

Universität Mannheim
Seminar 1998



Approximation mit Ridge-Funktionen, sigmoidale
Funktionen und neuronalen Netzen nach Will Light

Referent: Lin Mischa Himmelmann
Seminarleitung: Walz, Zeilfelder

1998

1

Approximation mit Ridge-Funktionen, sigmoidalen Funktionen und neuronalen Netzen nach Will Light

1.1 Überblick

Approximiert werden Funktionen der Art $f : D \rightarrow \mathbb{R}$, auf einem geeignetem Gebiet $D \subseteq \mathbb{R}^n$. Diese Funktionen werden mit Ridge-Funktionen und sigmoidalen Funktionen approximiert. Sigmoidale Funktionen spielen bei der Approximation mit neuronalen Netzen eine wichtige Rolle. Bedeutend ist hierbei die Existenz konstruktiver Beweise, dass gewisse neuronale Netze eine gegebene Funktionen f mit beliebiger Genauigkeit approximiert.

Eine wichtige Frage hierbei ist, ob für eine gegebene stetige sigmoidale Funktion σ , der Raum der Funktionen $\{x \mapsto \sigma(yx + \theta) : y \in \mathbb{R}^n, \theta \in \mathbb{R}\}$, $x \in \mathbb{R}^n$ dicht in $C(\mathbb{R}^n)$ ist. Den Beweis lieferte Cybenko. Daraufhin beschäftigten sich Sun und Cheney sowie Chui und Li mit der Frage, ob auch der Raum $\{x \mapsto \sigma(zx + k) : z \in \mathbb{Z}^n, k \in \mathbb{Z}\}$, $x \in K$ dicht in $C(K)$ ist ($K \subseteq \mathbb{R}^n$ kompakter Teilraum). Der Vorteil gegenüber Cybenko ist, dass der Raum wesentlich kleiner ist, sogar abzählbar.

1.2 Verwendete Grundlagen

Definition 1.1 Sei $\phi \in C(\mathbb{R}^m)$ und $x \in \mathbb{R}^m$. $\hat{\phi}(x) := \frac{1}{(2\pi)^{\frac{m}{2}}} \int_{\mathbb{R}^m} \phi(t) e^{-itx} d^m t$. $\hat{\phi}(x)$ heißt die Fourier-Transformierte von ϕ .

Satz 1.2 Ist $\phi \in C(\mathbb{R}^m)$ mit kompaktem Träger, dann gilt: $(\frac{\partial}{\partial t_\nu} \hat{\phi})(x) = ix_\nu \hat{\phi}(x)$.

Satz 1.3 "Hahn-Banach" Sei B ein Banach-Raum und U Unterraum von B . Dann sind äquivalent:

- (i) Für ein lineares Funktional auf B , mit $L(U) = \{0\}$ folgt $L(B) = \{0\}$.
(ii) U ist dicht in B .

Satz 1.4 Sei $A \in M(n \times m)$ mit $\dim L(A) = k$. Sei P die orthogonale Projektion von \mathbb{R}^m nach $L(A)$. Seien $E : L(A) \rightarrow \mathbb{R}^k$ und $F : L(A)^\perp \rightarrow \mathbb{R}^{m-k}$ die üblichen Basistransformationen. Jedes $x \in \mathbb{R}^m$ kann geschrieben werden als $x = (x', x'')$, mit $x' = EPx$ und $x'' = F(I-P)x$. Weiter existiert ein $B \in M(n \times m)$ mit $Ax = Bx', \forall x \in \mathbb{R}^m$.

Satz 1.5 Sei G ein fundamentaler Raum in $C(\mathbb{R})$ und Φ ein Unterraum von dem stetigen Dualraum X^* des Banachraums X . Wenn $\Phi_0 = \{\frac{\phi}{\|\phi\|} : \phi \in \Phi \setminus \{0\}\}$ dicht in $S(X^*)$ ist, dann ist $A(G, \Phi) = \{g \circ \phi : g \in G, \phi \in \Phi\}$ fundamental in $C(X)$.

1.3 Ergebnis

Zunächst bestimmt man, wann die Funktionenmenge

$$M(\Omega) := \{x \mapsto g(Ax) : A \in \Omega, g \in C(\mathbb{R}^n), x \in \mathbb{R}\} \quad (1.1)$$

fundamental in $C(\mathbb{R}^n)$ ist. Hierbei sei $1 \leq m \leq n - 1$ und Ω ein Teilraum von $M(n \times m)$.

