal stetig differenzierbar und  $t_{\sigma, \text{id}} \in \mathbb{R}$  so, dass  $\sigma'(t_{\sigma, \text{id}}) \neq 0$  ist. Dann gilt für das neuronale Netz

$$f_{\text{id}}(x) = \frac{R}{\sigma'(t_{\sigma, \text{id}})} \cdot \left( \sigma\left(\frac{x}{R} + t_{\sigma, \text{id}}\right) - \sigma(t_{\sigma, \text{id}}) \right)$$

für beliebiges  $x \in [-a, a]$ :

$$|f_{\text{id}}(x) - x| \leq \frac{\|\sigma''\|_{\infty} \cdot a^2}{2 \cdot |\sigma'(t_{\sigma, \text{id}})|} \cdot \frac{1}{R}.$$

- b) Angenommen  $\sigma$  ist dreimal stetig differenzierbar und  $t_{\sigma, \text{sq}} \in \mathbb{R}$  so, dass  $\sigma''(t_{\sigma, \text{sq}}) \neq 0$  ist. Dann gilt für das neuronale Netz

$$f_{\text{sq}}(x) = \frac{R^2}{\sigma''(t_{\sigma, \text{sq}})} \cdot \left( \sigma\left(\frac{2 \cdot x}{R} + t_{\sigma, \text{sq}}\right) - 2 \cdot \sigma\left(\frac{x}{R} + t_{\sigma, \text{sq}}\right) + \sigma(t_{\sigma, \text{sq}}) \right)$$

für beliebiges  $x \in [-a, a]$ :

$$|f_{\text{sq}}(x) - x^2| \leq \frac{5 \cdot \|\sigma'''\|_{\infty} \cdot a^3}{3 \cdot |\sigma''(t_{\sigma, \text{sq}})|} \cdot \frac{1}{R}.$$

*Beweis.* a) Wir wissen, dass  $f_{\text{id}}$  zweimal differenzierbar ist, da nach Voraussetzung  $\sigma$  zweimal stetig differenzierbar ist. Damit folgt mit Lemma 1.5 für  $f = f_{\text{id}}, N = 1, u = 0$  und  $I = [-a, a]$ , dass ein  $\xi$  zwischen 0 und  $x$  existiert, so dass,

$$f_{\text{id}}(x) = \sum_{k=0}^1 \frac{f_{\text{id}}^{(k)}(0)}{k!} x^k + \frac{f_{\text{id}}^{(2)}(\xi)}{2!} x^2.$$

Es gilt:

$$f'_{\text{id}}(x) = \frac{1}{\sigma'(t_{\sigma, \text{id}})} \cdot \sigma'\left(\frac{x}{R} + t_{\sigma, \text{id}}\right).$$

Mit  $f_{\text{id}}(0) = 0$  und  $f'_{\text{id}}(0) = 1$  erhalten wir:

$$\begin{aligned} & |f_{\text{id}}(x) - x| \\ &= \left| 0 + 1 \cdot x + \frac{1}{2} \cdot \frac{1}{R \cdot \sigma'(t_{\sigma, \text{id}})} \sigma''\left(\frac{\xi}{R} + t_{\sigma, \text{id}}\right) \cdot x^2 - x \right| \\ &= \left| \frac{\sigma''\left(\frac{\xi}{R} + t_{\sigma, \text{id}}\right) \cdot x^2}{2R \cdot \sigma'(t_{\sigma, \text{id}})} + x - x \right| \\ &\leq \frac{\|\sigma''\|_{\infty} \cdot a^2}{2 \cdot |\sigma'(t_{\sigma, \text{id}})|} \cdot \frac{1}{R}, \end{aligned}$$

wobei sich die letzte Ungleichung aus den Eigenschaften der Supremumsnorm und  $x^2 \leq a^2$  ergibt, was aus  $x \in [-a, a]$  folgt. Daraus erhalten wir die Behauptung.

- b) Die Funktion  $f_{\text{sq}}$  ist dreimal differenzierbar, da  $\sigma$  nach Voraussetzung dreimal stetig differenzierbar ist. Wie in a) folgt durch Lemma 1.5 mit  $f = f_{\text{sq}}, N = 2, u = 0$  und  $I = [-a, a]$ , dass ein  $\xi$  zwischen 0 und  $x$  existiert, so dass:

$$f_{\text{sq}}(x) = \sum_{k=0}^2 \frac{f_{\text{sq}}^{(k)}(0)}{k!} x^k + \frac{f_{\text{sq}}^{(3)}(\xi)}{3!} x^3.$$

Es gilt:

$$f'_{\text{sq}}(x) = \frac{R^2}{\sigma''(t_{\sigma, \text{sq}})} \cdot \left( \frac{2}{R} \cdot \sigma' \left( \frac{2x}{R} + t_{\sigma, \text{sq}} \right) - \frac{2}{R} \cdot \sigma' \left( \frac{x}{R} + t_{\sigma, \text{sq}} \right) \right)$$

und

$$f''_{\text{sq}}(x) = \frac{R^2}{\sigma''(t_{\sigma, \text{sq}})} \cdot \left( \frac{4}{R^2} \cdot \sigma'' \left( \frac{2x}{R} + t_{\sigma, \text{sq}} \right) - \frac{2}{R^2} \cdot \sigma'' \left( \frac{x}{R} + t_{\sigma, \text{sq}} \right) \right).$$

Mit  $f_{\text{sq}}(0) = 0$ ,  $f'_{\text{sq}}(0) = 0$  und  $f''_{\text{sq}}(0) = 2$  erhalten wir:

$$\begin{aligned} & |f_{\text{sq}}(x) - x^2| \\ &= \left| x^2 + \frac{1}{6} \cdot \frac{R^2}{\sigma''(t_{\sigma, \text{sq}})} \left( \frac{8}{R^3} \sigma''' \left( \frac{2\xi}{R} + t_{\sigma, \text{sq}} \right) - \frac{2}{R^3} \sigma''' \left( \frac{\xi}{R} + t_{\sigma, \text{sq}} \right) \right) \cdot x^3 - x^2 \right| \\ &\leq \frac{a^3}{6 \cdot |\sigma''(t_{\sigma, \text{sq}})|} \cdot \frac{1}{R} \cdot \left( 8 \cdot |\sigma''' \left( \frac{2\xi}{R} + t_{\sigma, \text{sq}} \right)| + 2 |\sigma''' \left( \frac{\xi}{R} + t_{\sigma, \text{sq}} \right)| \right) \\ &\leq \frac{10 \cdot a^3}{6 \cdot |\sigma''(t_{\sigma, \text{sq}})|} \cdot \frac{1}{R} \cdot \|\sigma'''\|_{\infty} \\ &= \frac{5 \cdot \|\sigma'''\|_{\infty} \cdot a^3}{3 \cdot |\sigma''(t_{\sigma, \text{sq}})|} \cdot \frac{1}{R}. \end{aligned}$$

Daraus folgt die Behauptung. □

Wir betrachten als Nächstes die Approximation einer Multiplikation zweier Werte auf einem Intervall durch das neuronale Netz  $f_{\text{mult}}: \mathbb{R}^2 \rightarrow \mathbb{R}$  mit Architektur  $(1, (4))$ , welches in Abbildung 1.3 veranschaulicht wird.

**Lemma 1.7.** Sei  $\sigma: \mathbb{R} \rightarrow [0, 1]$  2-zulässig. Zudem sei  $R > 0$  und  $a > 0$  beliebig. Dann gilt für das neuronale Netz

$$\begin{aligned} f_{\text{mult}}(x, y) = \frac{R^2}{4 \cdot \sigma''(t_{\sigma})} \cdot & \left( \sigma \left( \frac{2 \cdot (x+y)}{R} + t_{\sigma} \right) - 2 \cdot \sigma \left( \frac{x+y}{R} + t_{\sigma} \right) \right. \\ & \left. - \sigma \left( \frac{2 \cdot (x-y)}{R} + t_{\sigma} \right) + 2 \cdot \sigma \left( \frac{x-y}{R} + t_{\sigma} \right) \right) \end{aligned}$$

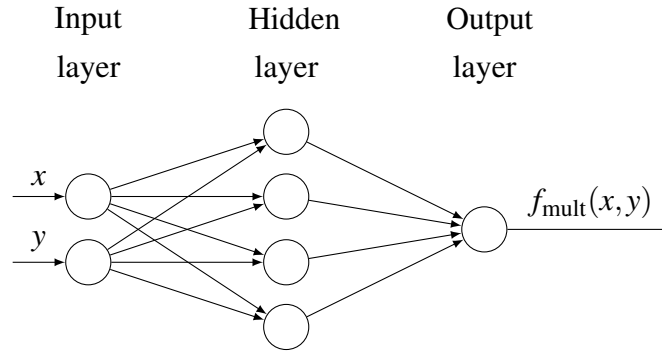


Abbildung 1.3: Neuronales Netz  $f_{\text{mult}}(x, y)$  mit Architektur  $(1, (4))$ .

für beliebige  $x, y \in [-a, a]$  folgende Ungleichung:

$$|f_{\text{mult}}(x, y) - x \cdot y| \leq \frac{20 \cdot \|\sigma'''\|_{\infty} \cdot a^3}{3 \cdot |\sigma''(t_{\sigma})|} \cdot \frac{1}{R}.$$

*Beweis.* Durch Einsetzen von  $f_{\text{sq}}$  in  $f_{\text{mult}}$  erhalten wir

$$f_{\text{mult}}(x, y) = \frac{1}{4} (f_{\text{sq}}(x + y) - f_{\text{sq}}(x - y))$$

und durch Umformungen

$$x \cdot y = \frac{1}{4} ((x + y)^2 - (x - y)^2).$$

Aus diesen beiden Gleichungen folgt durch Ausklammern von  $\frac{1}{4}$ , der Homogenität des Betrags und der Dreiecksungleichung:

$$\begin{aligned} |f_{\text{mult}}(x, y) - x \cdot y| &= \frac{1}{4} \cdot |f_{\text{sq}}(x + y) - f_{\text{sq}}(x - y) - (x + y)^2 + (x - y)^2| \\ &\leq \frac{1}{4} \cdot |f_{\text{sq}}(x + y) - (x + y)^2| + \frac{1}{4} \cdot |(x - y)^2 - f_{\text{sq}}(x - y)| \\ &\leq 2 \cdot \frac{1}{4} \cdot \frac{40 \cdot \|\sigma'''\|_{\infty} \cdot a^3}{3 \cdot |\sigma''(t_{\sigma})|} \cdot \frac{1}{R} \\ &= \frac{20 \cdot \|\sigma'''\|_{\infty} \cdot a^3}{3 \cdot |\sigma''(t_{\sigma})|} \cdot \frac{1}{R}, \end{aligned}$$

wobei wir bei der letzten Ungleichung Lemma 1.6b) mit  $x + y, x - y \in [-2a, 2a]$  verwendet haben. Daraus folgt die Behauptung.  $\square$

Wir betrachten als Nächstes die Approximation der Maximumsfunktion  $\max\{x, 0\}$  für  $x$  aus einem beschränkten Intervall durch das neuronale Netz  $f_{\text{ReLU}}: \mathbb{R} \rightarrow \mathbb{R}$  mit Architektur  $(2, (2, 4))$ , welches in Abbildung 1.4 veranschaulicht wird.



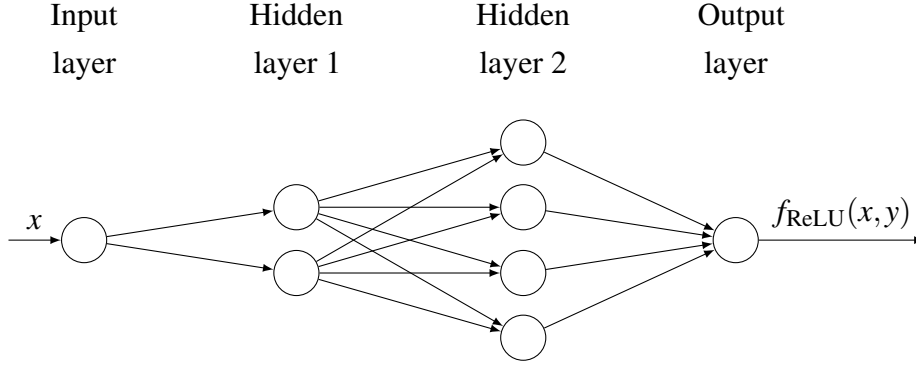


Abbildung 1.4: Neuronales Netz  $f_{\text{ReLU}}(x)$  mit Architektur  $(2, (2, 4))$ .

**Lemma 1.8.** Sei  $\sigma: \mathbb{R} \rightarrow [0, 1]$  2-zulässig und  $a \geq 1$ . Sei  $f_{\text{mult}}$  das neuronale Netz aus Lemma 1.7 und  $f_{\text{id}}$  das neuronale Netz aus Lemma 1.6. Angenommen es gelte die Ungleichung

$$R \geq \frac{\|\sigma''\|_{\infty} \cdot a}{2 \cdot |\sigma'(t_{\sigma})|}. \quad (1.6)$$

Dann erfüllt das neuronale Netz

$$f_{\text{ReLU}}(x) = f_{\text{mult}}(f_{\text{id}}(x), \sigma(R \cdot x)) \quad (1.7)$$

für alle  $x \in [-a, a]$  folgende Ungleichung:

$$|f_{\text{ReLU}}(x) - \max\{x, 0\}| \leq 56 \cdot \frac{\max\{|\sigma''|_{\infty}, \|\sigma'''\|_{\infty}, 1\}}{\min\{2 \cdot |\sigma'(t_{\sigma})|, |\sigma''(t_{\sigma})|, 1\}} \cdot a^3 \cdot \frac{1}{R}.$$

*Beweis.* Da  $\sigma$  nach Voraussetzung 2-zulässig ist, gilt für  $R \geq 0$  und  $x \in \mathbb{R} \setminus \{0\}$ :

$$|\sigma(R \cdot x) - 1| \leq \frac{1}{R \cdot x} \quad \text{für } x > 0$$

und

$$|\sigma(R \cdot x)| \leq \frac{1}{|R \cdot x|} \quad \text{für } x < 0.$$

Damit folgt aus der Homogenität des Betrags für alle  $x \neq 0$ :

$$|\sigma(R \cdot x) - \mathbb{1}_{[0, \infty)}(x)| \leq \frac{1}{|R \cdot x|} = \frac{1}{R \cdot |x|}. \quad (1.8)$$

Nach Lemma 1.6 und Lemma 1.7 gilt:

$$|f_{\text{id}}(x) - x| \leq \frac{\|\sigma''\|_{\infty} \cdot a^2}{2 \cdot |\sigma'(t_{\sigma})|} \cdot \frac{1}{R} \quad \text{für } x \in [-a, a] \quad (1.9)$$

und

$$|f_{\text{mult}}(x, y) - x \cdot y| \leq \frac{160 \cdot \|\sigma'''\|_{\infty} \cdot a^3}{3 \cdot |\sigma''(t_{\sigma})|} \cdot \frac{1}{R} \quad \text{für } x, y \in [-2a, 2a]. \quad (1.10)$$

Da nach Voraussetzung  $a \geq 1$  ist, gilt insbesondere  $[0, 1] \subseteq [-2a, 2a]$  und daher gilt insbesondere  $\sigma(Rx) \in [0, 1] \subseteq [-2a, 2a]$ . Zudem erhalten wir durch eine Nulladdition und der Dreiecksungleichung:

$$|f_{\text{id}}(x)| = |f_{\text{id}}(x) - x + x| \leq |f_{\text{id}}(x) - x| + |x|. \quad (1.11)$$

Aus Gleichung (1.9) und Voraussetzung (1.6), folgt

$$|f_{\text{id}}(x) - x| \leq \frac{\|\sigma''\|_{\infty} \cdot a^2}{2 \cdot |\sigma'(t_{\sigma})|} \cdot \frac{1}{R} \leq \frac{\|\sigma''\|_{\infty} \cdot a^2}{2 \cdot |\sigma'(t_{\sigma})|} \cdot \frac{2 \cdot |\sigma'(t_{\sigma})|}{\|\sigma''\|_{\infty} \cdot a} = a. \quad (1.12)$$

Aus  $x \in [-a, a]$  folgt schließlich mit den Ungleichungen (1.11) und (1.12):

$$|f_{\text{id}}(x)| \leq |f_{\text{id}}(x) - x| + |x| \leq 2 \cdot a.$$

Daraus folgt insbesondere  $f_{\text{id}}(x) \in [-2a, 2a]$ . Mithilfe von  $\max\{x, 0\} = x \cdot \mathbb{1}_{[0, \infty)}(x)$ , der Definition des Netzes (1.7), zweier Nulladditionen und zweimal der Dreiecksungleichung erhalten wir:

$$\begin{aligned} & |f_{\text{ReLU}}(x) - \max\{x, 0\}| \\ &= |f_{\text{mult}}(f_{\text{id}}(x), \sigma(R \cdot x)) - x \cdot \mathbb{1}_{[0, \infty)}(x)| \\ &\leq |f_{\text{mult}}(f_{\text{id}}(x), \sigma(R \cdot x)) - f_{\text{id}}(x) \cdot \sigma(R \cdot x)| \\ &\quad + |f_{\text{id}}(x) \cdot \sigma(R \cdot x) - x \cdot \sigma(R \cdot x)| + |x \cdot \sigma(R \cdot x) - x \cdot \mathbb{1}_{[0, \infty)}(x)|. \end{aligned} \quad (1.13)$$

Wenden wir nun auf den ersten Summanden in Gleichung (1.13) die Ungleichung (1.10) an, erhalten wir:

$$|f_{\text{mult}}(f_{\text{id}}(x), \sigma(R \cdot x)) - f_{\text{id}}(x) \cdot \sigma(R \cdot x)| \leq \frac{160 \cdot \|\sigma'''\|_{\infty} \cdot a^3}{3 \cdot |\sigma''(t_{\sigma})|} \cdot \frac{1}{R}. \quad (1.14)$$

Klammern wir nun im zweiten Summanden in Gleichung (1.13) den Faktor  $\sigma(R \cdot x) \in [0, 1]$  aus und wenden Ungleichung (1.9) an, erhalten wir:

$$|f_{\text{id}}(x) \cdot \sigma(R \cdot x) - x \cdot \sigma(R \cdot x)| \leq \frac{\|\sigma''\|_{\infty} \cdot a^2}{2 \cdot |\sigma'(t_{\sigma})|} \cdot \frac{1}{R} \cdot 1 \leq \frac{\|\sigma''\|_{\infty} \cdot a^3}{2 \cdot |\sigma'(t_{\sigma})|} \cdot \frac{1}{R}, \quad (1.15)$$

wobei wir verwendet haben, dass nach der Voraussetzung  $a \geq 1$  und damit  $a^2 \leq a^3$  gilt. Wenn wir nun im dritten Summanden in Gleichung (1.13) den Faktor  $|x|$  durch die Homogenität des Betrags ausklammern und Ungleichung (1.8) anwenden, erhalten wir:

$$|x \cdot \sigma(R \cdot x) - x \cdot \mathbb{1}_{[0, \infty)}(x)| \leq \frac{1}{R}. \quad (1.16)$$

Setzen wir nun Ungleichungen (1.14)- (1.16) in Ungleichung (1.13) ein, ergibt sich durch Ausklammern von  $\frac{1}{R}$  und Abschätzungen:

$$\begin{aligned}
 |f_{\text{ReLU}}(x) - \max\{x, 0\}| &\leq \left( \frac{160}{3} \cdot \frac{\|\sigma'''\|_{\infty} \cdot a^3}{|\sigma''(t_{\sigma})|} + \frac{\|\sigma''\|_{\infty} \cdot a^3}{2 \cdot |\sigma'(t_{\sigma})|} + \frac{a^3}{a^3} \right) \cdot \frac{1}{R} \\
 &\leq \left( \frac{160 \cdot \|\sigma'''\|_{\infty} \cdot a^3 + 3 \cdot \|\sigma''\|_{\infty} \cdot a^3 + 3 \cdot a^3}{3 \cdot \min\{2 \cdot |\sigma'(t_{\sigma})|, |\sigma''(t_{\sigma})|, 1\}} \right) \cdot \frac{1}{R} \\
 &\leq \frac{166}{3} \cdot \left( \frac{\max\{\|\sigma'''\|_{\infty}, \|\sigma''\|_{\infty}, 1\}}{\min\{2 \cdot |\sigma'(t_{\sigma})|, |\sigma''(t_{\sigma})|, 1\}} \right) \cdot a^3 \cdot \frac{1}{R}.
 \end{aligned}$$

Da  $\frac{166}{3} \leq 56$  gilt, folgt schließlich die Behauptung.  $\square$

Wir betrachten als Nächstes die Approximation des Positivteils einer Funktion auf einem Intervall durch das neuronale Netz  $f_{\text{hat},y}: \mathbb{R} \rightarrow \mathbb{R}$  mit Architektur  $(2, (6, 12))$ , welches in Abbildung 1.5 veranschaulicht wird.

