

Quantitative Analyse neuronaler Netze als universelle Funktionsapproximatoren

Stephan Trenn

Diplomarbeit - Verteidigung

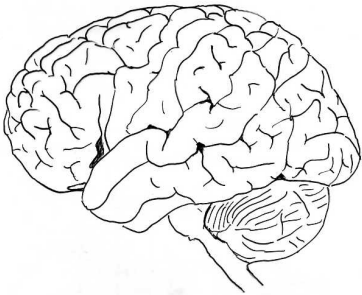
Ilmenau, 29. September 2006

Gliederung

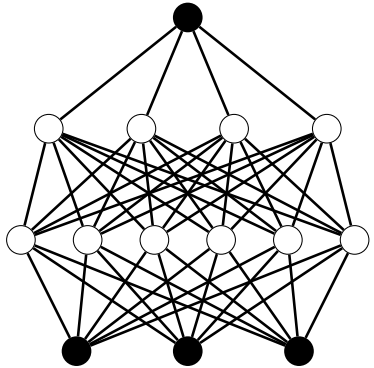


- 1 Bedeutung künstlicher neuronaler Netze (KNNs)
- 2 Approximationsgüte und Approximationsordnung
- 3 Anzahl der Koeffizienten des Taylorpolynoms
- 4 Anzahl der Netzwerkparameter
- 5 Hauptergebnis: Explizite Formel für notwendige Größe

Künstlichen neuronalen Netze (KNNs)



Natürliche neuronale Netze



Künstliche neuronale Netze

Anwendung künstlicher neuronaler Netze



Mustererkennung

- Texterkennung
- Spracherkennung
- Medizinische Diagnose
- Visuelle Qualitätsprüfung

Anwendung künstlicher neuronaler Netze



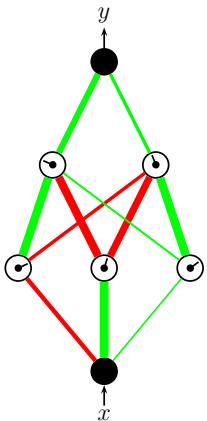
Mustererkennung

- Texterkennung
- Spracherkennung
- Medizinische Diagnose
- Visuelle Qualitätsprüfung

Funktionsapproximation

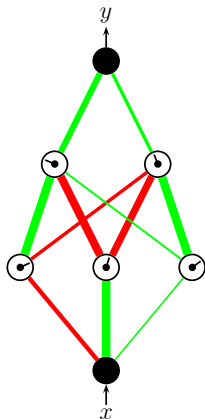
- *Mustererkennung* (Identifikatorfunktion)
- Modellierung
- Steuerung und Regelung