Zu $A = (a_{ij})_{\substack{i=1,\dots,m \\ j=1,\dots,n}} \in M(n \times m)$ definiere man

$$L(A) := \text{span}\left\{\begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix}, j = 1, \dots, n\right\} \subseteq \mathbb{R}^m, \text{ und} \quad (1.2)$$

$$L(\Omega) := \cup_{A \in \Omega} L(A). \quad (1.3)$$

Weiter sei H_k der Raum der homogenen Polynome vom Grad k auf \mathbb{R}^m , $k \in \mathbb{N}_0$. $H := \cup_{k \in \mathbb{N}_0} H_k$ ist somit der Raum der homogenen Polynome (mit natürlichem Homogenitätsgrad) im \mathbb{R}^m . Ein Polynom $p \in H_k$ besitzt somit folgende Darstellung:

$$p(x_1, \dots, x_m) = \sum_{\substack{\alpha_1, \dots, \alpha_m \in \mathbb{N}_0 \\ \alpha_1 + \dots + \alpha_m = k}} c_{\alpha_1, \dots, \alpha_m} x_1^{\alpha_1} \dots x_m^{\alpha_m}. \quad (1.4)$$

Notation: $p(x) = \sum_{|\alpha|=k} c_\alpha x^\alpha$. Mit den obigen Definitionen kann man folgende Charakterisierung geben:

Satz 1.6 *Der Raum $M(\Omega)$ ist fundamental in $C(\mathbb{R}^n)$ genau dann, wenn es kein $p \in H \setminus \{0\}$ gibt, mit $p(L(\Omega)) = \{0\}$.*

Beweis: “ \Rightarrow “ Zeige die Negation der Aussage.

Sei also $p \in H \setminus \{0\}$ mit $p(x) \equiv 0, \forall x \in L(\Omega)$.

p besitzt mit obiger Notation folgende Darstellung:

$$p(x) = \sum_{|\alpha|=k} c_\alpha x^\alpha, x \in \mathbb{R}^m, \alpha \in \mathbb{N}_0^m, \text{ wobei nicht alle } c_\alpha \text{ Null sind.}$$

Sei nun $\phi \in C^\infty(\mathbb{R}^m) \setminus \{0\}$ mit kompaktem Träger.

Existenz so einer Funktion:

Suche zunächst: $\tilde{\phi} \in C^\infty(\mathbb{R}) \setminus \{0\}$ mit kompaktem Träger:

$$\tilde{\phi}(x) := \begin{cases} e^{\frac{1}{x^2-1}} & , (-1, 1) \\ 0 & , \text{sonst.} \end{cases}$$

Die k -te Ableitung hat folgende Form:

$$\tilde{\phi}^{(k)}(x) = \begin{cases} r(x)e^{\frac{1}{x^2-1}} & , (-1, 1) \\ 0 & , \text{sonst.} \end{cases}$$

mit einer rationalen Funktion r .

$\tilde{\phi}(x) \in C^\infty(\mathbb{R}) \setminus \{0\}$ mit kompaktem Träger, da für $x \rightarrow \pm 1$ der Ausdruck $e^{\frac{1}{x^2-1}}$ gegen Null geht, und zwar stärker als jede rationale Funktion gegen $\pm\infty$ gehen kann.

Wähle nun $\phi \in C^\infty(\mathbb{R}^m) \setminus \{0\}$ mit kompaktem Träger, definiert durch $\phi(x_1, \dots, x_m) := \tilde{\phi}(x_1) \dots \tilde{\phi}(x_m)$. Dann besitzt ϕ die gewünschten Eigenschaften.

Bezeichne $D^\alpha := \frac{\partial^\alpha}{\partial x_1^{\alpha_1} \dots \partial x_m^{\alpha_m}}$, mit $\alpha = (\alpha_1, \dots, \alpha_m) \in \mathbb{N}_0^m$.

Dann definiere man zu ϕ wie oben:

$$\psi(x) := \sum_{|\alpha|=k} c_\alpha (D^\alpha \phi)(x), \text{ mit } x \in \mathbb{R}^m, c_\alpha \text{ wie oben.}$$

Dann gilt:

$\psi \in C^\infty(\mathbb{R}^m) \setminus \{0\}$, mit kompaktem Träger. Dies sieht man, da ϕ nur endlich oft in seinen m -Komponenten abgeleitet wird.