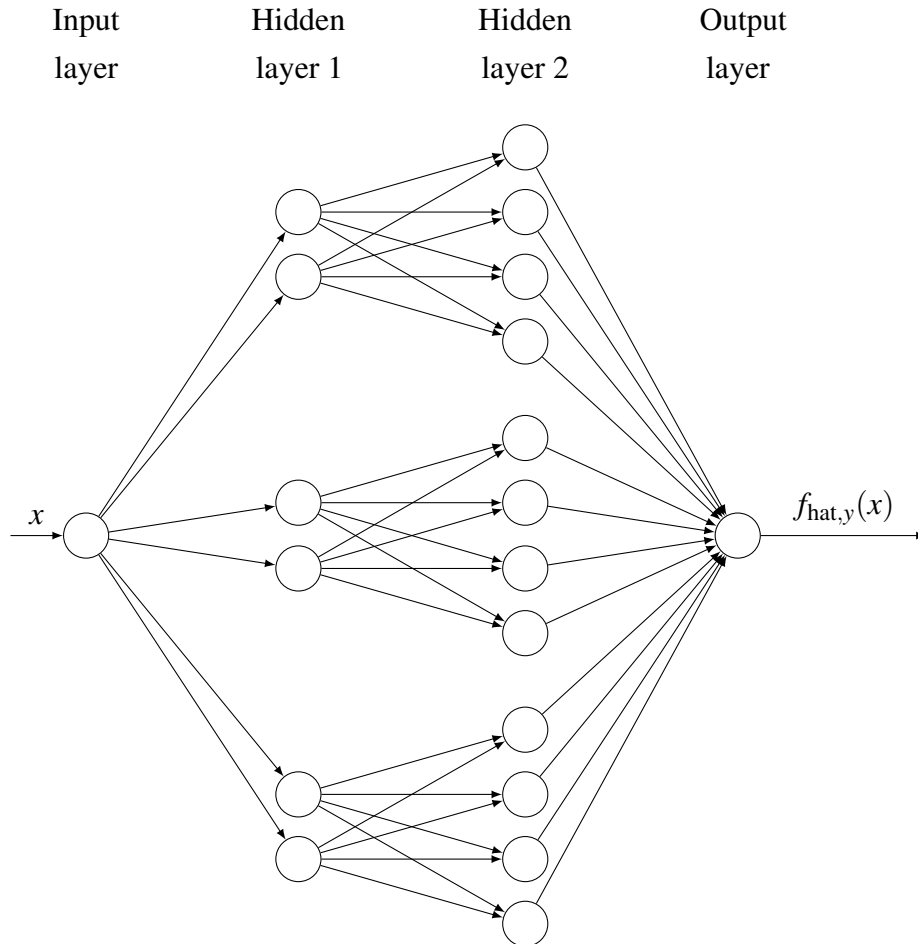


Abbildung 1.5: Neuronales Netz  $f_{\text{hat},y}(x)$  mit Architektur  $(2, (6, 12))$ .

**Lemma 1.9.** Sei  $M \in \mathbb{N}$  und sei  $\sigma: \mathbb{R} \rightarrow [0, 1]$  2-zulässig. Sei  $a > 0$  und

$$R \geq \frac{\|\sigma''\|_\infty \cdot (M+1)}{2 \cdot |\sigma'(t_\sigma)|},$$

sei  $y \in [-a, a]$  und  $f_{\text{ReLU}}$  das neuronale Netz aus Lemma 1.8. Dann erfüllt das neuronale Netz

$$\begin{aligned} f_{\text{hat},y}(x) = & f_{\text{ReLU}}\left(\frac{M}{2a} \cdot (x-y) + 1\right) - 2 \cdot f_{\text{ReLU}}\left(\frac{M}{2a} \cdot (x-y)\right) \\ & + f_{\text{ReLU}}\left(\frac{M}{2a} \cdot (x-y) - 1\right) \end{aligned}$$

für alle  $x \in [-a, a]$  mit  $z_+ = \max\{0, z\}$  ( $z \in \mathbb{R}$ ):

$$\left| f_{\text{hat},y}(x) - \left(1 - \frac{M}{2a} \cdot |x-y|\right)_+ \right| \leq 1792 \cdot \frac{\max\{\|\sigma''\|_\infty, \|\sigma'''\|_\infty, 1\}}{\min\{2 \cdot |\sigma'(t_\sigma)|, |\sigma''(t_\sigma)|, 1\}} \cdot M^3 \cdot \frac{1}{R}.$$

*Beweis.* Für  $x \in \mathbb{R}$  gilt die Gleichung:

$$\left(1 - \frac{M}{2a} \cdot |x|\right)_+ = \max\left\{\frac{M}{2a} \cdot x + 1, 0\right\} - 2 \cdot \max\left\{\frac{M}{2a} \cdot x, 0\right\} + \max\left\{\frac{M}{2a} \cdot x - 1, 0\right\}, \quad (1.17)$$

die wir im zweiten Teil dieses Beweises zeigen werden. Damit beweisen wir das Resultat mit Hilfe von Lemma 1.8, denn mit der Definition von  $f_{\text{hat},y}(x)$  und der Dreiecksungleichung folgt:

$$\begin{aligned} \left| f_{\text{hat},y}(x) - \left(1 - \frac{M}{2a} \cdot |x-y|\right)_+ \right| & \leq \left| f_{\text{ReLU}}\left(\frac{M}{2a} \cdot (x-y) + 1\right) - \max\left\{\frac{M}{2a} \cdot (x-y) + 1, 0\right\} \right| \\ & \quad + 2 \cdot \left| f_{\text{ReLU}}\left(\frac{M}{2a} \cdot (x-y)\right) - \max\left\{\frac{M}{2a} \cdot (x-y), 0\right\} \right| \\ & \quad + \left| f_{\text{ReLU}}\left(\frac{M}{2a} \cdot (x-y) - 1\right) - \max\left\{\frac{M}{2a} \cdot (x-y) - 1, 0\right\} \right| \\ & \leq 1792 \cdot \frac{\max\{\|\sigma''\|_\infty, \|\sigma'''\|_\infty, 1\}}{\min\{2 \cdot |\sigma'(t_\sigma)|, |\sigma''(t_\sigma)|, 1\}} \cdot M^3 \cdot \frac{1}{R}, \end{aligned}$$

wobei die letzte Ungleichung daraus folgt, dass wir auf jeden Summanden Lemma 1.8 mit

$$1 \leq a = M+1$$

angewendet haben, da aus  $x, y \in [-a, a]$  folgt, dass  $\frac{M}{2a} \cdot (x-y) \in [-M, M]$  gilt. Schließlich haben wir

$$(M+1)^3 \leq (2M)^3 = 8M^3$$

verwendet, da  $M \geq 1$  ist. (□)

Um Gleichung (1.17) zu zeigen unterscheiden wir vier Fälle.

Fall 1 ( $x < 0$ ) In diesem Fall hat die linke Seite von Gleichung (1.17) nach der Definition des Betrags die Gestalt

$$\max\{1 + \frac{M}{2a} \cdot x, 0\}$$

und die rechte Seite von Gleichung (1.17) die Form

$$\max\{\frac{M}{2a} \cdot x + 1, 0\} - 2 \cdot 0 + 0,$$

da  $x < 0$  und damit die letzten zwei Summanden Null sind. Es erfordert hier eine weitere Fallunterscheidung.

Fall 1.1 ( $0 > x \geq -\frac{2a}{M}$ ) In diesem Fall gilt für die linke und rechte Seite von Gleichung (1.17):

$$\max\{1 + \frac{M}{2a} \cdot x, 0\} = 1 + \frac{M}{2a} \cdot x.$$

Fall 1.2 ( $x < -\frac{2a}{M}$ ) In diesem Fall sind beide Seiten gleich Null, da  $1 + \frac{M}{2a} \cdot x \leq 0$  ist. ( $\square$ )

Fall 2 ( $x \geq 0$ ) In diesem Fall hat die linke Seite von Gleichung (1.17) nach der Definition des Betrags die Gestalt

$$\max\{1 - \frac{M}{2a} \cdot x, 0\}$$

und die rechte Seite von Gleichung (1.17) die Form

$$\max\{\frac{M}{2a} \cdot x + 1, 0\} - 2 \cdot \max\{\frac{M}{2a} \cdot x, 0\} + \max\{\frac{M}{2a} \cdot x - 1, 0\}$$

und erfordert daher eine weitere Fallunterscheidung.

Fall 2.1 ( $0 \leq x < \frac{2a}{M}$ ) In diesem Fall hat die linke Seite von Gleichung (1.17) die Gestalt

$$1 - \frac{M}{2a} \cdot x$$

und die rechte Seite von Gleichung (1.17) die Form

$$\frac{M}{2a} \cdot x + 1 - 2 \cdot \frac{M}{2a} \cdot x + 0 = 1 - \frac{M}{2a} \cdot x$$

und stimmt daher mit der linken Seite überein.

Fall 2.2 ( $x \geq \frac{2a}{M}$ ) In diesem Fall ist die linke Seite von Gleichung (1.17) gleich 0, da  $1 - \frac{M}{2a} \cdot x < 0$  ist und die rechte Seite besitzt die Form

$$\frac{M}{2a} \cdot x + 1 - 2 \cdot \frac{M}{2a} \cdot x + \frac{M}{2a} \cdot x - 1 = 0.$$

( $\square$ )

Durch diese Fallunterscheidung wurde die Gleichung (1.17) bewiesen und damit ist der Beweis vollständig.  $\square$

Im nächsten Kapitel werden wir die hier eingeführten neuronalen Netze als Bausteine verwenden, um unseren Neuronale-Netze-Regressionsschätzer zu konstruieren.

## Kapitel 2

# Konstruktion eines Neuronale-Netze-Schätzers

In dieser Arbeit behandeln wir Neuronale-Netze-Schätzer im Kontext der *nichtparametrischen Regression* mit *zufälligem Design*. In der Regressionsanalyse betrachtet man einen  $\mathbb{R}^d \times \mathbb{R}$ -wertigen Zufallsvektor  $(X, Y)$ . Man ist daran interessiert, wie der Wert der *Reaktionsvariable*  $Y$  vom Wert des *Beobachtungsvektors*  $X$  abhängt. Dies bedeutet, dass man eine (messbare) Funktion  $f: \mathbb{R}^d \rightarrow \mathbb{R}$  sucht, sodass  $f(X)$  eine „gute Approximation von  $Y$  ist“. Das wiederum bedeutet, dass  $f(X)$  nah an  $Y$  sein sollte, was in gewisser Weise gleichwertig ist, den Ausdruck  $|f(X) - Y|$  „klein“ zu machen. Da aber  $X$  und  $Y$  Zufallsvektoren sind und damit  $|f(X) - Y|$  auch zufällig ist, ist nicht klar was unter „ $|f(X) - Y|$  klein“ zu verstehen ist. Wir können dieses Problem lösen, indem wir das sogenannte  $L_2$ -Risiko

$$\mathbb{E}[|f(X) - Y|^2],$$

eingeführen und verlangen, dass es so klein wie möglich sein muss. Bei der nichtparametrischen Regression ist die Bauart der zu schätzenden Funktion komplett unbekannt, was den Vorteil besitzt, dass weniger Annahmen getroffen werden müssen, man aber dadurch noch mehr Daten benötigt, um eine Funktion zu schätzen. Das Problem der Regressions-schätzung bei zufälligem Design lässt sich wie folgt erläutern: In Anwendungsfällen ist üblicherweise die Verteilung von  $(X, Y)$  unbekannt, daher kann die *Regressionsfunktion*  $m(x) = \mathbb{E}[Y \mid X = x]$  nicht berechnet werden. Oft ist es aber möglich, Werte von  $(X, Y)$  zu beobachten. Für das  $L_2$ -Risiko einer beliebigen messbaren Funktion  $f: \mathbb{R}^d \rightarrow \mathbb{R}$  gilt:

$$\mathbb{E}[|f(X) - Y|^2] = \mathbb{E}[|m(X) - Y|^2] + \int |f(x) - m(x)|^2 \mathbb{P}_X(dx),$$

d.h. der mittlere quadratische Vorhersagefehler einer Funktion ist darstellbar als Summe des  $L_2$ -Risikos der Regressionsfunktion (unvermeidbarer Fehler) und des  $L_2$ -Fehlers, der

aufgrund der Verwendung von  $f$  an Stelle von  $m$  bei der Vorhersage bzw. Approximation des Wertes von  $Y$  entsteht. Im Hinblick auf die Minimierung des  $L_2$ -Risikos sollte dabei der  $L_2$ -Fehler der Schätzfunktion möglichst klein sein. Formal führt das daher auf folgende Problemstellung:

**Problem 2.1.**  $(X, Y), (X_1, Y_1), (X_2, Y_2), \dots$  seien u.i.v.  $\mathbb{R}^d \times \mathbb{R}$  wertige Zufallsvariablen mit  $\mathbb{E}[Y^2] < \infty$  und  $m: \mathbb{R}^d \rightarrow \mathbb{R}$  definiert durch  $m(x) = \mathbb{E}[Y \mid X = x]$  sei die zugehörige Regressionsfunktion. Gegeben sei die Datenmenge

$$\mathcal{D}_n = \{(X_1, Y_1), \dots, (X_n, Y_n)\}. \quad (2.1)$$

Gesucht ist eine Schätzung

$$m_n(\cdot) = m_n(\cdot, \mathcal{D}_n): \mathbb{R}^d \rightarrow \mathbb{R}$$

von  $m$ , für die der  $L_2$ -Fehler

$$\int |m_n(x) - m(x)|^2 \mathbb{P}_X(dx)$$

möglichst „klein“ ist [LG02, Kapitel 1.1 und Kapitel 1.2].

In diesem Kapitel werden wir mithilfe der gegebenen Datenmenge  $\mathcal{D}_n$  unseren Regressionsschätzer  $\tilde{m}_n$  konstruieren. In Kapitel 1 haben wir bereits in Definition 1.1 vorgestellt, was wir unter einem mehrschichtigen feedforward neuronalen Netz mit Architektur  $(L, \mathbf{k})$  und Aktivierungsfunktion  $\sigma$  verstehen.

Für die Konstruktion unseres Neuronale-Netze-Regressionsschätzers wählen wir den logistischen Squasher (1.2) als Aktivierungsfunktion  $\sigma$ , verwenden die gegebene Datenmenge  $\mathcal{D}_n$  und wählen die Gewichte des neuronalen Netzes so, dass die resultierende Funktion aus Definition 1.1 eine gute Schätzung für die Regressionsfunktion  $m$  ist. Dafür wählen wir die Gewichte bis auf die in der Ausgabeschicht fest und schätzen die Gewichte in der Ausgabeschicht, indem wir mit unserer Datenmenge (2.1) ein Kleinste-Quadrate Problem lösen.

Es ist bekannt (vgl. [LD96, Theorem 7.2 und Problem 7.2] und [LD80, Section 3]), dass man Glattheitsvoraussetzungen an die Regressionsfunktion stellen muss, um nichttriviale Konvergenzresultate für nichtparametrische Regressionsschätzer herzuleiten. Dafür verwenden wir die folgende Definition.

**Definition 2.2** ( $(p, C)$ -Glattheit). Sei  $p = q + s$  mit  $q \in \mathbb{N}_0$  und  $s \in (0, 1]$  (also  $p \in (0, \infty)$ ) und sei  $C > 0$ . Eine Funktion  $f: \mathbb{R}^d \rightarrow \mathbb{R}$  heißt  $(p, C)$ -glatt, falls für alle  $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}_0^d$  mit  $\sum_{j=1}^d \alpha_j = q$  die partielle Ableitung

$$\frac{\partial^q f}{\partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}}$$

existiert und falls für alle  $x, z \in \mathbb{R}^d$  die Abschätzung

$$\left| \frac{\partial^q f}{\partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}}(x) - \frac{\partial^q f}{\partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}}(z) \right| \leq C \cdot \|x - z\|^s,$$

gilt, wobei  $\|\cdot\|$  die euklidische Norm in  $\mathbb{R}^d$  bezeichnet.

## 2.1 Definition der Netzwerkarchitektur

In diesem Abschnitt stellen wir die *Netzwerkarchitektur* unsere Neuronale-Netze-Regressionsschätzers vor. Dafür legen wir die Architektur  $(L, \mathbf{k})$  fest und gehen auf die konkrete Konstruktion unseres Schätzers ein.

Zunächst fixieren wir die Multiindexnotation, die wir der Übersichtlichkeit halber im weiteren Verlauf dieser Arbeit verwenden werden. Sei  $a > 0$  fest und  $M \in \mathbb{N}$ .

Sei  $[M]^d := \{0, 1, \dots, M\}^d$ . Für  $(\mathbf{i}^{(1)}, \dots, \mathbf{i}^{(d)}) = \mathbf{i} \in [M]^d$  und  $x \in \mathbb{R}^d$  definieren wir

$$|\mathbf{i}|_1 := \sum_{k=1}^d \mathbf{i}^{(k)}, \quad \mathbf{i}! := \mathbf{i}^{(1)}! \dots \mathbf{i}^{(d)}! \quad \text{und} \quad x^{\mathbf{i}} := x_1^{\mathbf{i}^{(1)}} \dots x_d^{\mathbf{i}^{(d)}}.$$

Für  $f: \mathbb{R}^d \rightarrow \mathbb{R}$  ausreichend oft differenzierbar definieren wir

$$\partial^{\mathbf{i}} f(x) := \frac{\partial^{|\mathbf{i}|_1} f}{\partial \mathbf{i}^{(1)} x_1 \dots \partial \mathbf{i}^{(d)} x_d}(x).$$

Das nächste Lemma ist ein Resultat aus der Kombinatorik, welches wir in Kapitel 2.2 benötigen werden.

**Lemma 2.3.** *Sei  $d, N \in \mathbb{N}$  und  $k \in \mathbb{N}_0$  mit  $k < N$ . Dann gilt:*

$$\left| \left\{ \mathbf{j} \in [N]^d : |\mathbf{j}|_1 = k \right\} \right| = \binom{d+k-1}{k}.$$

*Beweis.* Diese Aussage folgt aus einer Analogie zu einem Urnenexperiment. Wir betrachten eine Urne mit  $d$  Kugeln, die wir mit  $1, \dots, d$  beschriften. Wir ziehen  $k$ -mal aus dieser Urne mit Zurücklegen und ohne Beachtung der Reihenfolge und konstruieren so einen Vektor  $\mathbf{j} = (j_1, \dots, j_d)$  aus der Menge mit  $|\mathbf{j}|_1 = k$ . Der Koeffizient  $j_i$  mit  $i = 1, \dots, d$  gibt an wie oft die Kugel mit der Nummer  $i$  gezogen wurde. Damit stimmt die Kardinalität der Menge auf der linken Seite mit der Anzahl aller Möglichkeiten überein die man erhält, wenn man  $k$ -mal aus dieser Urne mit Zurücklegen und ohne Beachtung der Reihenfolge zieht. □



Wir betrachten im Folgenden ein  $d$ -dimensionales äquidistantes Gitter im Würfel  $[-a, a]^d$  mit Schrittweite  $\frac{2a}{M}$ . Dann ordnen wir jedem Multiindex  $\mathbf{i} \in [M]^d$  einen Gitterpunkt

$$x_{\mathbf{i}} = \left( -a + \mathbf{i}^{(1)} \cdot \frac{2a}{M}, \dots, -a + \mathbf{i}^{(d)} \cdot \frac{2a}{M} \right) = -\mathbf{a} + \frac{2a}{M} \cdot \mathbf{i}, \quad (2.2)$$

zu, mit  $\mathbf{a} = (a, a, \dots, a) \in \mathbb{R}^d$ .

Hiermit lässt sich das zu  $m$  gehörige Taylorpolynom der Ordnung  $q \in \mathbb{N}_0$  mit Entwicklungspunkt  $x_{\mathbf{i}}$  schreiben als

$$p_{\mathbf{i}}^m(x) = \sum_{\substack{\mathbf{j} \in [q]^d \\ |\mathbf{j}|_1 \leq q}} \partial^{\mathbf{j}} m(x_{\mathbf{i}}) \cdot \frac{(x - x_{\mathbf{i}})^{\mathbf{j}}}{\mathbf{j}!}.$$

Zudem betrachten wir eine Funktion

$$P_m(x) = \sum_{\mathbf{i} \in [M]^d} p_{\mathbf{i}}^m(x) \prod_{j=1}^d \left( 1 - \frac{M}{2a} \cdot |x^{(j)} - x_{\mathbf{i}}^{(j)}| \right)_+, \quad (2.3)$$

für die wir im weiteren Verlauf dieser Arbeit zeigen werden, dass diese die Regressionsfunktion  $m$  approximiert.

Wir zeigen mithilfe des folgenden Lemmas, dass  $P_m(x)$  eine lokale Spline-Interpolation von Taylorpolynomen von  $m$  ist.

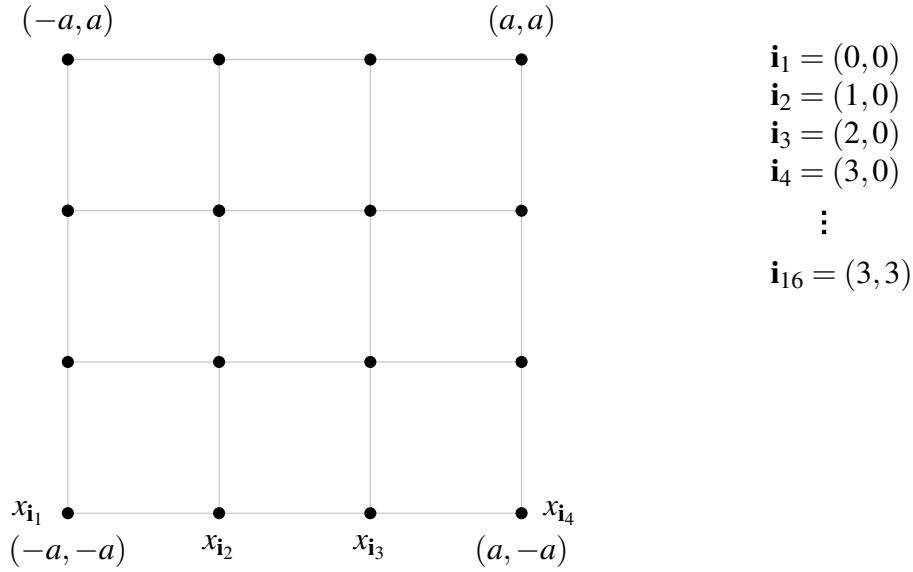
**Lemma 2.4.** *Sei  $a > 0$  und  $M \in \mathbb{N}$ . Dann sind für  $\mathbf{i} \in [M]^d$  die Funktionen*

$$B_{\mathbf{i}}(x) = \prod_{j=1}^d \left( 1 - \frac{M}{2a} \cdot |x^{(j)} - x_{\mathbf{i}}^{(j)}| \right)_+ \quad \text{für } x \in [-a, a]^d,$$

mit  $x_{\mathbf{i}}$  als Gitterpunkte aus Gleichung (2.2), B-Splines auf  $[-a, a]^d$ , für die die folgenden drei Bedingungen gelten:

- (i) Zerlegung der Eins:  $\sum_{\mathbf{i} \in [M]^d} B_{\mathbf{i}}(x) = 1$  für  $x \in [-a, a]^d$ .
- (ii) Nichtnegativität:  $B_{\mathbf{i}}(x) \geq 0$  für alle  $\mathbf{i} \in [M]^d$ .
- (iii) Lokaler Träger: Für  $\mathbf{i} \in [M]^d$  ist  $B_{\mathbf{i}}(x) > 0$  falls  $|x^{(j)} - x_{\mathbf{i}}^{(j)}| < \frac{2a}{M}$  für alle  $j \in \{1, \dots, d\}$  gilt und andernfalls  $B_{\mathbf{i}}(x) = 0$ .