Netzwerkparameter und Netzwerkfunktion

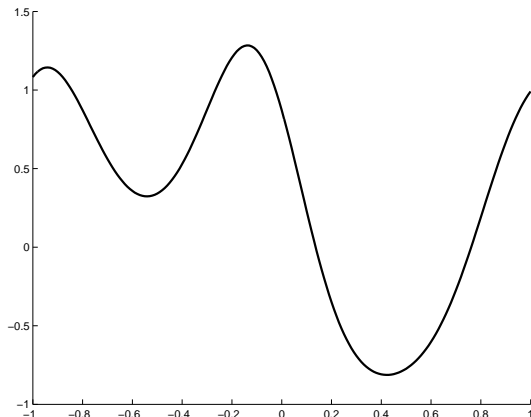


Parameter

Netzwerkparameter und Netzwerkfunktion

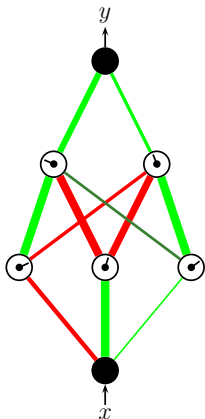


Parameter

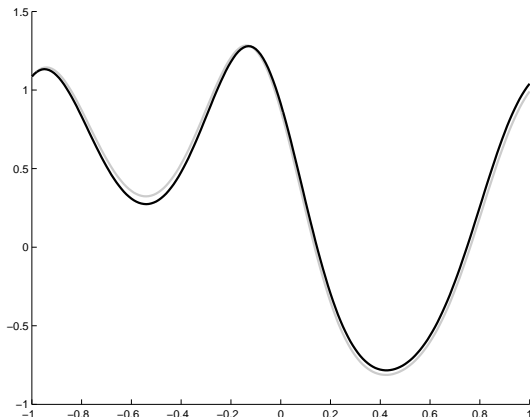


Netzwerkfunktion

Netzwerkparameter und Netzwerkfunktion

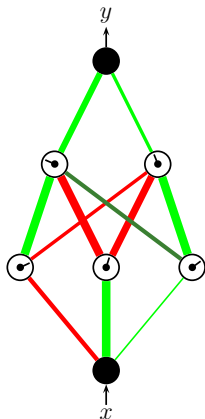


Parameter

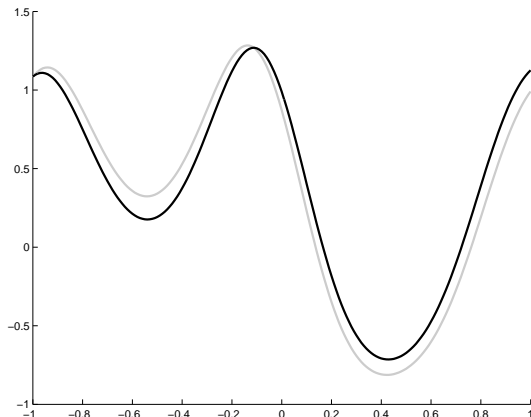


Netzwerkfunktion

Netzwerkparameter und Netzwerkfunktion

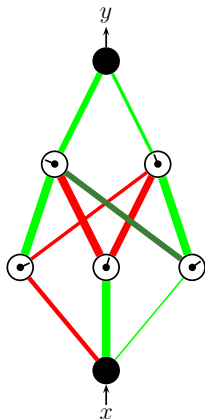


Parameter

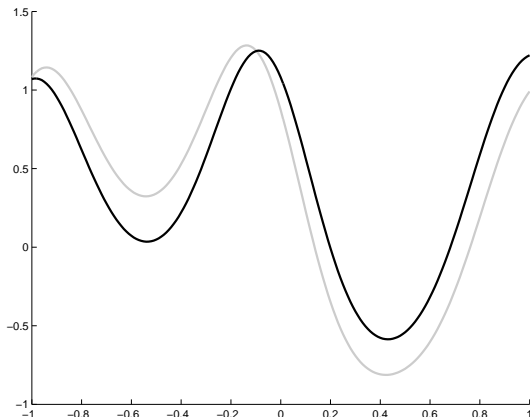


Netzwerkfunktion

Netzwerkparameter und Netzwerkfunktion

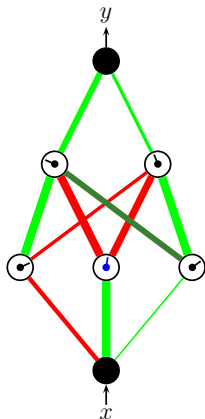


Parameter

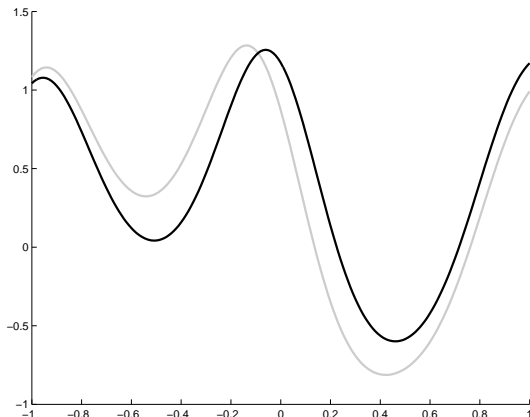


Netzwerkfunktion

Netzwerkparameter und Netzwerkfunktion

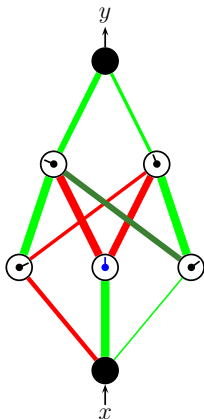


Parameter

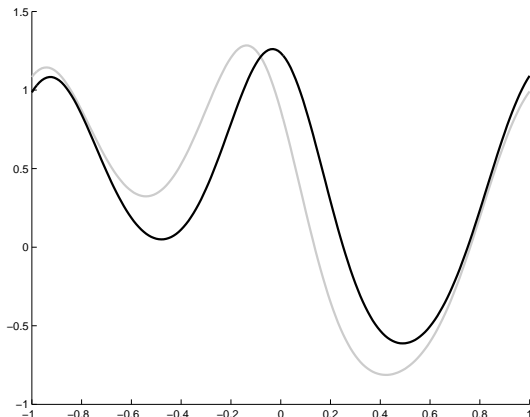


Netzwerkfunktion

Netzwerkparameter und Netzwerkfunktion

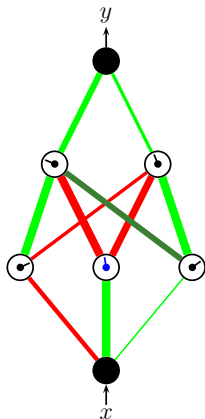


Parameter

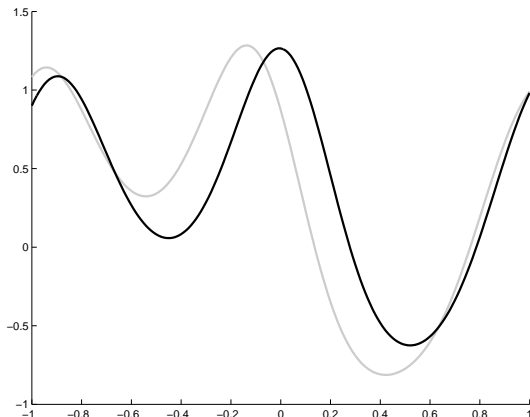


Netzwerkfunktion

Netzwerkparameter und Netzwerkfunktion



Parameter



Netzwerkfunktion

Gliederung



- 1 Bedeutung künstlicher neuronaler Netze (KNNs)
- 2 Approximationsgüte und Approximationsordnung
- 3 Anzahl der Koeffizienten des Taylorpolynoms
- 4 Anzahl der Netzwerkparameter
- 5 Hauptergebnis: Explizite Formel für notwendige Größe

Allgemeine Zielstellung