*Beweis.* Als erstes möchten wir für  $d = 2$  und  $M = 3$  eine Skizze angeben, um die Idee des Beweises zu veranschaulichen. Es ist ein Gitter mit  $(M+1)^d$  Gitterpunkten, die den  $x_{\mathbf{i}_k}$  entsprechen. Der Abstand zwischen zwei Gitterpunkten beträgt  $\frac{2a}{M}$ . Man betrachtet immer den Abstand zu den nächsten  $2^d$  Gitterpunkten, da  $(1 - \frac{M}{2a} \cdot |x^{(j)} - x_{\mathbf{i}}^{(j)}|)_+ = 0$  immer dann gilt, wenn der Abstand zwischen  $x^{(j)}$  und  $x_{\mathbf{i}}^{(j)}$  größer als  $\frac{2a}{M}$  ist.

Abbildung 2.1: Beispielhafte Darstellung der  $x_{i_k}$  für  $d = 2$  und  $M = 3$ .

(i) Im Folgenden wollen wir

$$\sum_{\mathbf{i} \in [M]^d} B_{\mathbf{i}}(x) = 1 \text{ für } x \in [-a, a]^d \quad (2.4)$$

per Induktion über  $d$  zeigen.

*Induktionsanfang (IA):* Für  $d = 1$  kann  $x$  nur zwischen zwei Gitterpunkten  $x_{i_1} \neq x_{i_2}$  liegen. Sei ohne Beschränkung der Allgemeinheit  $x_{i_1} \leq x \leq x_{i_2}$ , dann gilt mit der gleichen Begründung wie im einleitenden Beispiel:

$$\begin{aligned} \sum_{\mathbf{i} \in [M]^d} \left(1 - \frac{M}{2a} \cdot |x - x_{\mathbf{i}}|\right)_+ &= \left(1 - \frac{M}{2a} \cdot |x - x_{i_1}|\right)_+ + \left(1 - \frac{M}{2a} \cdot |x - x_{i_2}|\right)_+ \\ &= 1 + 1 - \frac{M}{2a} \cdot (x - x_{i_1} + x_{i_2} - x) \\ &= 1 + 1 - \frac{M}{2a} \cdot \frac{2a}{M} \\ &= 1, \end{aligned}$$

wobei wir unter anderem verwendet haben, dass beide Summanden unabhängig von dem Positivteil nichtnegativ sind, da der Abstand von  $x$  zu den beiden Gitterpunkten  $x_{i_1}$  und  $x_{i_2}$  kleiner gleich  $\frac{2a}{M}$  ist. Zudem haben wir verwendet, dass  $x_{i_2} - x_{i_1} = \frac{2a}{M}$  gilt, da beides Gitterpunkte sind.

*Induktionshypothese (IH):* Aussage (2.4) gelte für ein beliebiges aber festes  $d \in \mathbb{N}$ .

*Induktionsschritt (IS):* Wir nehmen ohne Beschränkung der Allgemeinheit an, dass  $x_{(0,\dots,0)} \leq x \leq x_{(1,\dots,1)}$  komponentenweise gilt. Das heißt also, dass  $x \in [-a, -a + \frac{2a}{M}]^{d+1}$

gilt. Im Folgenden zeigen wir

$$\sum_{\mathbf{i} \in [M]^{(d+1)}} B_{\mathbf{i}}(x) = \sum_{\mathbf{i} \in [M]^{(d+1)}} \prod_{j=1}^{d+1} \left( 1 - \frac{M}{2a} \cdot |x^{(j)} - x_{\mathbf{i}}^{(j)}| \right)_+ = 1.$$

Ein Summand ist Null, wenn ein  $j \in \{1, \dots, d+1\}$  existiert mit  $|x^{(j)} - x_{\mathbf{i}}^{(j)}| \geq \frac{2a}{M}$ . Zudem haben wir ohne Beschränkung der Allgemeinheit angenommen, dass  $x \in [-a, -a + \frac{2a}{M}]^{d+1}$  gilt. Damit haben wir also nur noch  $2^{d+1}$  Summanden, was der Anzahl der Gitterpunkte die am nächsten bei  $x$  liegen entspricht. Zudem wissen wir, dass alle Gitterpunkte, die in der  $(d+1)$ -ten Komponente den selben Wert haben, in dieser Dimension gleich weit von  $x^{(d+1)}$  entfernt sind. Daraus ergibt sich, dass der Faktor  $(1 - \frac{M}{2a} \cdot |x^{(d+1)} - x_{(0, \dots, 0)}^{(d+1)}|)$  bzw.  $(1 - \frac{M}{2a} \cdot |x^{(d+1)} - x_{(1, \dots, 1)}^{(d+1)}|)$  in jedem Summanden vorkommt, da

$$\left( 1 - \frac{M}{2a} \cdot |x^{(d+1)} - x_{\mathbf{i}}^{(d+1)}| \right) = \begin{cases} (1 - \frac{M}{2a} \cdot |x^{(d+1)} - x_{(0, \dots, 0)}^{(d+1)}|) & \mathbf{i} \in \{0, 1\}^d \times \{0\} \\ (1 - \frac{M}{2a} \cdot |x^{(d+1)} - x_{(1, \dots, 1)}^{(d+1)}|) & \mathbf{i} \in \{0, 1\}^d \times \{1\} \end{cases}$$

gilt. Daraus ergibt sich:

$$\begin{aligned} & \sum_{\mathbf{i} \in [M]^{(d+1)}} \prod_{j=1}^{d+1} \left( 1 - \frac{M}{2a} \cdot |x^{(j)} - x_{\mathbf{i}}^{(j)}| \right)_+ \\ &= \sum_{\mathbf{i} \in \{0, 1\}^{d+1}} \prod_{j=1}^{d+1} \left( 1 - \frac{M}{2a} \cdot |x^{(j)} - x_{\mathbf{i}}^{(j)}| \right) \\ &= \left( \sum_{\mathbf{i} \in \{0, 1\}^d \times \{0\}} \prod_{j=1}^d \left( 1 - \frac{M}{2a} \cdot |x^{(j)} - x_{\mathbf{i}}^{(j)}| \right) \right) \cdot \left( 1 - \frac{M}{2a} \cdot |x^{(d+1)} - x_{(0, \dots, 0)}^{(d+1)}| \right) \\ &\quad + \left( \sum_{\mathbf{i} \in \{0, 1\}^d \times \{1\}} \prod_{j=1}^d \left( 1 - \frac{M}{2a} \cdot |x^{(j)} - x_{\mathbf{i}}^{(j)}| \right) \right) \cdot \left( 1 - \frac{M}{2a} \cdot |x^{(d+1)} - x_{(1, \dots, 1)}^{(d+1)}| \right) \\ &\stackrel{\text{(IV)}}{=} 1 \cdot \left( 1 - \frac{M}{2a} \cdot |x^{(d+1)} - x_{(0, \dots, 0)}^{(d+1)}| \right) + 1 \cdot \left( 1 - \frac{M}{2a} \cdot |x^{(d+1)} - x_{(1, \dots, 1)}^{(d+1)}| \right) \\ &= 1 + 1 - \frac{M}{2a} \cdot |x^{(d+1)} - x_{(0, \dots, 0)}^{(d+1)} + x_{(1, \dots, 1)}^{(d+1)} - x^{(d+1)}| \\ &= 1 + 1 - 1 \\ &= 1, \end{aligned}$$

wobei wir bei der vorletzten Gleichung angewendet haben, dass  $x_{(1, \dots, 1)}^{(d+1)} - x_{(0, \dots, 0)}^{(d+1)} = \frac{2a}{M}$  ist, da beides Gitterpunkte sind.  $(\square)$

(ii) Es folgt  $\prod_{j=1}^d (1 - \frac{M}{2a} \cdot |x^{(j)} - x_{\mathbf{i}}^{(j)}|)_+ \geq 0$  für alle  $\mathbf{i} \in [M]^d$ , da

$$z_+ = \max\{z, 0\} \geq 0 \text{ für } z \in \mathbb{R}$$

gilt.

(iii) Es handelt sich hierbei um einen lokalen Träger, da nach der Konstruktion von  $B_{\mathbf{i}}(x)$  der Funktionswert genau dann Null ist, wenn ein  $j \in \{1, \dots, d\}$  existiert, sodass  $|x^{(j)} - x_{\mathbf{i}}^{(j)}| \geq \frac{2a}{M}$  gilt. Andernfalls erhalten wir mit Bedingung (ii), dass  $B_{\mathbf{i}}(x) > 0$  ist.  $\square$

Mit der Definition eines B-Splines aus Lemma 2.4 erhalten wir nun:

$$\begin{aligned} P_m(x) &= \sum_{\mathbf{i} \in [M]^d} p_{\mathbf{i}}^m(x) \prod_{j=1}^d \left( 1 - \frac{M}{2a} \cdot |x^{(j)} - x_{\mathbf{i}}^{(j)}| \right)_+ \\ &= \sum_{\mathbf{i} \in [M]^d} p_{\mathbf{i}}^m(x) \cdot B_{\mathbf{i}}(x). \end{aligned} \quad (2.5)$$

Damit wissen wir, dass  $P_m(x)$  eine Spline-Interpolation von Taylorpolynomen von  $m$  ist. Die Wahl der Architektur  $(L, \mathbf{k})$  unseres Neuronale-Netze-Regressionsschätzers und der Werte aller Gewichte bis auf die der Ausgabeschicht ist durch folgendes Approximationsresultat motiviert.

**Lemma 2.5.** *Sei  $M \in \mathbb{N}$ ,  $a > 0$  und  $f$  eine  $(p, C)$ -glatte Funktion, wobei  $p = q + s$  mit  $q \in \mathbb{N}_0, s \in (0, 1]$  und  $C > 0$  sind. Sei zudem  $P_f(x)$  analog zu Gleichung (2.5) eine lokale Spline-Interpolation von Taylorpolynomen von  $f$ . Dann gilt:*

$$\sup_{x \in [-a, a]^d} |f(x) - P_f(x)| \leq c \cdot \left( \frac{a}{M} \right)^p,$$

mit einer Konstante  $c$ , die von  $p, d, s$  und  $C$  abhängt.

*Beweis.* Nach Lemma 1.5 i) über die Lagrange Form des Restglieds existiert ein  $\xi$  auf der Verbindungsstrecke zwischen  $x_{\mathbf{i}}$  und  $x$ , so, dass

$$\begin{aligned} f(x) &= T_{q-1}f(x; x_{\mathbf{i}}) + R_q(x; x_{\mathbf{i}}) \\ &= \sum_{\substack{\mathbf{j} \in [q]^d \\ |\mathbf{j}|_1 \leq q-1}} \partial^{\mathbf{j}} f(x_{\mathbf{i}}) \cdot \frac{(x - x_{\mathbf{i}})^{\mathbf{j}}}{\mathbf{j}!} + \sum_{\substack{\mathbf{j} \in [q]^d \\ |\mathbf{j}|_1 = q}} \partial^{\mathbf{j}} f(\xi) \cdot \frac{(x - x_{\mathbf{i}})^{\mathbf{j}}}{\mathbf{j}!}. \end{aligned} \quad (2.6)$$

Nach der B-Spline Eigenschaft aus Gleichung (2.4) erhalten wir

$$f(x) = \sum_{\mathbf{i} \in [M]^d} f(x_{\mathbf{i}}) \prod_{j=1}^d \left( 1 - \frac{M}{2a} \cdot |x^{(j)} - x_{\mathbf{i}}^{(j)}| \right)_+.$$

Zudem wissen wir, dass man ohne Beschränkung der Allgemeinheit immer den Abstand zu den nächsten  $2^d$  Gitterpunkten betrachtet, da  $(1 - \frac{M}{2a} \cdot |x^{(j)} - x_{\mathbf{i}}^{(j)}|)_+ = 0$  immer dann gilt, wenn der Abstand zwischen  $x^{(j)}$  und  $x_{\mathbf{i}}^{(j)}$  größer als  $\frac{2a}{M}$  ist. Daher ergibt sich:

$$\sum_{\substack{\mathbf{j} \in [q]^d \\ |\mathbf{j}|_1 = q}} \partial^{\mathbf{j}} f(\xi) \frac{(x - x_{\mathbf{i}})^{\mathbf{j}}}{\mathbf{j}!} \leq \sum_{\substack{\mathbf{j} \in [q]^d \\ |\mathbf{j}|_1 = q}} \partial^{\mathbf{j}} f(\xi) \frac{1}{\mathbf{j}!} \cdot \left( \frac{2a}{M} \right)^q. \quad (2.7)$$

Mithilfe der Dreiecksungleichung und der Konstruktion von  $P_f(x)$  erhalten wir:

$$|f(x) - P_f(x)| \leq \sum_{\mathbf{i} \in [M]^d} \left| f(x) - \sum_{\substack{\mathbf{j} \in [q]^d \\ |\mathbf{j}|_1 \leq q}} \partial^{\mathbf{j}} f(x_{\mathbf{i}}) \cdot \frac{(x - x_{\mathbf{i}})^{\mathbf{j}}}{\mathbf{j}!} \right| \prod_{j=1}^d \left( 1 - \frac{M}{2a} \cdot |x^{(j)} - x_{\mathbf{i}}^{(j)}| \right)_+ . \quad (2.8)$$

Nach Gleichung (2.6) erhalten wir:

$$\begin{aligned} & \left| f(x) - \sum_{\substack{\mathbf{j} \in [q]^d \\ |\mathbf{j}|_1 \leq q}} \partial^{\mathbf{j}} f(x_{\mathbf{i}}) \cdot \frac{(x - x_{\mathbf{i}})^{\mathbf{j}}}{\mathbf{j}!} \right| \\ &= \left| \sum_{\substack{\mathbf{j} \in [q]^d \\ |\mathbf{j}|_1 \leq q-1}} \partial^{\mathbf{j}} f(x_{\mathbf{i}}) \cdot \frac{(x - x_{\mathbf{i}})^{\mathbf{j}}}{\mathbf{j}!} + \sum_{\substack{\mathbf{j} \in [q]^d \\ |\mathbf{j}|_1 = q}} \partial^{\mathbf{j}} f(\xi) \frac{(x - x_{\mathbf{i}})^{\mathbf{j}}}{\mathbf{j}!} \right. \\ & \quad \left. - \sum_{\substack{\mathbf{j} \in [q]^d \\ |\mathbf{j}|_1 \leq q}} \partial^{\mathbf{j}} f(x_{\mathbf{i}}) \cdot \frac{(x - x_{\mathbf{i}})^{\mathbf{j}}}{\mathbf{j}!} \right| \\ &= \left| \sum_{\substack{\mathbf{j} \in [q]^d \\ |\mathbf{j}|_1 = q}} \partial^{\mathbf{j}} f(\xi) \frac{(x - x_{\mathbf{i}})^{\mathbf{j}}}{\mathbf{j}!} - \sum_{\substack{\mathbf{j} \in [q]^d \\ |\mathbf{j}|_1 = q}} \partial^{\mathbf{j}} f(x_{\mathbf{i}}) \frac{(x - x_{\mathbf{i}})^{\mathbf{j}}}{\mathbf{j}!} \right|. \end{aligned} \quad (2.9)$$

Aus der  $(p, C)$ -Glattheit von  $f$  mit  $\xi, x_{\mathbf{i}} \in \mathbb{R}^d$  folgt durch (2.7) und (2.9):

$$\begin{aligned} & \left| \sum_{\substack{\mathbf{j} \in [q]^d \\ |\mathbf{j}|_1 = q}} \partial^{\mathbf{j}} f(\xi) \frac{(x - x_{\mathbf{i}})^{\mathbf{j}}}{\mathbf{j}!} - \sum_{\substack{\mathbf{j} \in [q]^d \\ |\mathbf{j}|_1 = q}} \partial^{\mathbf{j}} f(x_{\mathbf{i}}) \frac{(x - x_{\mathbf{i}})^{\mathbf{j}}}{\mathbf{j}!} \right| \\ & \leq \left( \frac{2a}{M} \right)^q \|\xi - x_{\mathbf{i}}\|^s \cdot C. \end{aligned} \quad (2.10)$$

Fassen wir die Gleichungen (2.8), (2.9) und (2.10) zusammen, erhalten wir:

$$\begin{aligned} |f(x) - P_f(x)| &\leq \sum_{\mathbf{i} \in [M]^d} \left| f(x) - \sum_{\substack{\mathbf{j} \in [q]^d \\ |\mathbf{j}|_1 \leq q}} \partial^{\mathbf{j}} f(x_{\mathbf{i}}) \cdot \frac{(x - x_{\mathbf{i}})^{\mathbf{j}}}{\mathbf{j}!} \right| \prod_{j=1}^d \left( 1 - \frac{M}{2a} \cdot |x^{(j)} - x_{\mathbf{i}}^{(j)}| \right)_+ \\ &\leq \left( \frac{2a}{M} \right)^q \|\xi - x_{\mathbf{i}}\|^s \cdot C \cdot \sum_{\mathbf{i} \in [M]^d} \prod_{j=1}^d \left( 1 - \frac{M}{2a} \cdot |x^{(j)} - x_{\mathbf{i}}^{(j)}| \right)_+ \\ &\leq \left( \frac{2a}{M} \right)^q \cdot \left( \frac{2a}{M} \right)^s \cdot d^{s/2} \cdot C \\ &= c \cdot \left( \frac{a}{M} \right)^p, \end{aligned}$$

wobei wir bei der letzten Ungleichung die Eigenschaft der Zerlegung der Eins aus Gleichung (2.4) und  $q + s = p$  verwendet haben, mit

$$c = 2^p \cdot d^{s/2} \cdot C.$$

Bilden wir schließlich noch das Supremum über  $x \in [-a, a]^d$ , erhalten wir die Behauptung.  $\square$

Durch geeignet gewählte  $a_{\mathbf{i}, \mathbf{j}} \in \mathbb{R}$  lässt sich  $P_m(x)$  in die Form

$$\sum_{\mathbf{i} \in [M]^d} \sum_{\substack{\mathbf{j} \in [q]^d \\ |\mathbf{j}|_1 \leq q}} a_{\mathbf{i}, \mathbf{j}} \cdot (x - x_{\mathbf{i}})^{\mathbf{j}} \prod_{j=1}^d \left( 1 - \frac{M}{2a} \cdot |x^{(j)} - x_{\mathbf{i}}^{(j)}| \right)_+$$

bringen, da sich jedes  $p_{\mathbf{i}}^m(x)$  als Polynom umordnen lässt und wir daher auch  $P_m(x)$  umschreiben können.

Als Nächstes wollen wir geeignete neuronale Netze  $f_{\text{net}, \mathbf{j}, \mathbf{i}}$  mit Architektur  $(L, \mathbf{k})$  definieren, die die Funktionen

$$x \mapsto (x - x_{\mathbf{i}})^{\mathbf{j}} \prod_{j=1}^d \left( 1 - \frac{M}{2a} \cdot |x^{(j)} - x_{\mathbf{i}}^{(j)}| \right)_+$$

approximieren. Zudem möchten wir die Architektur  $(L, \mathbf{k})$  so wählen, dass neuronale Netze der Form

$$\sum_{\mathbf{i} \in [M]^d} \sum_{\substack{\mathbf{j} \in [q]^d \\ |\mathbf{j}|_1 \leq q}} a_{\mathbf{i}, \mathbf{j}} \cdot f_{\text{net}, \mathbf{j}, \mathbf{i}}(x) \quad (a_{\mathbf{i}, \mathbf{j}} \in \mathbb{R})$$

in der Klasse  $\mathfrak{N}(L, \mathbf{k}, \sigma)$  enthalten sind. Um dies zu erreichen, sei

$$\sigma(x) = \frac{1}{(1 + \exp(-x))} \quad (x \in \mathbb{R})$$

der logistische Squasher (1.2). Zudem wählen wir  $R \geq 1$  und definieren die folgenden neuronalen Netze:

Das neuronale Netz

$$f_{\text{id}}(x) = 4R \cdot \sigma\left(\frac{x}{R}\right) - 2R, \quad (2.11)$$

mit Architektur  $(1, (1))$ , welches, wie in Lemma 1.6 gezeigt, die Funktion  $f(x) = x$  approximiert und in Abbildung 1.2 veranschaulicht wurde.

Das neuronale Netz

$$f_{\text{mult}}(x, y) = \frac{R^2}{4} \cdot \frac{(1 + \exp(-1))^3}{\exp(-2) - \exp(-1)} \cdot \left( \sigma\left(\frac{2(x+y)}{R} + 1\right) - 2 \cdot \sigma\left(\frac{x+y}{R} + 1\right) - \sigma\left(\frac{2(x-y)}{R} + 1\right) + 2 \cdot \sigma\left(\frac{x-y}{R} + 1\right) \right), \quad (2.12)$$

mit Architektur  $(1, (4))$ , welches, wie in Lemma 1.7 gezeigt, die Funktion  $f(x, y) = x \cdot y$  approximiert und in Abbildung 1.3 veranschaulicht wurde.