Allgemeine Zielstellung

Finde Parameter so, dass Netzwerkfunktion möglichst nah an einer Zielfunktion.

Allgemeine Zielstellung



Allgemeine Zielstellung

Finde Parameter so, dass Netzwerkfunktion möglichst nah an einer Zielfunktion.

Lernen

- Anpassung der Parameter anhand von Beispieldaten
- Backpropagation-Algorithmus
- **Nicht** Schwerpunkt der Diplomarbeit

Allgemeine Zielstellung



Allgemeine Zielstellung

Finde Parameter so, dass Netzwerkfunktion möglichst nah an einer Zielfunktion.

Lernen

- Anpassung der Parameter anhand von Beispieldaten
- Backpropagation-Algorithmus
- **Nicht** Schwerpunkt der Diplomarbeit

Welche Netzwerkgröße?

- Lernen ändert nur Parameter, nicht die Struktur
- Problem: Ist Struktur reichhaltig genug?

Wie gut ist eine Approximation?



Definition (Approximationsgüte)

$f: [-1, 1]^{n_0} \rightarrow \mathbb{R}$ approximiert $g: [-1, 1]^{n_0} \rightarrow \mathbb{R}$ mit Güte $\varepsilon > 0$

$:\Leftrightarrow$

$$\max_{\mathbf{x}} |f(\mathbf{x}) - g(\mathbf{x})| < \varepsilon$$

Wie gut ist eine Approximation?

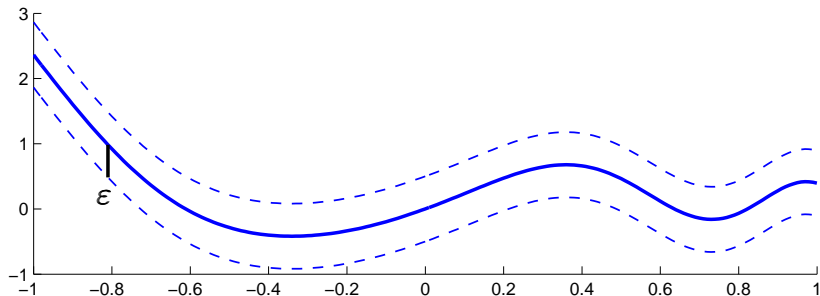


Definition (Approximationsgüte)

$f: [-1, 1]^{n_0} \rightarrow \mathbb{R}$ approximiert $g: [-1, 1]^{n_0} \rightarrow \mathbb{R}$ mit Güte $\varepsilon > 0$

$:\Leftrightarrow$

$$\max_{\mathbf{x}} |f(\mathbf{x}) - g(\mathbf{x})| < \varepsilon$$



Wie gut ist eine Approximation?