Das neuronale Netz

$$f_{\text{ReLU}}(x) = f_{\text{mult}}(f_{\text{id}}(x), \sigma(R \cdot x)), \quad (2.13)$$

mit Architektur  $(2, (2, 4))$ , welches, wie in Lemma 1.8 gezeigt, die Funktion  $f(x) = x_+$  approximiert und in Abbildung 1.4 veranschaulicht wurde und schließlich das neuronale Netz

$$f_{\text{hat},y}(x) = f_{\text{ReLU}}\left(\frac{M}{2a} \cdot (x-y) + 1\right) - 2 \cdot f_{\text{ReLU}}\left(\frac{M}{2a} \cdot (x-y)\right) + f_{\text{ReLU}}\left(\frac{M}{2a} \cdot (x-y) - 1\right), \quad (2.14)$$

mit Architektur  $(2, (6, 12))$ , welches, wie in Lemma 1.9 gezeigt, für fixes  $y \in \mathbb{R}$  die Funktion

$$f(x) = \left(1 - \left(\frac{M}{2a}\right) \cdot |x-y|\right)_+$$

approximiert und in Abbildung 1.5 veranschaulicht wurde.

Mit diesen neuronalen Netzen können wir nun  $f_{\text{net},\mathbf{j},\mathbf{i}}$  rekursiv definieren. Dafür wählen wir  $N \geq q$ , setzen  $s = \lceil \log_2(N+d) \rceil$  und definieren für  $k \in \{1, \dots, (M+1)^d\}$ ,  $\mathbf{i} \in [M]^d$  und  $\mathbf{j} \in [N]^d$ :

$$f_{\text{net},\mathbf{j},\mathbf{i}}(x) = f_1^{(0)}(x),$$

wobei

$$f_k^{(l)}(x) = f_{\text{mult}}\left(f_{2k-1}^{(l+1)}(x), f_{2k}^{(l+1)}(x)\right)$$

für  $k \in \{1, 2, \dots, 2^l\}$  und  $l \in \{0, \dots, s-1\}$ . Zudem ist

$$f_k^{(s)}(x) = f_{\text{id}}(f_{\text{id}}(x^{(l)} - x_{\mathbf{i}_k}^{(l)})) \quad (2.15)$$

für  $j_1 + j_2 + \dots + j_{l-1} + 1 \leq k \leq j_1 + j_2 + \dots + j_l$  und  $1 \leq l \leq d$  und

$$f_{|\mathbf{j}|_1+k}^{(s)}(x) = f_{\text{hat},x_{\mathbf{i}_k}^{(k)}}(x^{(k)})$$

für  $1 \leq k \leq d$  und

$$f_k^{(s)}(x) = 1$$

für  $|\mathbf{j}|_1 + d + 1 \leq k \leq 2^s$ .

Da das neuronale Netz  $f_{\text{net},\mathbf{j},\mathbf{i}}$  aus mehreren neuronalen Netzen zusammengebaut wurde, lässt sich dadurch auch die Anzahl an Schichten und Neuronen pro Schicht durch diese Struktur erklären. Aus der rekursiven Definition (2.15) entnimmt man, dass  $f_{\text{net},\mathbf{j},\mathbf{i}}$   $s+2$  verborgene Schichten, durch  $s$ -maliges Anwenden von  $f_{\text{mult}}$  und einer Anwendung von  $f_{\text{hat}}$  bzw.  $f_{\text{id}}(f_{\text{id}})$ , hat. Da  $f_{\text{hat}}$  zwei verborgene Schichten besitzt, ergibt sich daraus die Anzahl an verborgenen Schichten von  $f_{\text{net},\mathbf{j},\mathbf{i}}$ . Die Anzahl der Neuronen pro verborgener Schicht von  $f_{\text{net},\mathbf{j},\mathbf{i}}$  ergeben sich wie folgt:

- Die erste verborgene Schicht enthält maximal  $3 \cdot 2 \cdot 2^s = 6 \cdot 2^s$  Neuronen, da dies die erste verborgene Schicht von  $f_{\text{hat}}$  ist und maximal  $(2^s)$ -mal aufgerufen wird.
- Die zweite verborgene Schicht maximal  $3 \cdot 4 \cdot 2^s = 12 \cdot 2^s$  Neuronen, da dies die zweite verborgene Schicht von  $f_{\text{hat}}$  ist und maximal  $2^s$ -mal aufgerufen wird.
- Die verborgenen Schichten  $3, \dots, s+2$  enthalten maximal  $2^{s+1}, 2^s, \dots, 2^3, 2^2$  Neuronen, da wir  $s$ -mal  $f_{\text{mult}}$  verschachtelt aufrufen.

Da man, wie in Kapitel 1 bereits erwähnt, bei fully connected neuronalen Netzen die Gewichte der Verbindungen zwischen zwei Neuronen auf Null setzen kann, können auch nicht fully connected neuronale Netze in dieser Klasse enthalten sein. Daher liegt auch  $f_{\text{net},\mathbf{j},\mathbf{i}}$  nach Definition 1.1 in  $\mathfrak{N}(s+2, \{24 \cdot (N+d)\}^{s+2}, \sigma)$ , da die größte Anzahl an Neuronen in einer Schicht

$$12 \cdot 2^s = 12 \cdot 2^{\lceil \log_2(N+d) \rceil} \leq 12 \cdot 2^{\log_2(N+d)+1} = 24 \cdot (N+d)$$

ist. Weiterhin erkennt man durch die Zusammensetzung der neuronalen Netze, dass alle Gewichte im Betrag durch  $c \cdot \max\{\frac{M}{2a}, R^2\}$  beschränkt sind, wobei  $c > 0$  ist.

## 2.2 Bestimmung der Gewichte der Ausgabeschicht

Wir definieren unseren Neuronale-Netze-Regressionsschätzer  $\tilde{m}_n(x)$  durch:

$$\tilde{m}_n(x) := \sum_{\mathbf{i} \in [M]^d} \sum_{\substack{\mathbf{j} \in [N]^d \\ |\mathbf{j}|_1 \leq N}} a_{\mathbf{i},\mathbf{j}} \cdot f_{\text{net},\mathbf{j},\mathbf{i}}(x), \quad (2.16)$$

wobei  $n$  die Größe unserer gegebenen Datenmenge (2.1) ist und wir die Koeffizienten  $a_{\mathbf{i},\mathbf{j}}$  durch eine Tikhonov Regularisierung [Kre98, Kapitel 16.1] des Funktional

$$\varphi(a_{\mathbf{i},\mathbf{j}}) := \frac{1}{n} \sum_{i=1}^n |Y_i - \tilde{m}_n(X_i)|^2 + \frac{c}{n} \cdot \|a_{\mathbf{i},\mathbf{j}}\|_2^2 \quad (2.17)$$

mit Regularitätsparameter  $\frac{c}{n}$ , für eine von  $n$  unabhängigen Konstante  $c > 0$  erhalten.

Wir wollen Gleichung (2.17) als Gleichungssystem darstellen. Dafür definieren wir uns die Menge

$$\{U_s : s = 1, \dots, S\} := \left\{ f_{\text{net},\mathbf{j},\mathbf{i}}(x) : \mathbf{i} \in [M]^d \text{ und } |\mathbf{j}|_1 \leq N \text{ mit } \mathbf{j} \in [N]^d \right\}$$

wobei

$$S := |[M]^d| \cdot \binom{N+d}{d} = (M+1)^d \cdot \binom{N+d}{d}$$



die Kardinalität der Menge ist. Diese Kardinalität erhalten wir mit einem Kombinatorik Argument. Wir wissen, dass es insgesamt  $(M+1)^d$  Möglichkeiten gibt  $d$ -viele Zahlen aus einer Menge mit der Größe  $(M+1)$  zu ziehen, mit Zurücklegen und da wir Vektoren betrachten und die Komponenten nicht vertauschbar sind, ist auch die Reihenfolge der Ziehung zu beachten. Für jedes dieser  $(M+1)^d$  Möglichkeiten ist noch zu beachten, dass wir zusätzlich  $d$ -mal aus einer Menge mit  $(N+1)$ -vielen Zahlen ziehen und gleichzeitig die Bedingung beachten müssen, dass die Summe der gezogen  $d$  Elemente zwischen Null und  $N$  liegt. Gesucht ist also

$$\left| \left\{ \mathbf{j} \in [N]^d : |\mathbf{j}|_1 \leq N \right\} \right| =: H.$$

Wir stellen fest, dass

$$\begin{aligned} & \left\{ \mathbf{j} \in [N]^d : |\mathbf{j}|_1 \leq N \right\} \\ &= \left\{ \mathbf{j} \in [N]^d : |\mathbf{j}|_1 = 0 \right\} \cup \left\{ \mathbf{j} \in [N]^d : |\mathbf{j}|_1 = 1 \right\} \cup \dots \cup \left\{ \mathbf{j} \in [N]^d : |\mathbf{j}|_1 = N \right\} \end{aligned}$$

gilt. Mit Lemma 2.3 wissen wir, dass für  $d, N \in \mathbb{N}$  und  $k \in \mathbb{N}_0$  die Identität

$$\left| \left\{ \mathbf{j} \in [N]^d : |\mathbf{j}|_1 = k \right\} \right| = \binom{d+k-1}{k}$$

gilt. Damit erhalten wir

$$|H| = \sum_{k=0}^N \binom{N-1+k}{k} = \binom{N+d}{d},$$

mit der *Hockey-Stick Identität*

$$\binom{n}{k} = \sum_{i=0}^k \binom{n-k-1+i}{i} \quad (k, n \in \mathbb{N} \text{ mit } k < n).$$

Wir setzen nun

$$\mathbf{U} = (U_s(X_i))_{1 \leq i \leq n, 1 \leq s \leq S} \quad \text{und} \quad \mathbf{Y} = (Y_i)_{i=1, \dots, n}.$$

Im folgenden Lemma bestimmen wir den Koeffizientenvektor unseres Schätzers  $\tilde{m}_n$ .

**Lemma 2.6.** *Der Koeffizientenvektor unseres Schätzers  $\tilde{m}_n$  aus Gleichung (2.16) ist die eindeutige Lösung des linearen Gleichungssystems*

$$\left( \frac{1}{n} \mathbf{U}^T \mathbf{U} + \frac{c}{n} \cdot \mathbf{1} \right) \mathbf{a} = \frac{1}{n} \mathbf{U}^T \mathbf{Y} \quad (2.18)$$

Hierbei ist  $\mathbf{1}$  eine  $S \times S$ -Einheitsmatrix und  $\mathbf{a} \in \mathbb{R}^S$ , wobei  $\mathbf{a}$  wie in Gleichung (2.17) den Ausdruck:

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n |Y_i - \tilde{m}_n(X_i)|^2 + \frac{c}{n} \cdot \sum_{\mathbf{i} \in [M]^d} \sum_{\substack{\mathbf{j} \in [N]^d \\ \|\mathbf{j}\|_1 \leq N}} a_{\mathbf{i}, \mathbf{j}}^2 \\ &= \frac{1}{n} (\mathbf{Y} - \mathbf{U}\mathbf{a})^T (\mathbf{Y} - \mathbf{U}\mathbf{a}) + \frac{c}{n} \cdot \mathbf{a}^T \mathbf{a} \end{aligned} \quad (2.19)$$

minimiert.

*Beweis.* Der Schätzer aus Gleichung (2.16) lässt sich umschreiben zu

$$\tilde{m}_n(x) = \sum_{s=1}^S a_s \cdot U_s(x) \quad (2.20)$$

mit  $(a_s)_{s=1, \dots, S} = \mathbf{a} \in \mathbb{R}^S$ .

Es gilt  $\mathbf{a}^T \mathbf{U}^T \mathbf{Y} = \mathbf{Y}^T \mathbf{U}\mathbf{a}$ , da dieser Ausdruck eine reelle Zahl und damit insbesondere symmetrisch ist. Damit erhalten wir für die rechte Seite aus Gleichung (2.19):

$$\begin{aligned} & \frac{1}{n} (\mathbf{Y} - \mathbf{U}\mathbf{a})^T (\mathbf{Y} - \mathbf{U}\mathbf{a}) + \frac{c}{n} \cdot \mathbf{a}^T \mathbf{a} \\ &= \frac{1}{n} (\mathbf{Y}^T \mathbf{Y} - \mathbf{Y}^T \mathbf{U}\mathbf{a} - \mathbf{a}^T \mathbf{U}^T \mathbf{Y} + \mathbf{a}^T \mathbf{U}^T \mathbf{U}\mathbf{a}) + \frac{c}{n} \cdot \mathbf{a}^T \mathbf{a} \\ &= \frac{1}{n} (\mathbf{Y}^T \mathbf{Y} - 2\mathbf{Y}^T \mathbf{U}\mathbf{a}) + \mathbf{a}^T \left( \frac{1}{n} \mathbf{U}^T \mathbf{U} + \frac{c}{n} \cdot \mathbf{1} \right) \mathbf{a}. \end{aligned} \quad (2.21)$$

Die Matrix  $\mathbf{U}^T \mathbf{U} \in \mathbb{R}^{S \times S}$  ist positiv definit, denn aufgrund der Rechenregeln der Transponierten und des Standardskalarprodukts sowie der positiven Definitheit des Standardskalarprodukts gilt für alle  $x \in \mathbb{R} \setminus \{0\}$ :

$$\langle x, \mathbf{U}^T \mathbf{U} x \rangle = \langle \mathbf{U} x, \mathbf{U} x \rangle > 0.$$

Zudem wissen wir dass  $\frac{c}{n} \mathbf{1}$  durch die Wahl von  $c$  nur positive Eigenwerte besitzt und damit positiv definit ist. Daher wissen wir, dass die Matrix

$$\mathbf{A} := \frac{1}{n} \mathbf{U}^T \mathbf{U} + \frac{c}{n} \cdot \mathbf{1}$$

positiv definit und insbesondere invertierbar ist, da die Eigenwerte positiv sind. Das folgt daraus, dass die ohnehin schon positiven Eigenwerte von  $\frac{1}{n} \mathbf{U}^T \mathbf{U}$  um  $\frac{c}{n}$  verschoben werden und damit positiv bleiben. Zudem ist die Matrix  $\mathbf{A}$  symmetrisch. Mit

$$\mathbf{b} = \frac{1}{n} \cdot \mathbf{A}^{-1} \mathbf{U}^T \mathbf{Y} \in \mathbb{R}^S$$

und

$$\mathbf{b}^T \mathbf{A} \mathbf{a} = \mathbf{a}^T \mathbf{A} \mathbf{b} = \frac{1}{n} \cdot \mathbf{a}^T \mathbf{U}^T \mathbf{Y} \in \mathbb{R},$$

was aus der Symmetrie von  $\mathbf{A}$  folgt, erhalten wir mit

$$\mathbf{Y}^T \mathbf{U} \mathbf{a} = \frac{n}{2} \cdot \mathbf{b}^T \mathbf{A} \mathbf{a} + \frac{n}{2} \cdot \mathbf{a}^T \mathbf{A} \mathbf{b}$$

mithilfe von

$$0 = \mathbf{b}^T \mathbf{U}^T \mathbf{Y} - \mathbf{Y}^T \mathbf{U} \mathbf{b}$$

die Gleichung

$$\begin{aligned} \mathbf{Y}^T \mathbf{U} \mathbf{a} &= \frac{n}{2} (\mathbf{b}^T \mathbf{A} \mathbf{a} + \mathbf{a}^T \mathbf{A} \mathbf{b}) = \frac{n}{2} \left( \mathbf{b}^T \mathbf{A} \mathbf{a} + \mathbf{a}^T \mathbf{A} \mathbf{b} - \frac{1}{n} \mathbf{b}^T \mathbf{U}^T \mathbf{Y} + \frac{1}{n} \mathbf{Y}^T \mathbf{U} \mathbf{b} \right) \\ &= \frac{n}{2} \left( \mathbf{b}^T \mathbf{A} \mathbf{a} + \mathbf{a}^T \mathbf{A} \mathbf{b} - \frac{1}{n} \mathbf{b}^T \mathbf{A} \mathbf{A}^{-1} \mathbf{U}^T \mathbf{Y} + \frac{1}{n} \mathbf{Y}^T \mathbf{U} \mathbf{b} \right) \\ &= \frac{n}{2} \left( \mathbf{b}^T \mathbf{A} \mathbf{a} + \mathbf{a}^T \mathbf{A} \mathbf{b} - \mathbf{b}^T \mathbf{A} \mathbf{b} + \frac{1}{n^2} \mathbf{Y}^T \mathbf{U} \mathbf{A}^{-1} \mathbf{U}^T \mathbf{Y} \right). \end{aligned} \quad (2.22)$$

Damit erhalten wir in Gleichung (2.21):

$$\begin{aligned} &\frac{1}{n} (\mathbf{Y}^T \mathbf{Y} - 2 \mathbf{Y}^T \mathbf{U} \mathbf{a}) + \mathbf{a}^T \left( \frac{1}{n} \mathbf{U}^T \mathbf{U} + \frac{c}{n} \cdot \mathbf{1} \right) \mathbf{a} \\ &= \frac{1}{n} (\mathbf{Y}^T \mathbf{Y} - 2 \mathbf{Y}^T \mathbf{U} \mathbf{a}) + \mathbf{a}^T \mathbf{A} \mathbf{a} \\ &= \frac{1}{n} \mathbf{Y}^T \mathbf{Y} - \mathbf{b}^T \mathbf{A} \mathbf{a} - \mathbf{a}^T \mathbf{A} \mathbf{b} + \mathbf{b}^T \mathbf{A} \mathbf{b} - \frac{1}{n^2} \mathbf{Y}^T \mathbf{U} \mathbf{A}^{-1} \mathbf{U}^T \mathbf{Y} + \mathbf{a}^T \mathbf{A} \mathbf{a} \\ &= \left( \mathbf{a} - \frac{1}{n} \cdot \mathbf{A}^{-1} \mathbf{U}^T \mathbf{Y} \right)^T \mathbf{A} \left( \mathbf{a} - \frac{1}{n} \cdot \mathbf{A}^{-1} \mathbf{U}^T \mathbf{Y} \right) + \frac{1}{n} \mathbf{Y}^T \mathbf{Y} - \frac{1}{n^2} \mathbf{Y}^T \mathbf{U} \mathbf{A}^{-1} \mathbf{U}^T \mathbf{Y}. \end{aligned}$$

Die letzte Gleichung wird für  $\mathbf{a} = \frac{1}{n} \cdot \mathbf{A}^{-1} \mathbf{U}^T \mathbf{Y}$  minimal, da wir wissen, dass  $\mathbf{A}$  positiv definit ist und damit  $x^T \mathbf{A} x > 0$  für alle  $x \in \mathbb{R}^S$  mit  $x \neq 0$  gilt. Wir erhalten dann insbesondere  $(\mathbf{a} - \mathbf{b})^T \mathbf{A} (\mathbf{a} - \mathbf{b}) = 0$  genau dann, wenn  $\mathbf{a} = \mathbf{b}$  gilt. Dies zeigt also, dass der Koeffizientenvektor unseres Schätzers (2.16) die eindeutige Lösung des linearen Gleichungssystems (2.18) ist.  $\square$

*Bemerkung 2.7.* Da der Koeffizientenvektor  $\mathbf{a}$  die Gleichung (2.19) minimiert, erhalten wir, wenn wir den Koeffizientenvektor gleich Null setzen:

$$\frac{1}{n} (\mathbf{Y} - \mathbf{U} \mathbf{a})^T (\mathbf{Y} - \mathbf{U} \mathbf{a}) + \frac{c}{n} \cdot \mathbf{a}^T \mathbf{a} \leq \frac{1}{n} \sum_{i=1}^n Y_i^2,$$

was uns erlaubt eine obere Schranke für den absoluten Wert unserer Koeffizienten abzuleiten. Daraus können wir folgern, dass unser Neuronale-Netze-Regressionsschätzer  $\tilde{m}_n$  beschränkt ist, da die neuronalen Netze  $f_{\text{net}, \mathbf{j}, \mathbf{i}}$  nach Konstruktion ebenfalls beschränkt sind.

## Kapitel 3

# Resultat zur Konvergenzgeschwindigkeit

In diesem Kapitel stellen wir das Hauptresultat dieser Arbeit vor. Wir betrachten im Folgenden das Problem 2.1 und wollen nun eine Abschätzung des erwarteten  $L_2$ -Fehlers

$$\mathbb{E} \left[ \int |m_n(x) - m(x)|^2 \mathbb{P}_X(dx) \right]$$

im Falle des Schätzers

$$m_n(x) := T_{\beta_n} \tilde{m}_n(x), \quad (3.1)$$

mit  $\beta_n = c_1 \cdot \log(n)$  für eine hinreichend große und von  $n$  unabhängige Konstante  $c_1 > 0$  und unter Annahme einer  $(p, C)$ -glatten Regressionsfunktion  $m$  herleiten. Hierbei bezeichnet  $T_\beta z = \max\{\min\{z, \beta\}, -\beta\}$  für  $z \in \mathbb{R}$  und  $\beta > 0$ . Weiterhin bezeichnen wir mit  $\tilde{m}_n$  unseren Neuronale-Netze-Regresssionsschätzer aus Kapitel 2.2, welcher aus mehreren neuronalen Netzen konstruiert wurde. Für diesen gilt: Die Aktivierungsfunktion  $\sigma$  ist der logistische Squasher,  $N \geq q$ ,  $M = M_n = \lceil c_2 \cdot n^{1/(2p+d)} \rceil$  mit  $c_2 > 0$  und unabhängig von  $n$ ,  $R = R_n = n^{d+4}$  und  $a = a_n = (\log n)^{1/(6(N+d))}$ .

Nun kommen wir zu dem Hauptresultat dieser Arbeit.