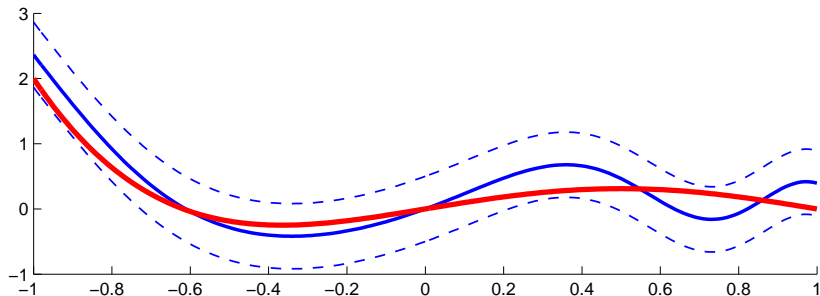


Definition (Approximationsgüte)

$f: [-1, 1]^{n_0} \rightarrow \mathbb{R}$ approximiert $g: [-1, 1]^{n_0} \rightarrow \mathbb{R}$ mit Güte $\varepsilon > 0$

$:\Leftrightarrow$

$$\max_{\mathbf{x}} |f(\mathbf{x}) - g(\mathbf{x})| < \varepsilon$$



Lokale Approximation



Definition (Approximationsordnung)

$f : [-1, 1]^{n_0} \rightarrow \mathbb{R}$ approximiert $g : [-1, 1]^{n_0} \rightarrow \mathbb{R}$ mit Ordnung $N > 0$

$:\Leftrightarrow$

$$f(0) = g(0), \quad f'(0) = g'(0), \quad \dots, \quad f^{(N)}(0) = g^{(N)}(0)$$

Lokale Approximation



Definition (Approximationsordnung)

$f: [-1, 1]^{n_0} \rightarrow \mathbb{R}$ approximiert $g: [-1, 1]^{n_0} \rightarrow \mathbb{R}$ mit Ordnung $N > 0$

$:\Leftrightarrow$

$$f(0) = g(0), \quad f'(0) = g'(0), \quad \dots, \quad f^{(N)}(0) = g^{(N)}(0)$$

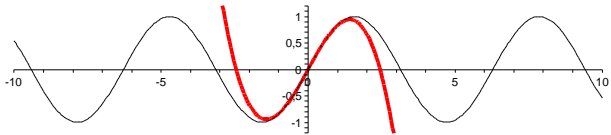
Bemerkung

Erste bis N -te Ableitung von f an der Stelle 0

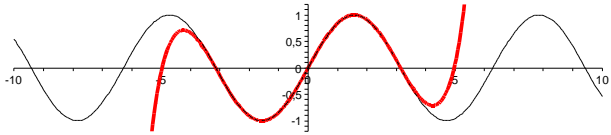
\Leftrightarrow

Taylorpolynom N -ten Grades von f an der Stelle 0

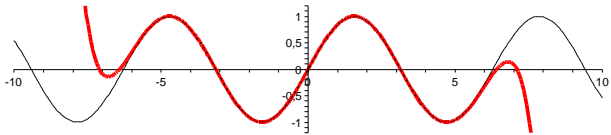
Taylorpolynome



$N = 3$

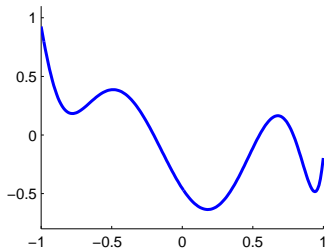


$N = 5$

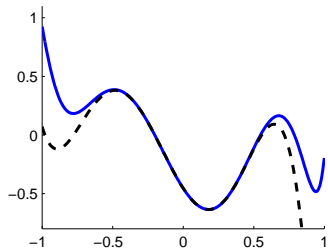


$N = 15$

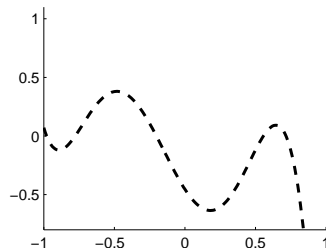
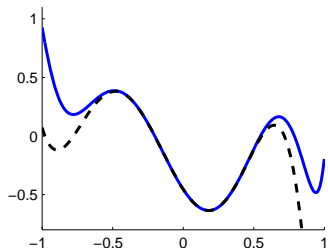
Veranschaulichung Approximationsordnung