**Hauptsatz 3.1** ([AB19, Theorem 1]). *Angenommen die Verteilung von  $Y$  erfüllt*

$$\mathbb{E} \left[ e^{c_3 \cdot |Y|^2} \right] < \infty$$

*für eine Konstante  $c_3 > 0$  und die Verteilung von  $X$  besitzt einen beschränkten Träger  $\text{supp}(\mathbb{P}_X)$ . Sei  $m(x) = \mathbb{E}[Y | X = x]$  die zu dem Tupel  $(X, Y)$  gehörige Regressionsfunktion. Angenommen  $m$  ist  $(p, C)$ -glatt, mit  $p = q + s$ , wobei  $q \in \mathbb{N}_0$ ,  $s \in (0, 1]$  und  $C > 0$  ist. Sei  $m_n$  aus Gleichung (3.1) gegeben.*

*Dann erhalten wir für hinreichend großes  $n$ :*

$$\mathbb{E} \left[ \int |m_n(x) - m(x)|^2 \mathbb{P}_X(dx) \right] \leq c_{\text{fin}} \cdot (\log n)^3 \cdot n^{-\frac{2p}{2p+d}},$$

wobei  $c_{\text{fin}} > 0$  und unabhängig von  $n$  ist.

Der folgende Abschnitt liefert Hilfsresultate, welche wir im Beweis unsere Hauptresultats gebrauchen werden.

### 3.1 Approximationsresultate für Hauptsatz 3.1

Die nächsten Definitionen und Lemmata benötigen wir für den Beweis unseres Hauptresultats, einer Aussage über die Konvergenzgeschwindigkeit unseres Neuronale-Netze-Regressionsschätzers. Die Lemmata werden hier nur der Vollständigkeit halber und ohne Beweis aufgeführt.

**Lemma 3.2** ([AB19, Lemma 5]). *Sei  $M \in \mathbb{N}$  und  $\sigma: \mathbb{R} \rightarrow [0, 1]$  2-zulässig. Sei  $a \geq 1$  und  $R \in \mathbb{R}$  mit*

$$R \geq \max \left\{ \frac{\|\sigma''\|_{\infty} \cdot (M+1)}{2 \cdot |\sigma'(t_{\sigma})|}, \frac{9 \cdot \|\sigma''\|_{\infty} \cdot a}{|\sigma'(t_{\sigma})|}, \frac{20 \cdot \|\sigma'''\|_{\infty}}{3 \cdot |\sigma''(t_{\sigma})|} \cdot 3^{3 \cdot 3^s} \cdot a^{3 \cdot 2^s}, 1792 \cdot \frac{\max\{\|\sigma''\|_{\infty}, \|\sigma'''\|_{\infty}, 1\}}{\min\{2 \cdot |\sigma'(t_{\sigma})|, |\sigma''(t_{\sigma})|, 1\}} \cdot M^3 \right\},$$

wobei  $s = \lceil \log_2(N+d) \rceil$  mit  $N \in \mathbb{N}$ . Sei  $y \in [-a, a]^d$  und  $\mathbf{j} \in [\mathbb{N}_0]^d$  so, dass  $|\mathbf{j}|_1 \leq N$  gilt. Seien  $f_{\text{id}}, f_{\text{mult}}$  und  $f_{\text{hat}, z}$  (für  $z \in \mathbb{R}$ ) die neuronalen Netze aus Lemma 1.6, Lemma 1.7 und Lemma 1.9. Wir definieren das Netz  $f_{\text{net}, \mathbf{j}, y}$  so wie  $f_{\text{net}, \mathbf{j}, \mathbf{i}}$  in Kapitel 2.1 Gleichung (2.15), wobei  $x_{i_k}$  durch  $y \in \mathbb{R}^d$  ersetzt wird. Dann erhalten wir für  $x \in [-a, a]^d$ :

$$\left| f_{\text{net}, \mathbf{j}, y}(x) - (x-y)^{\mathbf{j}} \prod_{j=1}^d \left(1 - \frac{M}{2a} \cdot |x^{(j)} - y^{(j)}|\right)_+ \right| \leq c \cdot 3^{3 \cdot 3^s} \cdot a^{3 \cdot 2^s} \cdot M^3 \cdot \frac{1}{R},$$

für eine von  $n$  unabhängige Konstante  $c > 0$ .

Als Nächstes geben wir eine Definition von Überdeckungszahlen an, da wir im Beweis für unser Hauptresultat eine Abschätzung einer  $L_p$ - $\varepsilon$ -Überdeckungszahl anwenden.

**Definition 3.3.** Sei  $(X, \delta)$  ein pseudometrischer Raum (siehe z.B. [Bar15, Definition 2.1.1]). Für  $x \in X$  und  $\varepsilon > 0$  sei:

$$U_{\varepsilon}(x) = \{z \in X : \delta(x, z) < \varepsilon\}$$

die offene Kugel um  $x$  mit Radius  $\varepsilon$ .

a)  $\{z_1, \dots, z_N\} \subseteq X$  heißt  $\varepsilon$ -Überdeckung einer Menge  $A \subseteq X$ , falls gilt:

$$A \subseteq \bigcup_{k=1}^N U_\varepsilon(z_k).$$

b) Ist  $A \subseteq X$  und  $\varepsilon > 0$ , so ist die sogenannte  $\varepsilon$ -Überdeckungszahl von  $A$  in  $(X, \delta)$  definiert als:

$$\mathcal{N}_{(X, \delta)}(\varepsilon, A) := \inf \{|U| : U \subseteq X \text{ ist } \varepsilon\text{-Überdeckung von } A\}.$$

**Definition 3.4.** Sei  $\mathcal{F}$  eine Menge von Funktionen  $f: \mathbb{R}^d \rightarrow \mathbb{R}$ , sei  $\varepsilon > 0$ ,  $1 \leq p < \infty$  und seien  $x_1, \dots, x_n \in \mathbb{R}^d$  und  $x_1^n = (x_1, \dots, x_n)$ . Dann ist die  $L_p$ - $\varepsilon$ -Überdeckungszahl von  $\mathcal{F}$  auf  $x_1^n$  definiert durch:

$$\mathcal{N}_p(\varepsilon, \mathcal{F}, x_1^n) := \mathcal{N}_{(X, \delta)}(\varepsilon, \mathcal{F}),$$

wobei der pseudometrische Raum  $(X, \delta)$  gegeben ist durch

- $X =$  Menge aller Funktionen  $f: \mathbb{R}^d \rightarrow \mathbb{R}$ ,
- $\delta(f, g) = \delta_p(f, g) = (\frac{1}{n} \sum_{i=1}^n |f(x_i) - g(x_i)|^p)^{1/p}$ .

Diese Definition benötigen wir, um nun im folgenden Lemma den erwarteten  $L_2$ -Fehler eines Schätzers mithilfe einer  $L_p$ - $\varepsilon$ -Überdeckungszahl abzuschätzen. Dieses Lemma ist ein Spezialfall von [AB19, Lemma 8].

**Lemma 3.5.** Sei  $\beta_n = c_1 \cdot \log(n)$  für eine hinreichend große Konstante  $c_1 > 0$ . Angenommen die Verteilung von  $Y$  erfüllt

$$\mathbb{E} \left[ e^{c_2 \cdot |Y|^2} \right] < \infty$$

für eine Konstante  $c_2 > 0$ . Zudem nehmen wir an, dass der Betrag der Regressionsfunktion  $m$  beschränkt ist. Sei  $\mathcal{F}_n$  eine Menge von Funktionen  $f: \mathbb{R}^d \rightarrow \mathbb{R}$  und wir nehmen an, dass der Schätzer  $\hat{m}_n$

$$\hat{m}_n = T_{\beta_n} \dot{m}_n = \max\{\min\{\dot{m}_n, \beta_n\}, -\beta_n\}$$

erfüllt, mit

$$\dot{m}_n(\cdot) = \dot{m}_n(\cdot, (X_1, Y_1), \dots, (X_n, Y_n)) \in \mathcal{F}_n$$

und

$$\frac{1}{n} \sum_{i=1}^n |Y_i - \dot{m}_n(X_i)|^2 \leq \frac{1}{n} \sum_{i=1}^n |Y_i - g_n(X_i)|^2 + \text{pen}_n(g_n) \quad (3.2)$$

mit einer deterministischen Funktion  $g_n: \mathbb{R}^d \rightarrow \mathbb{R}$  und deterministischem Penalty Term  $\text{pen}_n(g_n) \geq 0$ , wobei die Funktion  $g_n: \mathbb{R}^d \rightarrow \mathbb{R}$  unabhängig von  $(X_1, Y_1), \dots, (X_n, Y_n)$  ist. Dann gilt für den erwarteten  $L_2$ -Fehler die Ungleichung

$$\begin{aligned} & \mathbb{E} \left[ \int |\hat{m}_n(x) - m(x)|^2 \mathbb{P}_X(dx) \right] \\ & \leq \frac{c \cdot \log(n)^2 \cdot \left( \log \left( \sup_{x_1^n \in (\text{supp}(X))^n} \mathcal{N}_1 \left( \frac{1}{n \cdot \beta_n}, \mathcal{F}_n, x_1^n \right) \right) + 1 \right)}{n} \\ & \quad + 2 \cdot \mathbb{E} \left[ \int |g_n(x) - m(x)|^2 \mathbb{P}_X(dx) + \text{pen}_n(g_n) \right], \end{aligned} \quad (3.3)$$

für  $n > 1$  und eine von  $n$  unabhängige Konstante  $c > 0$ .

Das nächste Lemma benötigen wir, um in Ungleichung (3.3) die Überdeckungszahl  $\mathcal{N}_1 \left( \frac{1}{n \cdot \beta_n}, \mathcal{F}_n, x_1^n \right)$  weiter abzuschätzen.

**Lemma 3.6** ([AB19, Lemma 9]). *Seien  $a > 0$  und  $d, L, N, J_n \in \mathbb{N}$  so, dass  $J_n \leq n^{c_1}$  und setze  $\beta_n = c_2 \cdot \log(n)$ . Sei  $\sigma: \mathbb{R} \rightarrow [0, 1]$  2-zulässig. Sei  $\mathcal{F}$  die Menge aller Funktionen die durch Definition 1.1 definiert sind mit  $k_1 = k_2 = \dots = k_L = 24 \cdot (N + d)$  und einer Beschränkung des Betrags der Gewichte durch  $c_3 \cdot n^{c_4}$ .*

$$\mathcal{F}^{(J_n)} = \left\{ \sum_{j=1}^{J_n} a_j \cdot f_j : f_j \in \mathcal{F} \quad \text{und} \quad \sum_{j=1}^{J_n} a_j^2 \leq c_5 \cdot n^{c_6} \right\}.$$

Dann gilt für  $n > 1$ :

$$\log \left( \sup_{x_1^n \in [-a, a]^{d \cdot n}} \mathcal{N}_1 \left( \frac{1}{n \cdot \beta_n}, \mathcal{F}^{(J_n)}, x_1^n \right) \right) \leq c \cdot \log(n) \cdot J_n,$$

für eine Konstante  $c$  die nur von  $L, N, a$  und  $d$  abhängt.

Mit den Approximationsresultaten aus diesem Abschnitt verfügen wir nun über alle Bausteine, um unser Hauptresultat zu beweisen.

## 3.2 Der Beweis von Hauptsatz 3.1

Wir betrachten das Ereignis

$$A_n := \left[ \frac{1}{n} \sum_{i=1}^n Y_i^2 \leq 1 + \mathbb{E}[Y^2] \right]. \quad (3.4)$$

Damit können wir den erwarteten  $L_2$ -Fehler umschreiben zu:

$$\begin{aligned}
 \mathbb{E} \left[ \int |m_n(x) - m(x)|^2 \mathbb{P}_X(dx) \right] &= \mathbb{E} \left[ \int |m_n(x) - m(x)|^2 \mathbb{P}_X(dx) \cdot (\mathbb{1}_{A_n^c} + \mathbb{1}_{A_n}) \right] \\
 &= \mathbb{E} \left[ \int |m_n(x) - m(x)|^2 \mathbb{P}_X(dx) \mathbb{1}_{A_n^c} \right] \\
 &\quad + \mathbb{E} \left[ \int |m_n(x) - m(x)|^2 \mathbb{P}_X(dx) \mathbb{1}_{A_n} \right] \\
 &=: T_{1,n} + T_{2,n}.
 \end{aligned} \tag{3.5}$$

In den folgenden Abschnitten kümmern wir uns um die Abschätzung der Summanden  $T_{1,n}$  und  $T_{2,n}$ .

### 3.2.1 Abschätzung von $T_{1,n}$

Für zwei beliebige reelle Zahlen  $u, v \in \mathbb{R}$  gilt durch  $0 \leq (u - v)^2 = u^2 + v^2 - 2uv$ :

$$u^2 + v^2 \geq 2uv$$

und damit schließlich:

$$|(u - v)^2| = |u^2 - 2uv + v^2| \leq u^2 + 2uv + v^2 \leq 2u^2 + 2v^2. \tag{3.6}$$

Wir wissen, dass aufgrund der Unabhängigkeit und identischen Verteiltheit der  $\mathbb{R}^d \times \mathbb{R}$ -wertigen Zufallsvariablen  $(X, Y), (X_1, Y_1), (X_2, Y_2), \dots$  die Zufallsvariablen  $Y_1, \dots, Y_n$  ebenso wie die Zufallsvariablen  $X_1, \dots, X_n$  unabhängig und identisch verteilt sind. Mit Ungleichung (3.6) und durch die Unabhängigkeit von  $A_n$  von den Zufallsvariablen  $X, X_1, \dots, X_n$  erhalten wir für den ersten Summanden aus Gleichung (3.5):

$$\begin{aligned}
 T_{1,n} &= \mathbb{E} \left[ \int |m_n(x) - m(x)|^2 \mathbb{P}_X(dx) \mathbb{1}_{A_n^c} \right] = \mathbb{E} \left[ \int |m_n(x) - m(x)|^2 \mathbb{P}_X(dx) \right] \cdot \mathbb{P}(A_n^c) \\
 &\leq \mathbb{E} \left[ \int 2m_n(x)^2 + 2m(x)^2 \mathbb{P}_X(dx) \right] \cdot \mathbb{P}(A_n^c) \\
 &\leq \mathbb{E} \left[ \int 2\beta_n^2 + 2\beta_n^2 \mathbb{P}_X(dx) \right] \cdot \mathbb{P}(A_n^c).
 \end{aligned} \tag{3.7}$$

Da nach Voraussetzung  $\text{supp}(\mathbb{P}_X)$  beschränkt ist und definitionsgemäß immer abgeschlossen ist, wissen wir nach dem Satz von Heine-Borel (siehe z.B. [For17, Satz 5]), dass  $\text{supp}(\mathbb{P}_X)$  kompakt ist. Da  $m$  als  $(p, C)$ -glatte Funktion insbesondere stetig ist, wissen wir, dass sie auf einer kompakten Menge ein Maximum und Minimum annimmt. Dadurch können wir  $n$  so groß wählen, dass ohne Beschränkung der Allgemeinheit  $\|m\|_\infty \leq \beta_n$



gilt. Da wir nach Bemerkung 2.7 wissen, dass  $\tilde{m}_n$  beschränkt ist, ist  $m_n$  nach Konstruktion ebenfalls beschränkt. Wir haben daher bei Ungleichung (3.7) zudem verwendet, dass  $\max\{\|m\|_\infty, \|m_n\|_\infty\} \leq \beta_n$  nach Definition von  $m_n$  gilt. Im nächsten Schritt wollen wir  $\mathbb{P}(A_n^c)$  abschätzen.

Da die Zufallsvariablen  $Y_1, \dots, Y_n$  unabhängig und identisch verteilt sind, folgern wir daraus  $\mathbb{E}[\frac{1}{n} \sum_{i=1}^n Y_i^2] = \mathbb{E}[Y^2]$  mit der Linearität des Erwartungswerts. Mithilfe der Monotonie der Wahrscheinlichkeitsfunktion  $\mathbb{P}$  und der Chebyshev-Ungleichung für  $\varepsilon = 1$  (siehe z.B. [?, Satz 5.11]) erhalten wir:

$$\mathbb{P}(A_n^c) = \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n Y_i^2 - \mathbb{E}[Y^2] \geq 1\right) \leq \mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^n Y_i^2 - \mathbb{E}[Y^2]\right| \geq 1\right) \leq \mathbb{V}\left[\frac{1}{n} \sum_{i=1}^n Y_i^2\right].$$

Da die Zufallsvariablen  $Y_1, \dots, Y_n$  u.i.v. sind, folgt mit den Rechenregeln der Varianz:

$$\mathbb{P}(A_n^c) \leq \frac{n \cdot \mathbb{V}[Y^2]}{n^2} = \frac{\mathbb{V}[Y^2]}{n} = \frac{c_4}{n}, \quad (3.8)$$

wobei  $c_4 := \mathbb{V}[Y^2]$  ist.

Mit Ungleichung (3.8) erhalten wir in Ungleichung (3.7):

$$\mathbb{E}\left[\int 2\beta_n^2 + 2\beta_n^2 \mathbb{P}_X(dx)\right] \cdot \mathbb{P}(A_n^c) \leq 4\beta_n^2 \cdot \mathbb{P}(A_n^c) \stackrel{(3.8)}{\leq} \frac{4 \cdot c_4 \cdot \beta_n^2}{n},$$

wobei wir bei der ersten Ungleichung verwendet haben, dass  $\beta_n$  deterministisch und  $\mathbb{P}(X \in \text{supp}(\mathbb{P}_X)) = 1$  ist. Schließlich erhalten wir in Ungleichung (3.7), da für  $n$  hinreichend groß  $\log(n)^3 > 1$  gilt:

$$T_{1,n} \leq \frac{4 \cdot c_4 \cdot \beta_n^2}{n} \leq c_5 \cdot \log(n)^3 \cdot n^{-1} \leq c_5 \cdot \log(n)^3 \cdot n^{-\frac{2p}{2p+d}}, \quad (3.9)$$

mit einer von  $n$  unabhängigen Konstante  $c_5 := 4 \cdot c_4 \cdot c_1^2 > 0$ .

Damit haben wir  $T_{1,n}$  abgeschätzt entsprechend der rechten Seite von Hauptsatz 3.1 und als nächstes wollen wir  $T_{2,n}$  abschätzen.

### 3.2.2 Abschätzung von $T_{2,n}$

In diesem Abschnitt wollen wir für die Abschätzung von  $T_{2,n}$  Lemma 3.5 anwenden. Sei dafür

$$\hat{m}_n := \mathbb{1}_{A_n} m_n + \mathbb{1}_{A_n^c} T_{\beta_n} g_n = T_{\beta_n} (\mathbb{1}_{A_n} \tilde{m}_n + \mathbb{1}_{A_n^c} g_n),$$

wobei

$$g_n(x) := \sum_{\mathbf{i} \in [M_n]^d} \sum_{\substack{\mathbf{j} \in [q]^d \\ |\mathbf{j}|_1 \leq q}} \frac{1}{\mathbf{j}!} \cdot \partial^{\mathbf{j}} m(x_{\mathbf{i}}) \cdot f_{\text{net}, \mathbf{j}, \mathbf{i}}(x).$$

Durch unsere Definition von  $\hat{m}_n$  erhalten wir durch die Monotonie des Erwartungswerts und einer Abschätzung über den ganzen Raum:

$$\begin{aligned} T_{2,n} &= \mathbb{E} \left[ \int |m_n(x) - m(x)|^2 \mathbb{P}_X(dx) \mathbb{1}_{A_n} \right] = \mathbb{E} \left[ \int |\hat{m}_n(x) - m(x)|^2 \mathbb{P}_X(dx) \mathbb{1}_{A_n} \right] \\ &\leq \mathbb{E} \left[ \int |\hat{m}_n(x) - m(x)|^2 \mathbb{P}_X(dx) \right]. \end{aligned} \quad (3.10)$$

Da  $m$  nach Voraussetzung  $(p, C)$ -glatt ist, existiert für alle  $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}_0^d$  mit  $\sum_{j=1}^d \alpha_j = q$  die partielle Ableitung  $\partial^\alpha m$ . Daher existiert ein  $z \in \mathbb{R}$  mit

$$z := \max_{\mathbf{i} \in [M_n]^d, \mathbf{j} \in [q]^d, |\mathbf{j}|_1 \leq q} \left| \frac{1}{\mathbf{j}!} \cdot \partial^{\mathbf{j}} m(x_{\mathbf{i}}) \right| < \infty. \quad (3.11)$$

Sei  $\mathcal{F}$  die Menge aller Funktion aus Definition 1.1 mit Aktivierungsfunktion  $\sigma$  und  $L = s + 2 = \lceil \log_2(N + d) \rceil + 2$ , mit  $k_1 = k_2 = \dots = k_L = 24 \cdot (N + d)$  und einer Konstante  $c_6 > 0$  so, dass der Betrag der Gewichte durch  $n^{c_6}$  beschränkt ist. Wir definieren

$$\mathcal{F}^{(J_n)} := \left\{ \sum_{j=1}^{J_n} a_j \cdot f_j : f_j \in \mathcal{F} \text{ und } \sum_{j=1}^{J_n} a_j^2 \leq c_7 \cdot n \right\}$$

mit

$$J_n = (M_n + 1)^d \cdot \left| \left\{ \mathbf{j} \in [N]^d : |\mathbf{j}|_1 \leq N \right\} \right|$$

und

$$c_7 := \max \left\{ \frac{1 + \mathbb{E}[Y^2]}{c_8}, c_2^d \cdot (N + 1)^d \cdot z^2 \right\}. \quad (3.12)$$

Hierbei bezeichnet  $c_8$  die Konstante des Regularitätsterms aus Gleichung (2.17). Da wir uns in der nichtparametrischen Regressionsschätzung befinden, gilt unter anderem die Bedingung  $\mathbb{E}[Y^2] < \infty$  und daher ist  $c_7$  auch wohldefiniert. Weiterhin folgt wie in Kapitel 2.2 mit  $J_n = S$ ,

$$J_n = (M_n + 1)^d \cdot \binom{N + d}{d} \leq (M_n + 1)^d \cdot (N + 1)^d. \quad (3.13)$$

Da nach Konstruktion  $f_{\text{net}, \mathbf{j}, \mathbf{i}} \in \mathcal{F}$  ist, folgt mit Gleichung (3.12), dass für  $n$  hinreichend groß

$$\begin{aligned} \sum_{\mathbf{i} \in [M_n]^d} \sum_{\substack{\mathbf{j} \in [q]^d \\ |\mathbf{j}|_1 \leq q}} \left| \frac{1}{\mathbf{j}!} \cdot \partial^{\mathbf{j}} m(x_{\mathbf{i}}) \right|^2 &\leq (M_n + 1)^d (N + 1)^d \cdot z^2 \\ &\leq (2c_2 \cdot n^{1/2p+d})^d \cdot (N + 1)^d \cdot z^2 \\ &\leq c_7 \cdot n, \end{aligned}$$

gilt und damit  $g_n$  in  $\mathcal{F}^{(J_n)}$  liegt.

Wir zeigen nun  $\tilde{m}_n \in \mathcal{F}^{(J_n)}$ . Nach Gleichung (2.20) können wir unseren Schätzer  $\tilde{m}_n$  darstellen durch:

$$\tilde{m}_n(x) = \sum_{j=1}^{J_n} \hat{a}_j \cdot f_j$$

für geeignete  $f_j \in \mathcal{F}$  und  $\hat{a}_j$  welche

$$\frac{c_8}{n} \sum_{j=1}^{J_n} \hat{a}_j^2 = \frac{c_8}{n} \sum_{\mathbf{i} \in [M_n]^d} \sum_{\substack{\mathbf{j} \in [q]^d \\ |\mathbf{j}|_1 \leq q}} a_{\mathbf{i}, \mathbf{j}}^2 \leq \frac{1}{n} \sum_{i=1}^n |Y_i - \tilde{m}_n(X_i)|^2 + \frac{c_8}{n} \sum_{\mathbf{i} \in [M_n]^d} \sum_{\substack{\mathbf{j} \in [q]^d \\ |\mathbf{j}|_1 \leq q}} a_{\mathbf{i}, \mathbf{j}}^2 \leq \sum_{i=1}^n Y_i^2,$$

erfüllen, wobei wir bei der letzten Ungleichung wie in Kapitel 2 die minimierende Eigenschaft von  $a_{\mathbf{i}, \mathbf{j}}$  verwendet haben und zum Schluss die Koeffizienten Null gesetzt haben. Da  $c_8 > 0$  ist, erhalten wir dass, die Koeffizienten  $\hat{a}_j$  die Eigenschaft

$$\sum_{j=1}^{J_n} \hat{a}_j^2 \leq \frac{1}{n} \sum_{i=1}^n Y_i^2 \cdot \frac{n}{c_8} \quad (3.14)$$

erfüllen müssen. Da Ungleichung 3.14 insbesondere auf dem Ereignis  $A_n$  gilt, folgt aus der Definition von  $A_n$  in Gleichung (3.4) und der Definition der Konstante  $c_7$  in Gleichung (3.12):

$$\sum_{j=1}^{J_n} \hat{a}_j^2 \leq \frac{1 + \mathbb{E}[Y^2]}{c_8} \cdot n \leq c_7 \cdot n,$$

woraus durch  $f_j \in \mathcal{F}$  schließlich  $\tilde{m}_n \in \mathcal{F}^{(J_n)}$  folgt. Als nächstes wollen wir Ungleichung (3.2) aus Lemma 3.5 für  $\tilde{m}_n$  und  $g_n$  zeigen.

Die Funktionen  $\tilde{m}_n$  und  $g_n$  unterscheiden sich in den Vorfaktoren von  $f_j$ . Die Koeffizienten  $a_{\mathbf{i}, \mathbf{j}}$  von  $\tilde{m}_n$  haben wir durch Minimierung des Funktional in Gleichung (2.17) erhalten. Nach Voraussetzung ist  $N \geq q$  und damit gilt dann insbesondere  $\{0, \dots, q\} \subseteq \{0, \dots, N\}$ . Bei der Minimierung des Funktional hat man daher auch die Koeffizienten von  $g_n$  betrachtet. Daher erhalten wir:

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n |Y_i - \tilde{m}_n(X_i)|^2 \\ & \leq \frac{1}{n} \sum_{i=1}^n |Y_i - \tilde{m}_n(X_i)|^2 + \frac{c_8}{n} \cdot \sum_{\mathbf{i} \in [M_n]^d} \sum_{\substack{\mathbf{j} \in [N]^d \\ |\mathbf{j}|_1 \leq N}} a_{\mathbf{i}, \mathbf{j}}^2 \\ & \leq \frac{1}{n} \sum_{i=1}^n |Y_i - g_n(X_i)|^2 + \frac{c_8}{n} \cdot \sum_{\mathbf{i} \in [M_n]^d} \sum_{\substack{\mathbf{j} \in [q]^d \\ |\mathbf{j}|_1 \leq q}} \left| \frac{1}{\mathbf{j}!} \cdot \partial^{\mathbf{j}} m(x_{\mathbf{i}}) \right|^2 \\ & \leq \frac{1}{n} \sum_{i=1}^n |Y_i - g_n(X_i)|^2 + c_9 \cdot \frac{(M_n + 1)^d}{n}, \end{aligned} \quad (3.15)$$

mit  $c_9 = c_8 \cdot (q+1)^d \cdot z^2$  als Konstante die unabhängig von  $n$  ist. Wir erhalten damit für  $\bar{m}_n \in \{\tilde{m}_n, g_n\} \subseteq \mathcal{F}^{(J_n)}$ :

$$\frac{1}{n} \sum_{i=1}^n |Y_i - \bar{m}_n(X_i)|^2 \leq \frac{1}{n} \sum_{i=1}^n |Y_i - g_n(X_i)|^2 + c_9 \cdot \frac{(M_n+1)^d}{n}, \quad (3.16)$$

da für  $\bar{m}_n = g_n$  die Ungleichung unmittelbar folgt. Da  $g_n$  nach Definition deterministisch, damit also unabhängig von  $(X_1, Y_1), \dots, (X_n, Y_n)$  ist, sind mit der Abschätzung (3.16) für  $\hat{m}_n = T_{\beta_n} \bar{m}_n$  mit  $\bar{m}_n \in \mathcal{F}^{(J_n)}$  und dem *Penalty Term*  $\text{pen}_n(g_n) = c_9 \cdot \frac{(M_n+1)^d}{n} > 0$  die Voraussetzungen für Lemma 3.5 erfüllt. Wir erhalten durch dessen Anwendung:

$$\begin{aligned} & \mathbb{E} \left[ \int |\hat{m}_n(x) - m(x)|^2 \mathbb{P}_X(dx) \right] \\ & \leq \frac{c_9 \cdot \log(n)^2 \cdot \left( \log \left( \sup_{x_1^n \in (\text{supp}(X))^n} \mathcal{N}_1 \left( \frac{1}{n \cdot \beta_n}, \mathcal{F}^{(J_n)}, x_1^n \right) \right) + 1 \right)}{n} \\ & \quad + 2 \cdot \mathbb{E} \left[ \int |g_n(x) - m(x)|^2 \mathbb{P}_X(dx) + c_9 \cdot \frac{(M_n+1)^d}{n} \right] \\ & = \frac{c_9 \cdot \log(n)^2 \cdot \left( \log \left( \sup_{x_1^n \in (\text{supp}(X))^n} \mathcal{N}_1 \left( \frac{1}{n \cdot \beta_n}, \mathcal{F}^{(J_n)}, x_1^n \right) \right) + 1 \right)}{n} \\ & \quad + 2 \int |g_n(x) - m(x)|^2 \mathbb{P}_X(dx) + 2 \cdot c_9 \cdot \frac{(M_n+1)^d}{n} \\ & =: T_{2,A,n} + 2 \cdot T_{2,B,n} + 2 \cdot c_9 \cdot \frac{(M_n+1)^d}{n}, \end{aligned} \quad (3.17)$$

wobei wir bei der letzten Gleichheit verwendet haben, dass der letzte Summand deterministisch ist. Zudem wissen wir, dass  $c_8$  unabhängig von  $n$  ist und  $n > 1$ , da wir  $n$  hinreichend groß wählen. Des weiteren erhalten wir, da für  $n$  hinreichend groß  $\log(n)^3 > 1$  gilt, mit analogem Vorgehen:

$$c_9 \cdot \frac{(M_n+1)^d}{n} \leq c_{10} \cdot \frac{n^{\frac{d}{2p+d}}}{n} \leq c_{10} \cdot \log(n)^3 \cdot n^{-\frac{2p}{2p+d}}, \quad (3.18)$$

mit einer von  $n$  unabhängigen Konstanten  $c_{10} := c_9 \cdot (2c_2)^d > 0$ . Mithilfe von Ungleichung (3.18) erhalten wir nun in Gleichung (3.17):

$$\mathbb{E} \left[ \int |\hat{m}_n(x) - m(x)|^2 \mathbb{P}_X(dx) \right] \leq T_{2,A,n} + 2 \cdot T_{2,B,n} + 2 \cdot c_{10} \cdot \log(n)^3 \cdot n^{-\frac{2p}{2p+d}}.$$

Wir wollen die Summanden  $T_{2,A,n}$  und  $T_{2,B,n}$  weiter abschätzen.

### 3.2.2.1 Abschätzung von $T_{2,A,n}$

Als Erstes überprüfen wir die Voraussetzungen von Lemma 3.6, um damit dann den Summanden, welcher die Überdeckungszahl  $\mathcal{N}_1$  enthält, in Ungleichung (3.17) weiter

abzuschätzen. Nach Voraussetzung ist  $\beta_n = c_1 \cdot \log(n)$  und  $a_n = (\log n)^{1/(6(N+d))} > 0$  für hinreichend großes  $n$ . Nach Voraussetzung sind zudem  $d, N, J_n \in \mathbb{N}$  und es gilt nach Gleichung (3.13):

$$J_n \leq (M_n + 1)^d \cdot (N + 1)^d \leq n^c,$$

für hinreichend großes  $n$  und Konstante  $c > 0$ . Wir betrachten hier den logistischen Squasher  $\sigma$  welcher nach Lemma 1.4 insbesondere 2-zulässig ist. Da die hier betrachtete Menge von Funktionen  $\mathcal{F}^{(J_n)}$  identisch mit der aus Lemma 3.6 ist, sind nun alle Voraussetzungen für Lemma 3.6 erfüllt. Nach Voraussetzung wissen wir, dass  $\text{supp}(\mathbb{P}_X)$  beschränkt ist und wir können  $n$  so groß wählen, dass wir ohne Beschränkung der Allgemeinheit annehmen können, dass  $\text{supp}(\mathbb{P}_X) = \{x \in \mathbb{R}^d \mid \forall \varepsilon > 0 : \mathbb{P}_X(S_\varepsilon(x)) > 0\} \subseteq [-a_n, a_n]^d$  ist, mit  $S_\varepsilon$  als  $\varepsilon$ -Umgebung um  $x \in \mathbb{R}^d$ . In der nächsten Ungleichungskette fassen wir die Konstanten die sich von  $c_9$  unterscheiden und die unabhängig von  $n$  sind, zu einer Konstante  $c$  zusammen. Wir erhalten damit durch Lemma 3.6 für hinreichend großes  $n$ :

$$\begin{aligned} T_{2,A,n} &= \frac{c_9 \cdot \log(n)^2 \cdot \left( \log \left( \sup_{x_1^n \in (\text{supp}(X))^n} \mathcal{N}_1 \left( \frac{1}{n \cdot \beta_n}, \mathcal{F}^{(J_n)}, x_1^n \right) \right) + 1 \right)}{n} \\ &\leq \frac{c_9 \cdot \log(n)^2 \cdot \left( (c \cdot \log(n) \cdot (M_n + 1)^d \cdot (N + 1)^d) + 1 \right)}{n} \\ &\leq \frac{c_9 \cdot \log(n)^2 \cdot (2 \cdot c \cdot \log(n) \cdot (M_n + 1)^d \cdot (N + 1)^d)}{n} \\ &\leq c_{11} \cdot \frac{\log(n)^3 \cdot (M_n + 1)^d \cdot (N + 1)^d}{n}, \end{aligned} \tag{3.19}$$

wobei  $c_{11} := 2c_9 \cdot c$  eine von  $n$  unabhängige Konstante ist. Durch einsetzen der Definition von  $M_n$  erhalten wir für hinreichend großes  $n$

$$(M_n + 1)^d \leq (c_2 \cdot n^{\frac{1}{2p+d}} + 2)^d \leq 2^d \cdot c_2^d \cdot n^{\frac{d}{2p+d}} = 2^d \cdot c_2^d \cdot n^{\frac{d}{2p+d}-1} = 2^d \cdot c_2^d \cdot n^{-\frac{2p}{2p+d}}$$

und damit schließlich für Ungleichung (3.19):

$$\begin{aligned} &c_{11} \cdot \frac{\log(n)^3 \cdot (M_n + 1)^d \cdot (N + 1)^d}{n} \\ &= c_{11} \cdot (N + 1)^d \cdot 2^d \cdot c_2^d \cdot \log(n)^3 \cdot n^{-\frac{2p}{2p+d}} \\ &= c_{12} \cdot \log(n)^3 \cdot n^{-\frac{2p}{2p+d}}, \end{aligned} \tag{3.20}$$

mit einer von  $n$  unabhängigen Konstante  $c_{12} := c_{11} \cdot (N + 1)^d \cdot 2^d \cdot c_2^d > 0$ . Damit haben wir den Summanden aus Ungleichung (3.17), welcher die Überdeckungszahl  $\mathcal{N}_1$  enthält, weiter abgeschätzt.

### 3.2.2.2 Abschätzung von $T_{2,B,n}$

Sei

$$P_{m,n}(x) := \sum_{\mathbf{i} \in [M_n]^d} \sum_{\substack{\mathbf{j} \in [q]^d \\ |\mathbf{j}|_1 \leq q}} \frac{1}{\mathbf{j}!} \cdot \partial^{\mathbf{j}} m(x_{\mathbf{i}}) \cdot (x - x_{\mathbf{i}})^{\mathbf{j}} \prod_{j=1}^d \left(1 - \frac{M_n}{2a_n} \cdot |x^{(j)} - x_{\mathbf{i}}^{(j)}|\right)_+,$$

wie in Gleichung (2.3) eine lokale Spline-Interpolation von Taylorpolynomen von  $m$ . Mit Ungleichung (3.6), zusammen mit einer Nulladdition und der Linearität des Integrals erhalten wir:

$$\begin{aligned} T_{2,B,n} &= \int |g_n(x) - m(x)|^2 \mathbb{P}_X(dx) \\ &= \int |g_n(x) - P_{m,n}(x) + P_{m,n}(x) - m(x)|^2 \mathbb{P}_X(dx) \\ &\leq \int 2|g_n(x) - P_{m,n}(x)|^2 + 2|P_{m,n}(x) - m(x)|^2 \mathbb{P}_X(dx). \end{aligned}$$

Aus der Supremumseigenschaft folgt

$$\begin{aligned} &\int |g_n(x) - m(x)|^2 \mathbb{P}_X(dx) \\ &\leq 2 \int \sup_{x \in [-a_n, a_n]^d} |g_n(x) - P_{m,n}(x)|^2 \mathbb{P}_X(dx) + 2 \int \sup_{x \in [-a_n, a_n]^d} |P_{m,n}(x) - m(x)|^2 \mathbb{P}_X(dx) \\ &= 2 \sup_{x \in [-a_n, a_n]^d} |g_n(x) - P_{m,n}(x)|^2 + 2 \sup_{x \in [-a_n, a_n]^d} |P_{m,n}(x) - m(x)|^2 \\ &=: 2 \sup_{x \in [-a_n, a_n]^d} T_{2,B_1,n}^2 + 2 \sup_{x \in [-a_n, a_n]^d} T_{2,B_2,n}^2, \end{aligned} \tag{3.21}$$

wobei wir im letzten Schritt  $\text{supp}(\mathbb{P}_X) \subseteq [-a_n, a_n]^d$  und  $\mathbb{P}(X \in \text{supp}(\mathbb{P}_X)) = 1$  verwendet haben. Um die letzten beiden Summanden der Ungleichung (3.21) weiter abzuschätzen, möchten wir Lemma 3.2 anwenden.

#### Abschätzung von $T_{2,B_1,n}$

Wir überprüfen, ob für Lemma 3.2 alle Voraussetzungen erfüllt sind. Wir betrachten wieder den logistischen Squasher  $\sigma$  aus Gleichung 1.2, welcher nach Lemma 1.4 insbesondere 2-zulässig ist. Zudem ist für hinreichend großes  $n$  die Bedingung

$$\begin{aligned} R_n \geq \max \left\{ \frac{\|\sigma''\|_{\infty} \cdot (M_n + 1)}{2 \cdot |\sigma'(t_{\sigma})|}, \frac{9 \cdot \|\sigma''\|_{\infty} \cdot a_n}{|\sigma'(t_{\sigma})|}, \right. \\ \left. \frac{20 \cdot \|\sigma'''\|_{\infty}}{3 \cdot |\sigma''(t_{\sigma})|} \cdot 3^{3 \cdot 3^s} \cdot a^{3 \cdot 2^s}, 1792 \cdot \frac{\max\{\|\sigma''\|_{\infty}, \|\sigma'''\|_{\infty}, 1\}}{\min\{2 \cdot |\sigma'(t_{\sigma})|, |\sigma''(t_{\sigma})|, 1\}} \cdot M_n^3 \right\} \end{aligned}$$

erfüllt. Weiterhin ist unser neuronales Netz (2.15) mit  $x_i \in [-a_n, a_n]^d$  identisch mit der Definition aus Lemma 3.2 und damit sind alle Voraussetzungen für Lemma 3.2 erfüllt. Wir erhalten damit für  $x \in [-a_n, a_n]^d$  und  $n$  hinreichend groß:

$$\begin{aligned}
 \left| f_{\text{net}, \mathbf{j}, \mathbf{i}}(x) - (x - x_{\mathbf{i}})^{\mathbf{j}} \cdot \prod_{j=1}^d \left(1 - \frac{M_n}{2a_n} \cdot |x^{(j)} - x_{\mathbf{i}}^{(j)}| \right)_+ \right| &\leq c \cdot 3^{3 \cdot 3^s} \cdot a_n^{3 \cdot 2^s} \cdot M_n^3 \cdot \frac{1}{R_n} \\
 &\leq c \cdot a_n^{3 \cdot (N+d) \cdot 2} \cdot \frac{M_n^3}{R_n} \\
 &= c \cdot \log(n) \cdot \frac{M_n^3}{R_n},
 \end{aligned} \tag{3.22}$$

wobei wir in  $c$  alle von  $n$  unabhängigen Konstanten zusammenfassen, welche insbesondere die Konstante aus Lemma 3.2 enthält. Wir haben unter anderem verwendet, dass für hinreichend großes  $n$

$$a_n^{2^{\lceil \log_2(N+d) \rceil}} \leq a_n^{2^{\log_2(N+d)+1}} = a_n^{(N+d) \cdot 2}$$

gilt. Im letzten Schritt haben wir in Ungleichung (3.22) die Definition von  $a_n$  eingesetzt. Mit Ungleichung (3.22) erhalten wir nun:

$$\begin{aligned}
 T_{2, B_1, n} &= |g_n(x) - P_{m, n}(x)| \\
 &= \left| \sum_{\mathbf{i} \in [M_n]^d} \sum_{\substack{\mathbf{j} \in [q]^d \\ |\mathbf{j}|_1 \leq q}} \frac{1}{\mathbf{j}!} \cdot \partial^{\mathbf{j}} m(x_{\mathbf{i}}) \right| \cdot \left| f_{\text{net}, \mathbf{j}, \mathbf{i}}(x) - (x - x_{\mathbf{i}})^{\mathbf{j}} \cdot \prod_{j=1}^d \left(1 - \frac{M_n}{2a_n} \cdot |x^{(j)} - x_{\mathbf{i}}^{(j)}| \right)_+ \right| \\
 &\stackrel{(3.11)}{\leq} (M_n + 1)^d \cdot (q + 1)^d \cdot z \cdot \left| f_{\text{net}, \mathbf{j}, \mathbf{i}}(x) - (x - x_{\mathbf{i}})^{\mathbf{j}} \cdot \prod_{j=1}^d \left(1 - \frac{M_n}{2a_n} \cdot |x^{(j)} - x_{\mathbf{i}}^{(j)}| \right)_+ \right| \\
 &\leq (M_n + 1)^d \cdot (q + 1)^d \cdot c \cdot \log(n) \cdot \frac{M_n^3}{R_n},
 \end{aligned} \tag{3.23}$$

wobei wir in  $c$  alle von  $n$  unabhängigen Konstanten zusammenfassen, welche insbesondere die Konstante  $c$  aus Lemma 1.4 einschließt. Durch Einsetzen der Definitionen von  $M_n$  und  $R_n$  erhalten wir für  $n$  hinreichend groß:

$$\frac{(M_n + 1)^{8d}}{R_n^2} \leq c \cdot \frac{n^{\frac{8d}{2p+d}}}{n^{2d+8}} = c \cdot n^{\frac{8d}{2p+d} - 2d - 8} \leq c \cdot n^{\frac{8d}{2p+d} - 8 \frac{2p+d}{2p+d}} = c \cdot n^{-\frac{16p}{2p+d}} = c \cdot n^{-\frac{2p}{2p+d}},$$

wobei in dieser Ungleichungskette  $c = (2c_2)^{8d}$  gilt. Bei der letzten Ungleichung haben wir verwendet, dass  $\frac{16p}{2p+d} > \frac{2p}{2p+d}$ , da  $p > 0$  ist. Damit erhalten wir:

$$\begin{aligned}
\sup_{x \in [-a_n, a_n]^d} |g_n(x) - P_{m,n}(x)|^2 &\leq \left( (M_n + 1)^d \cdot (q + 1)^d \cdot c \cdot \log(n) \cdot \frac{M_n^3}{R_n} \right)^2 \\
&\leq c^2 \cdot (M_n + 1)^{2d} \cdot \log(n)^2 \cdot \frac{M_n^6}{R_n^2} \\
&\leq c^2 \cdot (M_n + 1)^{2d} \cdot \log(n)^2 \cdot \frac{(M_n + 1)^{6d}}{R_n^2} \\
&\leq c^2 \cdot \log(n)^2 \cdot \frac{(M_n + 1)^{8d}}{R_n^2} \\
&\leq c_{13} \cdot n^{-\frac{2p}{2p+d}} \cdot \log(n)^3,
\end{aligned} \tag{3.24}$$

mit  $c_{13} = c^2 \cdot (2c_2)^{8d} > 0$  und unabhängig von  $n$  ist. Bei der letzten Ungleichung haben wir verwendet, dass  $\log(n)^2 < \log(n)^3$  für  $n$  hinreichend groß gilt. Damit haben wir  $\sup_{x \in [-a_n, a_n]^d} T_{2,B_1,n}^2$  aus Ungleichung (3.21) abgeschätzt.