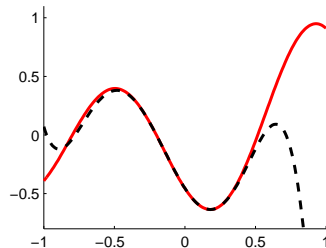
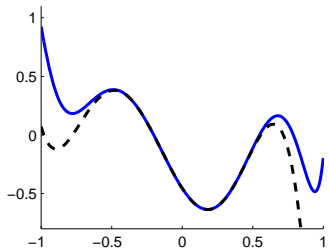
Veranschaulichung Approximationsordnung



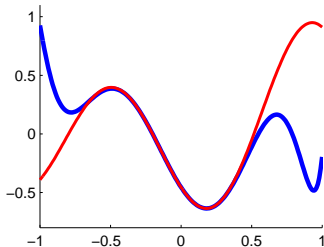
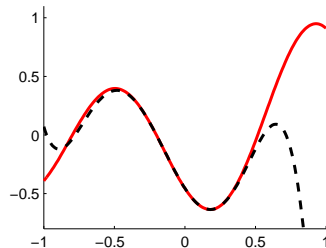
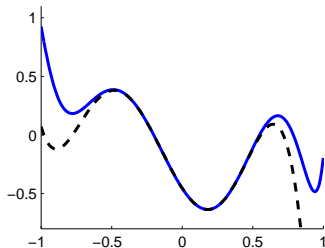
Veranschaulichung Approximationsordnung



Veranschaulichung Approximationsordnung



Veranschaulichung Approximationsordnung



Aufgabenstellung



Gegeben

- Hinreichend glatte Funktion $f : [-1, 1]^{n_0} \rightarrow \mathbb{R}$
- Gewünschte Approximationsordnung N

Aufgabenstellung



Gegeben

- Hinreichend glatte Funktion $f : [-1, 1]^{n_0} \rightarrow \mathbb{R}$
- Gewünschte Approximationsordnung N

Gesucht

Künstliches neuronales Netzwerk, welches f mit Ordnung N approximiert.

Aufgabenstellung



Gegeben

- Hinreichend glatte Funktion $f : [-1, 1]^{n_0} \rightarrow \mathbb{R}$
- Gewünschte Approximationsordnung N

Gesucht

Künstliches neuronales Netzwerk, welches f mit Ordnung N approximiert.

Wichtige Frage

Welche Größe des Netzwerkes ist mindestens erforderlich?

Aufgabenstellung



Gegeben

- Hinreichend glatte Funktion $f : [-1, 1]^{n_0} \rightarrow \mathbb{R}$
- Gewünschte Approximationsordnung N

Gesucht

Künstliches neuronales Netzwerk, welches f mit Ordnung N approximiert.

Wichtige Frage

Welche Größe des Netzwerkes ist mindestens erforderlich?

Konkreter: Wie viele Schichten und wie viele Neuronen pro Schicht werden benötigt?

Notwendige Größe eines Netzwerkes



Stellschrauben

- Kantengewichte
- Schwellwerte

Einzustellende Werte

Koeffizienten des Taylorpolynoms der Zielfunktion.

Notwendige Größe eines Netzwerkes



Stellschrauben

- Kantengewichte
- Schwellwerte

Einzustellende Werte

Koeffizienten des Taylorpolynoms der Zielfunktion.

Notwendige Bedingung - Intuition

Anzahl Stellschrauben \geq Anzahl einzustellender Werte

Gliederung



- 1 Bedeutung künstlicher neuronaler Netze (KNNs)
- 2 Approximationsgüte und Approximationsordnung
- 3 Anzahl der Koeffizienten des Taylorpolynoms**
- 4 Anzahl der Netzwerkparameter
- 5 Hauptergebnis: Explizite Formel für notwendige Größe

Anzahl Koeffizienten eines Polynoms



Beispiel

Ein Polynom vom Grad 2 in den 2 Variablen x_1 und x_2 :

$$1 + x_1 + 2x_2 - 3x_1x_2 + x_1^2 - x_2^2,$$

Anzahl Koeffizienten eines Polynoms



Beispiel

Ein Polynom vom Grad 2 in den 2 Variablen x_1 und x_2 :

$$1 + x_1 + 2x_2 - 3x_1x_2 + x_1^2 - x_2^2,$$

allgemeiner:

$$a_{00} + a_{10}x_1 + a_{01}x_2 + a_{11}x_1x_2 + a_{20}x_1^2 + a_{02}x_2^2.$$

Anzahl Koeffizienten eines Polynoms



Beispiel

Ein Polynom vom Grad 2 in den 2 Variablen x_1 und x_2 :

$$1 + x_1 + 2x_2 - 3x_1x_2 + x_1^2 - x_2^2,$$

allgemeiner:

$$a_{00} + a_{10}x_1 + a_{01}x_2 + a_{11}x_1x_2 + a_{20}x_1^2 + a_{02}x_2^2.$$

Anzahl Koeffizienten: **6**

Anzahl Koeffizienten eines Polynoms



Beispiel

Ein Polynom vom Grad 2 in den 2 Variablen x_1 und x_2 :

$$1 + x_1 + 2x_2 - 3x_1x_2 + x_1^2 - x_2^2,$$

allgemeiner:

$$a_{00} + a_{10}x_1 + a_{01}x_2 + a_{11}x_1x_2 + a_{20}x_1^2 + a_{02}x_2^2.$$

Anzahl Koeffizienten: **6**

Offensichtlich: Anzahl hängt von der Variablenanzahl und vom Grad ab.

Formel für Koeffizientenanzahl



Satz

Sei n_0 die Variablenanzahl und N der Grad eines Polynoms. Die Anzahl der Koeffizienten ist dann

$$\binom{N + n_0}{n_0} = \frac{(N + n_0)!}{N!n_0!}$$

Gliederung



- 1 Bedeutung künstlicher neuronaler Netze (KNNs)
- 2 Approximationsgüte und Approximationsordnung
- 3 Anzahl der Koeffizienten des Taylorpolynoms
- 4 Anzahl der Netzwerkparameter
- 5 Hauptergebnis: Explizite Formel für notwendige Größe

Fragestellung



Erinnerung: Notwendige Bedingung

Anzahl Stellschrauben \geq Anzahl einzustellender Werte

Fragestellung



Erinnerung: Notwendige Bedingung

Anzahl Stellschrauben \geq Anzahl einzustellender Werte (**bekannt**)

Fragestellung



Erinnerung: Notwendige Bedingung

Anzahl Stellschrauben \geq Anzahl einzustellender Werte (**bekannt**)

Anzahl Netzwerkparameter (Stellschrauben) abhängig von

- Anzahl Schichten des Netzwerkes
- Anzahl Neuronen pro Schicht

Fragestellung



Erinnerung: Notwendige Bedingung

Anzahl Stellschrauben \geq Anzahl einzustellender Werte (**bekannt**)

Anzahl Netzwerkparameter (Stellschrauben) abhängig von

- Anzahl Schichten des Netzwerkes
- Anzahl Neuronen pro Schicht

Problem

Zuordnung

Anzahl Parameter \rightarrow Anzahl Schichten
Anzahl Neuronen pro Schicht

nicht eindeutig.

Eine weitere Bedingung



Neue Bedingung

Unter den Netzwerken, die eine bestimmte Anzahl von Parametern habe, wähle dasjenige mit der **minimalen Gesamtzahl von Neuronen**.

Eine weitere Bedingung



Neue Bedingung

Unter den Netzwerken, die eine bestimmte Anzahl von Parametern habe, wähle dasjenige mit der **minimalen Gesamtzahl von Neuronen**.

Satz (Formel für Anzahl Parameter)

$$P_{\max}(n) = \begin{cases} (n_0 + 2)n, & \text{falls } n \leq n_0 + 1 + \sqrt{4n_0 + m}, \\ \frac{(n + n_0 + 3)^2 - m}{4} - 2(n_0 + 1), & \text{ansonsten,} \end{cases}$$

wobei n_0 die Anzahl der Eingänge, n die Gesamtzahl der (versteckten) Neuronen und $m = n + n_0 \pmod{2} \in \{0, 1\}$.

Gliederung



- 1 Bedeutung künstlicher neuronaler Netze (KNNs)
- 2 Approximationsgüte und Approximationsordnung
- 3 Anzahl der Koeffizienten des Taylorpolynoms
- 4 Anzahl der Netzwerkparameter
- 5 Hauptergebnis: Explizite Formel für notwendige Größe

Erstes Teilergebnis



Erinnerung: Notwendige Bedingung

Anzahl Stellschrauben \geq Anzahl einzustellender Werte

Im Folgenden:

- n_0 - Anzahl der Eingänge des Netzwerkes
- N - gewünschte Approximationsordnung

Erstes Teilergebnis



Erinnerung: Notwendige Bedingung

Anzahl Stellschrauben \geq Anzahl einzustellender Werte

Im Folgenden:

- n_0 - Anzahl der Eingänge des Netzwerkes
- N - gewünschte Approximationsordnung

Satz (Anzahl Schichten)

Falls $\binom{N + n_0}{n_0} \leq (n_0 + 2)(n_0 + 1 + 2\sqrt{n_0})$, dann benutze **eine** (versteckte) Schicht, sonst **zwei**.