#### Abschätzung von $T_{2,B_2,n}$

Da nach Konstruktion  $a_n > 0$ ,  $m$   $(p, C)$ -glatt und  $P_{m,n}(x)$  nach Lemma 2.4 eine Spline Interpolation von Taylorpolynomen von  $m$  ist, erhalten wir mit Lemma 2.5:

$$T_{2,B_2,n} = |P_{m,n}(x) - m(x)| \leq c_{14} \cdot \frac{a_n^p}{M_n^p} \leq c_{14} \cdot \log(n) \cdot \frac{1}{M_n^p}. \tag{3.25}$$

In dieser Ungleichung ist  $c_{14} > 0$  eine von  $n$  unabhängige Konstante aus Lemma 2.5 und wir haben zudem verwendet, dass:

$$a_n^p = a_n^{q+s} \leq a_n^{N+d} \leq a_n^{6 \cdot (N+d)} = \log(n),$$

für  $p = q + s$  für hinreichend großes  $n$  gilt, da nach Voraussetzung  $N \geq q$  und  $d \geq s$  mit  $s \in (0, 1]$  ist. Durch Quadrieren bleiben die Ungleichungen (3.23) und (3.25) auch erhalten und, da in beiden Ungleichungen die rechte Seite unabhängig von  $x$  ist, gelten die Ungleichungen ebenfalls für das Supremum über  $x \in [-a_n, a_n]^d$ . Für  $n$  hinreichend groß, erhalten wir durch Einsetzen der Definition von  $M_n$ :

$$\begin{aligned}
\sup_{x \in [-a_n, a_n]^d} |P_{m,n}(x) - m(x)|^2 &\leq \left( c_{14} \cdot \log(n) \cdot \frac{1}{M_n^p} \right)^2 \\
&\leq c_{14}^2 \cdot \log(n)^2 \cdot c_2^{-2p} \cdot n^{-\frac{2p}{2p+d}} \\
&\leq \left( \frac{c_{14}}{c_2^p} \right)^2 \cdot \log(n)^3 \cdot n^{-\frac{2p}{2p+d}}.
\end{aligned} \tag{3.26}$$



Hiermit haben wir  $\sup_{x \in [-a_n, a_n]^d} T_{2, B_2, n}^2$  aus Ungleichung (3.21) abgeschätzt. Dadurch wurde schließlich auch  $T_{2, B, n}$  und damit auch  $T_{2, n}$  abgeschätzt. Mit der Abschätzung von  $T_{1, n}$  haben wir nun alle Summanden von Ungleichung (3.5) abgeschätzt und erhalten schließlich:

$$\mathbb{E} \left[ \int |m_n(x) - m(x)|^2 \mathbb{P}_X(dx) \right] \leq c_{\text{fin}} \cdot \log(n)^3 \cdot n^{-\frac{2p}{2p+d}},$$

mit

$$c_{\text{fin}} = c_5 + 2 \cdot c_{10} + c_{12} + 2 \cdot \left( 2 \cdot \left( \frac{c_{14}}{c_2^p} \right)^2 + 2 \cdot c_{13} \right),$$

wobei  $c_{\text{fin}}$  als Summe nichtnegativer und positiver Konstanten, die unabhängig von  $n$  sind, nichtnegativ und unabhängig von  $n$  ist. Damit haben wir unser Hauptresultat bewiesen.  $\square$

## Kapitel 4

# Anwendungsbeispiel auf simulierte Daten

In diesem Kapitel betrachten wir die Leistung unseres Neuronale-Netze-Regressionsschätzers aus Kapitel 2, bei endlicher Stichprobengröße auf simulierte Daten in *Python* [VRDJ95, Version 3.7.3] mithilfe der *Keras* Bibliothek [C<sup>+</sup>15].

Die simulierten Daten welchen wir verwenden werden, sehen wie gefolgt aus: Wir wählen  $X$  gleichverteilt auf  $[-2, 2]^d$ , wobei  $d$  die Dimension des Inputs ist, zudem wählen wir  $\varepsilon$  als standardnormalverteilt und unabhängig von  $X$  und wir definieren  $Y$  durch:

$$Y = m_j(X) + \sigma \cdot \lambda_j \cdot \varepsilon,$$

mit  $m_j: [-2, 2]^d \rightarrow \mathbb{R}$  für  $j \in \{1, 2\}$  wie unten definiert,  $\lambda_j > 0$  als Skalierungsfaktor welcher wie unten definiert wird und einen Rauschfaktor (*noise*)  $\sigma = 0.05$ . Als Regressionsfunktionen verwenden wir die Funktionen:

$$m_1(x) = \sin(0.2 \cdot x^2) + \exp(0.5 \cdot x) + x^3$$

und

$$m_2(x_0, x_1) = \sin\left(\sqrt[2]{x_0^2 + x_1^2}\right).$$

Wir wählen  $\lambda_j$  als Interquartilsabstand einer Stichprobe von  $m(X)$  der Größe  $N = 800$ . Mit diesen Daten lässt sich nun auch  $Y$  darstellen.

Um die Leistung unseres Neuronalen-Netze-Regressionsschätzers zu überprüfen, haben wir erstmals die Funktion  $m_1$  und die jeweilige Schätzung durch unseren Neuronale-Netze-Regressionsschätzer für  $\sigma = 0.05$  zeichnen lassen. Visuelle kann man mit Abbildung 4.1 erkennen, dass der Schätzer eine sehr gute Approximation der Funktion liefert, aber um genauer beurteilen zu können wie gut die Schätzung wirklich ist und wie gut unser Schätzer im Vergleich zu anderen Schätzern abschneidet, betrachten wir in Tabelle 4.1

und Tabelle 4.2 den Interquartilsabstand (*IQR*) und den Median der skalierten  $L_2$ -Fehler der einzelnen Schätzer.

Unser Vorgehen zum Vergleich der drei hier betrachteten Regressionsschätzern gestaltet sich wie gefolgt: Wir bestimmen erst den  $L_2$ -Fehler der einzelnen Schätzer approximativ durch den empirisch arithmetischen  $L_2$ -Fehler  $\varepsilon_{L_2,N}$  auf einer unabhängigen Stichprobe von  $X$  der Größe  $N = 1000$ .

Da wir unsere Regressionsfunktionen kennen und der Fehler stark vom Verhalten der korrekten Funktion von  $m_j$  abhängt, betrachten wir den empirischen  $L_2$ -Fehler im Verhältnis zum einfachsten Schätzer von  $m_j$ , einer konstanten Funktion. Der Wert dieses *konstanten* Schätzers bestimmen wir in dem wir das empirische Mittel der beobachtete Daten  $Y$  nehmen. Wir erhalten damit einen skaliertes Fehlermaß  $\varepsilon_{L_2,N}(m_{n,i})/\bar{\varepsilon}_{L_2,N}(avg)$  mit  $\bar{\varepsilon}_{L_2,N}(avg)$  als Median von 50 unabhängigen Realisierungen von  $\varepsilon_{L_2,N}$ .

Wir bestimmten  $\varepsilon_{L_2,N}(\cdot)$  als empirisches Mittel von 25 quadratischen Fehlern der Schätzung des konstanten Schätzers. Dieses skalierte Fehlermaß ist so zu deuten, dass ein großer Fehler durch einen der drei Regressionsschätzer im Falle dass der Fehler des konstanten Schätzers klein ist, auf eine noch schlechtere Performance hindeutet. Der Fehler  $\varepsilon_{L_2,N}(m_{n,i})$  wird also durch  $\bar{\varepsilon}_{L_2,N}(avg)$  gewichtet.

Die resultierenden skalierten Fehler hängen noch von der Stichprobe von  $(X, Y)$  ab und um diese Werte besser vergleichen zu können, führen wir die Fehlerberechnung jeweils 50 mal durch und geben dann den Median und Interquartilsabstand für die Schätzung der betrachteten Regressionsschätzer aus. Wir teilen für jeden Schätzer die Stichprobe auf in ein *learning sample* der Größe  $n_l = 0.8 \cdot N$  und in ein *testing sample* der Größe  $n_t = 0.2 \cdot N$ . Wir bestimmen den Schätzer für alle Parameterwerte mit dem learning sample und bestimmen das korrespondierende  $L_2$ -Risiko auf dem testing sample und wählen dann die Parameter die zu einem minimalen empirischen  $L_2$ -Risiko auf dem testing sample führen. Unser erster Schätzer *fc\_neural\_1\_estimate* ist ein neuronales Netz mit einer verborgenen Schicht. Dieser Schätzer hat eine feste Anzahl an Neuronen die wir aus der Menge  $\{5, 10, 25, 50, 75\}$  auswählen die bei der Simulation zu einem minimalen empirischen  $L_2$ -Risiko führt.

Unser zweiter Schätzer *nearest\_neighbor\_estimate* ist ein Nächste-Nachbar-Schätzer bei der die Anzahl an nächsten Nachbarn aus der Menge  $\{1, 2, 3\} \cup \{4, 8, 12, 16, \dots, 4 \cdot \lfloor \frac{n_l}{4} \rfloor\}$  ausgewählt wird.

Unser letzter Schätzer ist der in Kapitel 2 vorgestellte Neuronale-Netze-Regressionsschätzer den wir mit *new\_neural\_network\_estimate* bezeichnen. Hier haben wir die Parameter je nachdem welche Regressionsfunktion wir betrachtet haben entsprechend angepasst. Für die Schätzung von  $m_1$  setzten wir:  $d = 1$ ,  $N = 3$ ,  $q = 2$ ,  $R = 10^6$ ,  $a = 2$ ,  $M = 2$ . Für die Schätzung von  $m_2$  wählten wir:  $d = 2$ ,  $N = 2$ ,  $q = 2$ ,  $R = 10^6$ ,  $a = 2$ ,  $M = 2$ .

	$m_1$	
<i>noise</i>	5%	10%
$\bar{\epsilon}_{L_2,N}(avg)$	13.4482362	13.3925910
<i>approach</i>	Median (IQR)	Median (IQR)
new_neural_network_estimate	<b>2.482e-05(1.612e-05)</b>	<b>6.908e-05(3.936e-05)</b>
fc_neural_1_estimate	4.384e-04(2.14e-03)	7.261e-04(4.57e-03)
nearest_neighbor_estimate	2.9527e-04(9.312e-05)	9.0864e-04(2.8952e-04)

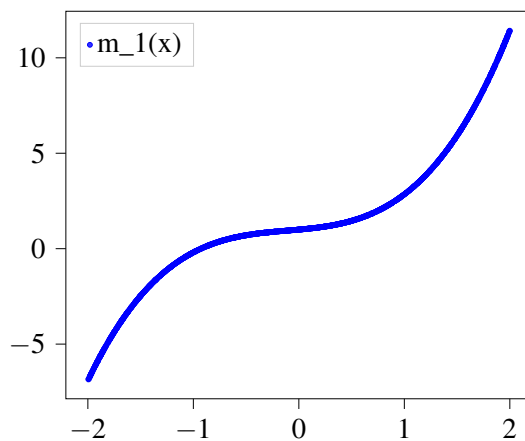
Tabelle 4.1: Median und IQR des skalierten empirischen  $L_2$ -Fehlers für Schätzungen für  $m_1$  mit Stichprobengröße  $N = 1000$ .

	$m_2$	
<i>noise</i>	5%	10%
$\bar{\epsilon}_{L_2,N}(avg)$	0.0324	0.0311
<i>approach</i>	Median (IQR)	Median (IQR)
new_neural_network_estimate	<b>0.003961(0.000932)</b>	<b>0.00431(0.000973)</b>
fc_neural_1_estimate	0.0257(0.3803)	0.0559(0.52033)
nearest_neighbor_estimate	0.01616(0.005906)	0.01763(0.007081)

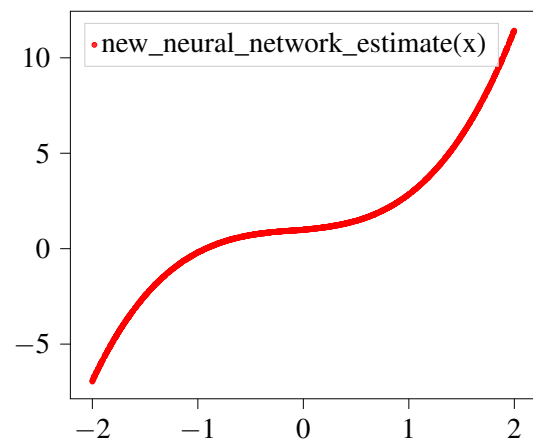
Tabelle 4.2: Median und IQR des skalierten empirischen  $L_2$ -Fehlers für Schätzungen  $m_2$  mit Stichprobengröße  $N = 1000$ .

Da wir zum Beispiel den Grad der zu schätzenden Funktion kennen und dies bei unserem Schätzer berücksichtigen können.

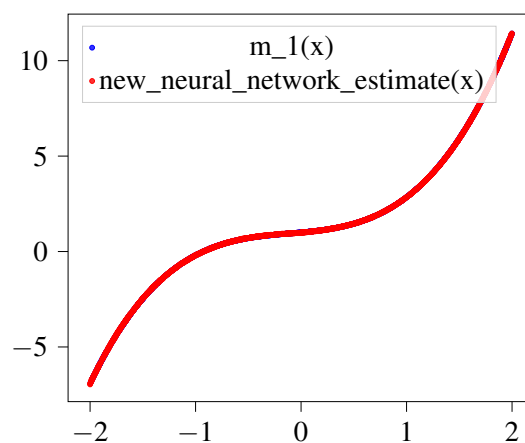
Wie wir in Tabelle 4.1 und Tabelle 4.2 anhand des skalierten  $L_2$ -Fehlers sehen können, übertrifft unserer Neuronale-Netze-Regressionsschätzer in allen Fällen die Leistung der anderen Schätzer.



(a) Zu schätzende Funktion  $m_1$ .



(b) Schätzung von  $m_1$  durch unseren Neuronale-Netze-Regressionsschätzer.



(c) Visueller Vergleich der Funktion  $m_1$  mit ihrer Schätzung.

Abbildung 4.1: Darstellung der Funktion  $m_1$  und dessen Schätzung durch den Schätzer  $\text{new\_neural\_network\_estimate}$ .

# Appendix

Der Programmcode ist wie folgt aufgebaut:

- `main.py` ist das Hauptprogramm welches alle Schätzer aufruft und die Outputs generiert.
- `data_gen.py` generiert die Daten die wir für unsere Simulation benötigen.
- `help_neural_networks.py` fasst alle Hilfsfunktionen zusammen.
- `new_neural_network.py` enthält unseren Neuronale-Netze-Regressionsschätzer.
- `fc_neural_network.py` enthält das neuronale Netz mit einer verborgenen Schicht.
- `nearest_neighbor.py` enthält einen Nächste-Nachbar-Schätzer.
- `constant.py` enthält den konstanten Schätzer.

Listing 4.1: `main.py`

```
1 """
2 Main Datei die die Simulation und damit den Vergleich der implementierten Schätzer
3 durchführt.
4 """
5 import numpy as np
6 import matplotlib.pyplot as plt
7 import pandas as pd
8 import tikzplotlib
9 from scipy.stats import iqr
10 from data_gen import gen_data_Y
11 from constant import constant_estimate
12 from new_neural_network import new_neural_network_estimate
13 from nearest_neighbor import nearest_neighbor_estimate
14 from fc_neural_network import fc_neural_l_estimate
15
16 n = 1000
17 n_train = int(n * 0.8)
18 n_test = int(n * 0.2)
19
20 '''
21 EINDIMENSIONALER FALL (d = 1) wird geplottet
```

```

22 '''
23 # Wahl der Parameter für unseren neuen Neuronale-Netze-Regressionschätzer
24
25 N = 3
26 q = 2
27 R = 10 ** 6
28 a = 2
29 M = 2
30 d = 1
31
32 sigma = 0.05
33
34 X_train = np.random.uniform(low=-2,high=2,size=(int(n_train),d))
35 m_X_train, Y_train = gen_data_Y(X_train, sigma)
36
37 X_test = np.random.uniform(low=-2,high=2,size=(int(n_test),d))
38
39 Y_pred_new_nn = new_neural_network_estimate(X_train, Y_train, X_test, N, q, R, d, M, a,)
40 m_X_test, dummy = gen_data_Y(X_test, sigma)
41
42
43 area = 4
44 plt.scatter(X_test, m_X_test, s=area, color = 'blue', label='m_1', alpha=0.5)
45 plt.scatter(X_test, Y_pred_new_nn, s=area, color = 'red', label='new_neural_network_
    estimate', alpha=0.5)
46 plt.title('...')
47 plt.legend(loc='upper left')
48 plt.xlabel('x')
49 plt.ylabel('y')
50 tikzplotlib.save("mytikz_d1.tex")
51
52 '''
53 ein Vergleich des emp. L2 Fehler gemacht für d = 1
54 '''
55 # Wahl der Parameter für unseren neuen Neuronale-Netze-Regressionschätzer
56
57 N = 3
58 q = 2
59 R = 10 ** 6
60 a = 2
61 M = 2
62 d = 1
63
64 spreads = [0.05, 0.1]
65
66 scaled_error = np.empty((50, 3,))
67 scaled_error[:] = np.nan
68
69 e_L2_avg = np.zeros(25)
70 e_L2_avg[:] = np.nan
71
72 for sigma in spreads:
73     for i in range(0, np.size(scaled_error, 0), 1):
74
75         X_train = np.random.uniform(low=-2,high=2,size=(int(n_train),d))

```

```

76     m_X_train , Y_train = gen_data_Y(X_train , sigma)
77
78     X_test = np.random.uniform(low=-2,high=2,size=(int(n_test),d))
79
80     Y_pred_new_nn = new_neural_network_estimate(X_train , Y_train , X_test , N, q, R, d
81         , M, a,)
82     Y_pred_fc_nn_1 = fc_neural_1_estimate(X_train , Y_train , X_test)
83     Y_pred_nearest_neighbor = nearest_neighbor_estimate(X_train , Y_train , X_test)
84
85
86     m_X_test , not_needed = gen_data_Y(X_test , sigma)
87
88     e_L2_new_nn = np.mean(abs(Y_pred_new_nn - m_X_test) ** 2)
89     e_L2_fc_nn_1 = np.mean(abs(Y_pred_fc_nn_1 - m_X_test) ** 2)
90     e_L2_nearest_neighbor = np.mean(abs(Y_pred_nearest_neighbor - m_X_test) ** 2)
91
92     for j in range(0,np.size(e_L2_avg),1):
93
94         X = np.random.uniform(low=-2,high=2,size=(n_test,d))
95         m_X, Y = gen_data_Y(X,sigma)
96         Y_pred_constant = constant_estimate(Y)
97
98         e_L2_avg[j] = np.mean(abs(Y_pred_constant - m_X) ** 2)
99
100     scaled_error[i,0] = e_L2_new_nn / np.median(e_L2_avg)
101     scaled_error[i,1] = e_L2_fc_nn_1 / np.median(e_L2_avg)
102     scaled_error[i,2] = e_L2_nearest_neighbor / np.median(e_L2_avg)
103
104     iqr_new_nn = iqr(scaled_error[:,0])
105     iqr_fc_nn_1 = iqr(scaled_error[:,1])
106     iqr_nearest_neighbor = iqr(scaled_error[:,2])
107
108     median_new_nn = np.median(scaled_error[:,0])
109     median_fc_nn_1 = np.median(scaled_error[:,1])
110     median_nearest_neighbor = np.median(scaled_error[:,2])
111
112     rows = [ "noise", "e_L2_avg", "approach", "new_nn", "fc_nn_1", "nearest_neighbor" ]
113
114     if sigma == 0.05:
115         series_noise_1 = pd.Series([repr(sigma)+'%', np.median(e_L2_avg), "(Median, IQR)"
116             ,(median_new_nn, iqr_new_nn), (median_fc_nn_1, iqr_fc_nn_1), (median_nearest_
117             _neighbor, iqr_nearest_neighbor)], index=rows)
118         series_noise_1.name = ""
119         print("Der empirische L2 Fehler für d = 1 und sigma = 0.05 ist berechnet worden!"
120             ")
121
122     else:
123         series_noise_2 = pd.Series([repr(sigma), np.median(e_L2_avg), "(Median, IQR)", (
124             median_new_nn, iqr_new_nn), (median_fc_nn_1, iqr_fc_nn_1), (median_nearest_
125             _neighbor, iqr_nearest_neighbor)], index=rows)
126         series_noise_2.name = ""
127         print("Der empirische L2 Fehler für d = 1 und sigma = 0.1 ist berechnet worden!"
128             ")
129
130 error_df = pd.concat([series_noise_1, series_noise_2], axis=1)