Erstes Teilergebnis



Erinnerung: Notwendige Bedingung

Anzahl Stellschrauben \geq Anzahl einzustellender Werte

Im Folgenden:

- n_0 - Anzahl der Eingänge des Netzwerkes
- N - gewünschte Approximationsordnung

Satz (Anzahl Schichten)

Falls $\binom{N + n_0}{n_0} \leq (n_0 + 2)(n_0 + 1 + 2\sqrt{n_0})$, dann benutze **eine** (versteckte) Schicht, sonst **zwei**.

Bemerkung: Insbesondere sind mehr als zwei Schichten nicht notwendig.

Hauptergebnis, 1. Teil



Satz (Anzahl Neuronen)

Falls $\binom{N + n_0}{n_0} \leq (n_0 + 2)(n_0 + 1 + 2\sqrt{n_0})$, dann wird eine Schicht mit mindestens

$$\frac{\binom{N + n_0}{n_0}}{n_0 + 2}$$

Neuronen benötigt.

Hauptergebnis, 1. Teil



Satz (Anzahl Neuronen)

Falls $\binom{N + n_0}{n_0} \leq (n_0 + 2)(n_0 + 1 + 2\sqrt{n_0})$, dann wird eine Schicht mit mindestens

$$\frac{\binom{N + n_0}{n_0}}{n_0 + 2}$$

Neuronen benötigt.

Bemerkung

Diese Anzahl gilt auch und ohne Einschränkung, wenn man sich von vornherein auf eine Schicht festlegt.

Hauptergebnis, 2. Teil



Satz (Anzahl Neuronen)

Falls $\binom{N + n_0}{n_0} > (n_0 + 2)(n_0 + 1 + 2\sqrt{n_0})$ werden zwei Schichten mit mindestens

$$\left\lceil \frac{n + n_0 - 1}{2} \right\rceil \text{ (in der ersten Schicht) und}$$

$$\left\lceil \frac{n - n_0 + 1}{2} \right\rceil \text{ (in der zweiten Schicht)}$$

Neuronen benötigt, wobei

$$n = \left\lceil 2\sqrt{\binom{N + n_0}{n_0} + 2(n_0 + 1)} - n_0 - 3 \right\rceil.$$

Beispiele für notwendige Neuronenanzahl



Ordnung	Anzahl Eingänge					
	1	2	3	4	5	6
1	1	1	1	1	1	1
	1	1	1	1	1	1
2	1	2	2	3	3	4
	1	2	2	3	3	4
3	2	3	4	6	8	11
	2	3	4	6	8	11
4	2	4	7	12	18	27
	2	4	7	(7,4)	(10,6)	(13,8)
5	2	6	12	21	36	58
	2	6	(6,4)	(10,7)	(14,11)	(20,15)

Zusammenfassung und Ausblick



- Ausführliche qualitative Analyse von KNNs
- Approximationsordnung
 - Zusammenhang zur Approximationsgüte
 - Bedingungen für Äquivalenz
 - Hinreichende Voraussetzungen an Netzwerk für Äquivalenz
- Explizite Berechnung der notwendigen Größe
- Wichtig für Vermeidung von Overfitting bei gestörten Daten
- Simulationen

- Zukünftige Fragen:
 - Hinreichende Größe
 - Spezielle Lernverfahren

Zusammenfassung und Ausblick



- Ausführliche qualitative Analyse von KNNs
- Approximationsordnung
 - Zusammenhang zur Approximationsgüte
 - Bedingungen für Äquivalenz
 - Hinreichende Voraussetzungen an Netzwerk für Äquivalenz
- Explizite Berechnung der notwendigen Größe
- Wichtig für Vermeidung von Overfitting bei gestörten Daten
- Simulationen

- Zukünftige Fragen:
 - Hinreichende Größe
 - Spezielle Lernverfahren

Zusammenfassung und Ausblick



- Ausführliche qualitative Analyse von KNNs
- Approximationsordnung
 - Zusammenhang zur Approximationsgüte
 - Bedingungen für Äquivalenz
 - Hinreichende Voraussetzungen an Netzwerk für Äquivalenz
- Explizite Berechnung der notwendigen Größe
- Wichtig für Vermeidung von Overfitting bei gestörten Daten
- Simulationen