```



```

124 error_df.to_csv('out_d_1.csv', index = True)
125
126 '''
127 ein Vergleich des emp. L2 Fehler gemacht für d = 2
128 '''
129 # Wahl der Parameter für unseren neuen Neuronale-Netze-Regressionschätzer
130
131 N = 2
132 q = 2
133 R = 10 ** 6
134 a = 2
135 M = 2
136 d = 2
137
138 spreads = [0.05, 0.1]
139
140 scaled_error = np.empty((50, 3,))
141 scaled_error[:] = np.nan
142
143 e_L2_avg = np.zeros(25)
144 e_L2_avg[:] = np.nan
145
146 for sigma in spreads:
147     for i in range(0, np.size(scaled_error, 0), 1):
148
149         X_train = np.random.uniform(low=-2, high=2, size=(int(n_train), d))
150         m_X_train, Y_train = gen_data_Y(X_train, sigma)
151
152         X_test = np.random.uniform(low=-2, high=2, size=(int(n_test), d))
153
154         Y_pred_new_nn = new_neural_network_estimate(X_train, Y_train, X_test, N, q, R, d
155             , M, a,)
156         Y_pred_fc_nn_1 = fc_neural_1_estimate(X_train, Y_train, X_test)
157         Y_pred_nearest_neighbor = nearest_neighbor_estimate(X_train, Y_train, X_test)
158
159         m_X_test, not_needed = gen_data_Y(X_test, sigma)
160
161         e_L2_new_nn = np.mean(abs(Y_pred_new_nn - m_X_test) ** 2)
162         e_L2_fc_nn_1 = np.mean(abs(Y_pred_fc_nn_1 - m_X_test) ** 2)
163         e_L2_nearest_neighbor = np.mean(abs(Y_pred_nearest_neighbor - m_X_test) ** 2)
164
165         for j in range(0, np.size(e_L2_avg), 1):
166
167             X = np.random.uniform(low=-2, high=2, size=(n_test, d))
168             m_X, Y = gen_data_Y(X, sigma)
169             Y_pred_constant = constant_estimate(Y)
170
171             e_L2_avg[j] = np.mean(abs(Y_pred_constant - m_X) ** 2)
172
173             scaled_error[i, 0] = e_L2_new_nn / np.median(e_L2_avg)
174             scaled_error[i, 1] = e_L2_fc_nn_1 / np.median(e_L2_avg)
175             scaled_error[i, 2] = e_L2_nearest_neighbor / np.median(e_L2_avg)
176
177 iqr_new_nn = iqr(scaled_error[:, 0])
178 iqr_fc_nn_1 = iqr(scaled_error[:, 1])

```

```

178     iqr_nearest_neighbor = iqr(scaled_error[:,2])
179
180     median_new_nn = np.median(scaled_error[:,0])
181     median_fc_nn_1 = np.median(scaled_error[:,1])
182     median_nearest_neighbor = np.median(scaled_error[:,2])
183
184     rows = ["noise", "e_L2_avg", "approach", "new_nn", "fc_nn_1", "nearest_neighbor"]
185
186     if sigma == 0.05:
187         series_noise_1 = pd.Series([repr(sigma), np.median(e_L2_avg), "(Median, IQR)", (
188             median_new_nn, iqr_new_nn), (median_fc_nn_1, iqr_fc_nn_1), (median_nearest_
189             neighbor, iqr_nearest_neighbor)], index=rows)
190         series_noise_1.name = ""
191         print("Der empirische L2 Fehler für d = 2 und sigma = 0.05 ist berechnet worden!")
192
193     else:
194         series_noise_2 = pd.Series([repr(sigma)+'%', np.median(e_L2_avg), "(Median, IQR)"
195             , (median_new_nn, iqr_new_nn), (median_fc_nn_1, iqr_fc_nn_1), (median_nearest_
196             _neighbor, iqr_nearest_neighbor)], index=rows)
197         series_noise_2.name = ""
198         print("Der empirische L2 Fehler für d = 2 und sigma = 0.1 ist berechnet worden!")
199
200 error_df = pd.concat([series_noise_1, series_noise_2], axis=1)
201 error_df.to_csv('out_d_2.csv', index = True)

```

#### Listing 4.2: data\_gen.py

```

1  """
2  Generieren der Daten die wir für einen Vergleich von Regressionsschätzern benötigen
3  """
4  # Wir wählen x gleichverteilt auf [-2,2]^d, wobei d die Dimension des Inputs ist
5  # n ist die Größe der Stichprobe
6
7  import numpy as np
8  from scipy.stats import iqr
9
10 # Regressionsfunktionen
11 #
12 # x: Ein Vektor x der Dimension d
13 # d: Dimension des Vektors x
14
15 def m_d(x, d):
16
17     sin = np.sin
18     exp = np.exp
19
20     if d == 1:
21         return sin(0.2 * x[0] ** 2) + exp(0.5 * x[0]) + x[0] ** 3
22
23     elif d == 2:
24         return np.sin(np.sqrt(x[0] ** 2 + x[1] ** 2))
25
26     else:
27         print("Your data has the wrong dimension!")

```

```

28
29 # Generiert den Vektor  $Y_1, \dots, Y_n$  für den Datensatz  $(X_1, Y_1), \dots, (X_n, Y_n)$ 
30 #
31 #  $X$ : Inputdaten der Form  $(X_1, \dots, X_n)$ , wobei  $X_i \in \mathbb{R}^d$  für  $i = 1, \dots, n$ 
32 #  $\sigma$ : Streuungsfaktor  $\in \{0.05, 0.1\}$ 
33
34 def gen_data_Y (X, sigma):
35
36     n = np.size(X, 0)
37     d = np.size(X, 1)
38
39     m_X = np.zeros((n,1))
40     m_X[:] = np.nan
41
42     S = np.random.standard_normal(size=(n,1))
43     for t in range(0,n):
44         m_X[t] = m_d(X[t], d)
45
46     Y = m_X + sigma * iqr(m_X) * S
47     return (m_X, Y)

```

Listing 4.3: help\_neural\_networks.py

```

1 """
2 Implementation von Neuronale-Netzen welche wir für die Konstruktion unseres
3 Neuronale-Netze-Regressionschätzers benötigen
4 """
5 import numpy as np
6
7 # Sigmoidfunktion
8 #
9 #  $x$ :  $x \in \mathbb{R}$ 
10
11 def sigmoid (x):
12
13     return 1 / (1 + np.exp(-x))
14
15 # Neuronales Netz welches die Funktion  $f(x) = x$  approximiert
16 #
17 #  $x$ : reelle Zahl
18 #  $R$ : reelle Zahl  $\geq 1$ 
19
20 def f_id (x, R):
21
22     return 4 * R * sigmoid(x / R) - 2 * R
23
24 # Neuronales Netz welches die Funktion  $f(x, y) = x * y$  approximiert
25 #
26 #  $x$ : reelle Zahl
27 #  $y$ : reelle Zahl
28 #  $R$ : reelle Zahl  $\geq 1$ 
29
30 def f_mult (x, y, R):
31
32     return (((R ** 2) / 4) * (((1 + np.exp(-1)) ** 3) / (np.exp(-2) - np.exp(-1)))) \
33         * (sigmoid(((2 * (x + y)) / R) + 1) - 2 * sigmoid(((x + y) / R) + 1) \

```

```

34     - sigmoid(((2 * (x - y)) / R) + 1) + 2 * sigmoid(((x - y) / R) + 1))
35
36 # Neuronales Netz welches die Funktion f(x) = max(x,0) approximiert
37 #
38 # x: reelle Zahl
39 # R: reelle Zahl >= 1
40
41 def f_relu (x, R):
42
43     return f_mult(f_id(x, R), sigmoid(R * x), R)
44
45 # Neuronales Netz welches die Funktion f(x) = max(1 - (M/(2 * a)) * abs(x - y), 0)
    approximiert
46 #
47 # x: reelle Zahl
48 # y: fixe reelle Zahl
49 # R: reelle Zahl >= 1
50 # M: fixe natürliche Zahl
51 # a: fixe Zahl > 0
52
53 def f_hat (x, y, R, M, a):
54
55     return f_relu((M / (2 * a)) * (x - y) + 1, R) - 2 * f_relu((M / (2 * a)) * (x - y),
        R) + 2 * f_relu((M / (2 * a)) * (x - y) - 1, R)

```

#### Listing 4.4: new\_neural\_network.py

```

1  """
2  Implementation unseres Neuronale-Netze-Regressionschätzers fully connected neuronalen
    Netzes mit einer verborgenen Schicht.
3  """
4  import scipy.special
5  import numpy as np
6  import itertools
7  from help_neural_networks import f_id, f_mult, f_hat
8  import math
9
10 # Neuronales Netz welches die Funktion f(x) = (x^(1)- x_ik^(1))^j1 * ... *
11 # (x^(d) - x_ik^(d))^jd * \prod_{j = 1}^d max((1 - (M/2a) * abs(x^(j) - x_ik^(j))), 0)
12 #
13 # x: Eingabevektor für das neuronale Netz x \in [-a,a]^d
14 # d: Ist die Dimension des Eingabevektors d > 0
15 # j_1_d: Ist ein d-dimensionaler Vektor j_1, ..., j_d \in {0,1,...,N}
16 # X_i: Ist eine d x (M+1)^d Matrix.
17 # N: Natürliche Zahl >= q
18 # q:
19 # s: [log_2(N + d)]
20 # R: Zahl >= 1
21 # M: M \in N
22 # a: > 0
23
24 def f_net (x, d, j_1_d, X_i, N, q, s, R, M, a):
25     #initialize f_l_k
26
27     f_l_k = np.empty((s + 1, (2 ** s) + 1,))
28     f_l_k[:] = np.nan

```

```

29
30     # Rekursive Definition des neuronalen Netzes f_net nach Kapitel 2
31
32     for k in range(np.sum(j_1_d) + d + 1, (2 ** s) + 1, 1):
33         f_l_k[s, k] = 1
34
35     for k in range(1, d + 1, 1):
36         f_l_k[s, np.sum(j_1_d) + k] = f_hat(x[k - 1], X_i[k - 1], R, M, a)
37
38     for l in range(1, d + 1, 1):
39         k = j_1_d[range(0, l - 1, 1)].sum() + 1
40         while k in range(j_1_d[range(0, l - 1, 1)].sum() + 1, j_1_d[range(0, l, 1)].sum() + 1, 1):
41             f_l_k[s, k] = f_id(f_id(x[l - 1] - X_i[l - 1], R), R)
42             k += 1
43
44     for l in range(s - 1, -1, -1):
45         for k in range((2 ** l), 0, -1):
46             f_l_k[l, k] = f_mult(f_l_k[l + 1, (2 * k) - 1], f_l_k[l + 1, 2 * k], R)
47
48     return f_l_k[0, 1]
49
50 # Bestimmung der Gewichte der Ausgabeschicht durch lösen eines regularisierten
51 # Kleinste-Quadrate Problems
52 #
53 # X: Eingabevektoren der Form (X_1, ..., X_n) für das neuronale Netz aus dem Datensatz (X
54 #   _1, Y_1), ..., (X_n, Y_n)
55 # Y: Eingabevektoren der Form (Y_1, ..., Y_n) für das neuronale Netz aus dem Datensatz (X
56 #   _1, Y_1), ..., (X_n, Y_n)
57 # N: Natürliche Zahl >= q
58 # q:
59 # R: Zahl >= 1
60 # d: Ist die Dimension des Eingabevektors d > 0
61 # M: M \in \mathbb{N}
62 # a: > 0
63
64 def output_weights(X, Y, N, q, R, d, M, a):
65
66     s = math.ceil(math.log2(N + d))
67
68     # Anzahl der Eingabevektoren X_1, ..., X_n
69
70     n = np.size(X, 0)
71
72     # Eine beliebige constante > 0
73
74     c_3 = 0.01
75
76     # Anzahl der Spalten der Matrix für das Kleinste-Quadrate Problem
77     # In den Spalten sind die Funktionswerte von f_net eingespeichert
78
79     J = int(((1 + M) ** d) * scipy.special.binom(N + d, d))
80
81     # Für die Konstruktion der Matrix brauchen wir erstmal alle Inputparameter
82     # für f_net, da wir dort nur den Funktionswert für einen Vektor j_1, ..., j_d

```

```

    einsetzen
81 # müssen wir erstmals alle möglichen Vektoren dieser Art konstruieren die die
    Bedingung  $0 \leq j_1 + \dots + j_d \leq N$  erfüllen
82 #  $X_{ik}$  hat in den Zeilen die Vektoren  $X_i$  aus dem Paper
83
84  $X_{ik} = \text{np.transpose}(\text{np.empty}((d, (1 + M) ** d)))$ 
85  $X_{ik}[:,] = \text{np.nan}$ 
86
87  $I_k = \text{np.array}(\text{list}(\text{itertools.product}(\text{range}(0, M + 1), \text{repeat} = d)))$ 
88  $X_{ik}[:,] = (I_k[:,] * ((2 * a) / M)) - a$ 
89
90  $\text{all\_j1\_jd} = \text{np.array}(\text{list}(\text{itertools.product}(\text{range}(0, N + 1), \text{repeat} = d)))$ 
91  $\text{all\_j1\_jd\_by\_cond} = \text{all\_j1\_jd}[\text{all\_j1\_jd.sum}(\text{axis}=1) \leq N]$ 
92
93  $B = \text{np.empty}((n, J))$ 
94  $B[:,] = \text{np.nan}$ 
95
96 for i in range(0, n):
97     j = 0
98     for k in range(0, ((M + 1) ** d)):
99         for z in range(0, int(scipy.special.binom(N + d, d))):
100              $B[i, j] = f_{\text{net}}(X[i], d, \text{all\_j1\_jd\_by\_cond}[z], X_{ik}[k], N, q, s, R, M, a)$ 
101             j += 1
102
103  $\text{weights} = \text{np.linalg.solve}(\text{np.dot}(\text{np.transpose}(B), B) + (c\_3) * \text{np.identity}(J), \text{np.dot}$ 
     $(\text{np.transpose}(B), Y))$ 
104
105 return (weights, J, all_j1_jd_by_cond, X_ik)
106
107 # Bestimmung des Funktionswerts des Neuronale-Netze-Regressionsschätzers.
108 #
109 # x: Eingabe für einen Vektor der Form  $[-a, a]^d$  für den eine Schätzung bestimmt werden
    soll
110 # X: Eingabevektoren der Form  $(X_1, \dots, X_n)$  für das neuronale Netz aus dem Datensatz  $(X$ 
     $_1, Y_1), \dots, (X_n, Y_n)$ 
111 # Y: Eingabevektoren der Form  $(Y_1, \dots, Y_n)$  für das neuronale Netz aus dem Datensatz  $(X$ 
     $_1, Y_1), \dots, (X_n, Y_n)$ 
112 # N: Natürliche Zahl  $\geq q$ 
113 # q:
114 # s:  $\lceil \log_2(N + d) \rceil$ 
115 # R: Zahl  $\geq 1$ 
116 # d: Ist die Dimension des Eingabevektors  $d > 0$ 
117 # M:  $M \in \mathbb{N}$ 
118 # a:  $> 0$ 
119
120 def new_neural_network_estimate(X_train, Y_train, X_test, N, q, R, d, M, a):
121
122      $Y_{\text{pred}} = \text{np.empty}((\text{len}(X_{\text{test}}), 1))$ 
123      $Y_{\text{pred}}[:,] = \text{np.nan}$ 
124
125      $s = \text{math.ceil}(\text{math.log2}(N + d))$ 
126
127     weights, J, all_j1_jd_by_cond, X_ik = output_weights(X_train, Y_train, N, q, R, d, M
        , a)
128

```

```

129 F_net = np.empty((1, J,))
130 F_net[:] = np.nan
131
132 for u in range(0, len(X_test), 1):
133     j = 0
134     while j < J:
135         for k in range(0, ((M + 1) ** d)):
136             for z in range(0, int(scipy.special.binom(N + d, d))):
137                 F_net[0,j] = f_net(X_test[u], d, all_jl_jd_by_cond[z], X_ik[k], N, q
                                     , s, R, M, a)
138                 j += 1
139
140 Y_pred[u] = np.sum(np.transpose(weights) * F_net)
141
142 return Y_pred

```

Listing 4.5: fc\_neural\_network.py

```

1 """
2 Implementation eines fully connected neuronalen Netzes mit einer verborgenen Schicht.
3 """
4 import numpy as np
5 from keras.models import Sequential
6 from keras.layers import Dense
7
8 # Fully connected neuronales Netz mit einer verborgenen Schicht welches die
9 # Anzahl der Neuronen adaptiv, durch Minimierung des L2-Fehlers, aus der Menge \{5, 10,
10 # 25, 50, 75\} wählt.
11 #
12 # X: Eingabevektor der Form (X_1,...,X_n) für das neuronale Netz aus dem Datensatz (X_1,
13 # Y_1), ..., (X_n, Y_n)
14 # Y: Eingabevektor der Form (Y_1,...,Y_n) für das neuronale Netz aus dem Datensatz (X_1,
15 # Y_1), ..., (X_n, Y_n)
16
17 def fc_neural_1_estimate (X_train, Y_train, X_test):
18
19     Ynew = np.empty((len(X_train), len([5,10,25,50,75]),))
20     Ynew[:] = np.nan
21
22     count = 0
23     n_neurons = [5,10,25,50,75]
24
25     d = np.size(X_train, 1)
26
27     for j in n_neurons:
28         model = Sequential()
29         model.add(Dense(j, input_dim=d, activation='relu'))
30         model.add(Dense(1, activation='linear'))
31         model.compile(loss='mse', optimizer='adam')
32         model.fit(X_train, Y_train, epochs=1000, verbose=0)
33
34         Ynew[:,count] = model.predict(X_train)[:,:0]
35         count += 1
36
37     Diff = Ynew[:] - Y_train[:]
38     best_n_neurons = n_neurons [(1/len(X_train) *(Diff.sum(axis=0) ** 2)).argmin()]

```

```

36
37     model = Sequential()
38     model.add(Dense(best_n_neurons, input_dim=d, activation='relu'))
39     model.add(Dense(1, activation='linear'))
40     model.compile(loss='mse', optimizer='adam')
41     model.fit(X_train, Y_train, epochs=1000, verbose=1)
42
43     return model.predict(X_test)

```

#### Listing 4.6: nearest\_neighbor.py

```

1  """
2  Implementation eines Nächste-Nachbarn-Schätzer
3  """
4
5  from sklearn import neighbors
6  from sklearn.model_selection import GridSearchCV
7  import warnings
8
9  warnings.simplefilter(action='ignore', category=FutureWarning)
10
11 # Implementierung des k-Nächste-Nachbarn-Schätzer. Dieser bestimmt auch selber bei einer
12   Liste von Anzahlen an Nachbarn die betrachtet werden
13 # sollen welches die beste Wahl ist. Dieser gibt die Schätzung für X_test aus.
14 #
15 # X_train: Inputvektor für das Training des Schätzers
16 # Y_train: Inputvektor für das Training des Schätzers
17 # X_test: Inputvektor der geschätzt werden soll
18
19 def nearest_neighbor_estimate (X_train, Y_train, X_test):
20
21     params = {'n_neighbors':[2,3,4,5,6,7,8,9], 'weights': ['uniform', 'distance']}
22
23     knn = neighbors.KNeighborsRegressor()
24
25     knn_gridsearch_model = GridSearchCV(knn, params, cv=5)
26     knn_gridsearch_model.fit(X_train, Y_train)
27
28     return knn_gridsearch_model.predict(X_test)

```

#### Listing 4.7: constant.py

```

1  """
2  Implementation des konstanten Schätzers.
3  """
4
5  import numpy as np
6  from scipy import mean
7
8  # Gibt den Mittelwert der Funktionswerte einer Funktion als Schätzung zurück
9  #
10 # Y: Datensatz der Form (Y_1,...) wobei Y_i \in \R für i = 1,...
11
12 def constant_estimate(Y):
13     m = np.zeros((len(Y),1))
14     m[:] = mean(Y)
15     return m

```



# Literaturverzeichnis

- [AB19] A. BRAUN, M. KOHLER UND A. KRZYZAK: *Analysis of the rate of convergence of neural network regression estimates which are easy to implement*. 2019.
- [Bar15] BARTSCH, R.: *Allgemeine Topologie*. De Gruyter, Berlin, 2. Auflage, 2015.
- [Bos13] BOSCH, S.: *Algebra*. Springer-Lehrbuch. Springer, Berlin, 8. Auflage, 2013.
- [C<sup>+</sup>15] CHOLLET, FRANÇOIS et al.: *Keras*. <https://keras.io>, 2015.
- [For17] FORSTER, O.: *Analysis I - Differential- und Integralrechnung einer Veränderlichen*. Springer, Berlin, 11. Auflage, 2017.
- [Kre98] KRESS, R.: *Numerical analysis*. Springer, Berlin, 1. Auflage, 1998.
- [Kö04] KÖNIGSBERGER, K.: *Analysis 2*. Springer, Berlin, 5. Auflage, 2004.
- [LD80] L. DEVROYE, T.J. WAGNER: *Distribution-free consistency results in non-parametric discrimination and regression function estimation*. *Annals of Statistic*, 8:231–239, 1980.
- [LD96] L. DEVROYE, L. GYÖRFI, G. LUGOSI: *A Probabilistic Theory of Pattern Recognition*. Springer, New York, 1. Auflage, 1996.
- [LG02] L. GYÖRFI, M. KOHLER, A. KRZYZAK UND H. WALK: *A Distribution-Free Theory of Nonparametric Regression*. Springer series in statistics. Springer, Berlin, 1. Auflage, 2002.
- [VRDJ95] VAN ROSSUM, GUIDO und FRED L DRAKE JR: *Python tutorial*. Centrum voor Wiskunde en Informatica Amsterdam, The Netherlands, 1995.