- Zukünftige Fragen:
 - Hinreichende Größe
 - Spezielle Lernverfahren

Zusammenfassung und Ausblick



- Ausführliche qualitative Analyse von KNNs
- Approximationsordnung
 - Zusammenhang zur Approximationsgüte
 - Bedingungen für Äquivalenz
 - Hinreichende Voraussetzungen an Netzwerk für Äquivalenz
- Explizite Berechnung der notwendigen Größe
- Wichtig für Vermeidung von Overfitting bei gestörten Daten
- Simulationen

- Zukünftige Fragen:
 - Hinreichende Größe
 - Spezielle Lernverfahren

Zusammenfassung und Ausblick



- Ausführliche qualitative Analyse von KNNs
- Approximationsordnung
 - Zusammenhang zur Approximationsgüte
 - Bedingungen für Äquivalenz
 - Hinreichende Voraussetzungen an Netzwerk für Äquivalenz
- Explizite Berechnung der notwendigen Größe
- Wichtig für Vermeidung von Overfitting bei gestörten Daten
- Simulationen
- Zukünftige Fragen:
 - Hinreichende Größe
 - Spezielle Lernverfahren

Zusammenfassung und Ausblick



- Ausführliche qualitative Analyse von KNNs
- Approximationsordnung
 - Zusammenhang zur Approximationsgüte
 - Bedingungen für Äquivalenz
 - Hinreichende Voraussetzungen an Netzwerk für Äquivalenz
- **Explizite Berechnung der notwendigen Größe**
- Wichtig für Vermeidung von Overfitting bei gestörten Daten
- Simulationen

- Zukünftige Fragen:
 - Hinreichende Größe
 - Spezielle Lernverfahren

Zusammenfassung und Ausblick



- Ausführliche qualitative Analyse von KNNs
- Approximationsordnung
 - Zusammenhang zur Approximationsgüte
 - Bedingungen für Äquivalenz
 - Hinreichende Voraussetzungen an Netzwerk für Äquivalenz
- **Explizite Berechnung der notwendigen Größe**
- Wichtig für Vermeidung von Overfitting bei gestörten Daten
- Simulationen

- Zukünftige Fragen:
 - Hinreichende Größe
 - Spezielle Lernverfahren

Zusammenfassung und Ausblick



- Ausführliche qualitative Analyse von KNNs
- Approximationsordnung
 - Zusammenhang zur Approximationsgüte
 - Bedingungen für Äquivalenz
 - Hinreichende Voraussetzungen an Netzwerk für Äquivalenz
- **Explizite Berechnung der notwendigen Größe**
- Wichtig für Vermeidung von Overfitting bei gestörten Daten
- Simulationen

- Zukünftige Fragen:
 - Hinreichende Größe
 - Spezielle Lernverfahren

Zusammenfassung und Ausblick



- Ausführliche qualitative Analyse von KNNs
- Approximationsordnung
 - Zusammenhang zur Approximationsgüte
 - Bedingungen für Äquivalenz
 - Hinreichende Voraussetzungen an Netzwerk für Äquivalenz
- **Explizite Berechnung der notwendigen Größe**
- Wichtig für Vermeidung von Overfitting bei gestörten Daten
- Simulationen

- Zukünftige Fragen:
 - Hinreichende Größe
 - Spezielle Lernverfahren

Zusammenfassung und Ausblick



- Ausführliche qualitative Analyse von KNNs
- Approximationsordnung
 - Zusammenhang zur Approximationsgüte
 - Bedingungen für Äquivalenz
 - Hinreichende Voraussetzungen an Netzwerk für Äquivalenz
- **Explizite Berechnung der notwendigen Größe**
- Wichtig für Vermeidung von Overfitting bei gestörten Daten
- Simulationen

- Zukünftige Fragen:
 - Hinreichende Größe
 - Spezielle Lernverfahren

Zusammenfassung und Ausblick



- Ausführliche qualitative Analyse von KNNs
- Approximationsordnung
 - Zusammenhang zur Approximationsgüte
 - Bedingungen für Äquivalenz
 - Hinreichende Voraussetzungen an Netzwerk für Äquivalenz
- **Explizite Berechnung der notwendigen Größe**
- Wichtig für Vermeidung von Overfitting bei gestörten Daten
- Simulationen

- Zukünftige Fragen:
 - Hinreichende Größe
 - Spezielle Lernverfahren