Zu ϕ und ψ existiert die sogenannte Fourier-Transformierte $\hat{\phi}$ und $\hat{\psi}$, da diese die oben genannten Eigenschaften besitzen. Diese Fourier-Transformierten (vgl. Definition 1.1) sind definiert durch: $\hat{\phi} = \frac{1}{(2\pi)^{\frac{m}{2}}} \int_{\mathbb{R}^m} \phi(t) e^{-itx} d^m t$.¹

¹Siehe auch Forster Analysis III, §12

Dann gilt nach dem fundamentalen Satz 1.1 aus der Fourier-Theorie:

$$\hat{\psi} = i^k \hat{\phi}(x)p(x).^2$$

Für den nächsten Schritt braucht man Satz 1.3 von Hahn-Banach:

Sei B ein Banach-Raum und U Unterraum von B . Dann sind äquivalent:

- (i) Für ein lineares Funktional auf B , mit $L(U) = \{0\}$ folgt $L(B) = \{0\}$.
- (ii) U ist dicht in B .

Jetzt kann man ein spezielles nichttriviales, stetiges, lineares Funktional L auf dem Raum $C(\mathbb{R}^m)$ angeben:

$$\begin{aligned} L : C(\mathbb{R}^m) &\longrightarrow \mathbb{R} \\ f &\longmapsto \int_{\mathbb{R}^m} f(x)\psi(x)dx \end{aligned}$$

Zeige: $\int_{\mathbb{R}^m} g(Ax)\psi(x)dx \stackrel{!}{=} 0, \forall A \in \Omega, g \in C(\mathbb{R}^n)$.

Dann hätte man ein lineares nichttriviales Funktional L auf dem \mathbb{R}^m , welches $\text{span}M(\Omega) \subseteq \mathbb{R}^m$ vernullt. Nach Satz 1.3 von Hahn Banach, kann dann $\text{span}M(\Omega)$ nicht dicht in $C(\mathbb{R}^m)$ sein. Und somit auch nicht in $C(\mathbb{R}^n)$, was die eine Richtung des Satz 1.6 beweist.

Sei $A \in \Omega, k = \dim(L(A)) \leq m$.

Zerlege $x \in \mathbb{R}^m$ gemäß Satz 1.4 in $x = (x', x'')$.

Sei $a \in \mathbb{R}^k$ beliebig und $v_a := (a, 0) \in \mathbb{R}^m$.

Nach Satz 1.4 ist $v_a \in L(A)$.

$$\begin{aligned} 0 &= i^k \hat{\phi}(v_a) \overbrace{p(v_a)}^{=0} \\ &= \hat{\psi}(v_a) \\ &= \frac{1}{(2\pi)^{\frac{m}{2}}} \int_{\mathbb{R}^m} \psi(x) e^{-iv_a x} dx \\ &= \frac{1}{(2\pi)^{\frac{m}{2}}} \int_{\mathbb{R}^k} \int_{\mathbb{R}^{m-k}} \psi(x', x'') e^{-iax'} dx'' dx' \\ &=: \frac{1}{(2\pi)^{\frac{m}{2}}} \int_{\mathbb{R}^k} h(x') e^{-iax'} dx' \\ &\stackrel{(*)}{=} \hat{h}(a) \end{aligned}$$

(*) Fourier-Transformation existiert, da h kompakten Träger hat.

$$\hat{h}(a) \equiv 0 \quad \forall a \in \mathbb{R}^m \Rightarrow h \equiv 0.$$

²Siehe Forster Analysis III, §12, Satz 1.c

Schließlich gilt:

$$\begin{aligned}
 & \int_{\mathbb{R}^m} g(A(x))\psi(x)dx \\
 &= \int_{\mathbb{R}^k} \int_{\mathbb{R}^{m-k}} g(Bx')\psi(x', x'')dx'' dx' \\
 &= \int_{\mathbb{R}^k} g(Bx') \underbrace{h(x')}_{=0} dx' \\
 &= 0.
 \end{aligned}$$

“ \Leftarrow “ Zeige: $H_k \subseteq \text{span}M(\Omega)$.

Sei $d_k := \dim H_k$, dann existieren $b_1, \dots, b_{d_k} \in L(\Omega)$ derart:

Falls $p \in H_k$ mit $p(b_i) = 0$ folgt $p \equiv 0$.

Die b_i 's haben folgende Darstellung: $b_i = y_i * A_i$, mit $y_i \in \mathbb{R}^n$, $A_i \in \Omega$

geeignet. Also:

$(b_i x)^k = (y_i A_i x)^k =: g_i(A_i x)$, mit $g_i \in C(\mathbb{R}^n)$, was zeigt,

dass $x \mapsto (b_i x)^k \in M(\Omega)$ ist.

Zeige: $\text{span}_{i=1, \dots, d_k} \{(b_i x)^k\} = H_k$.

Es genügt zu zeigen, dass diese Funktionen einen linear unabhängigen Raum aufspannen, da sie in H_k liegen und d_k -Stück sind.

Angenommen $\{(b_i x)^k\}_{i=1, \dots, d_k}$ sind linear abhängig.

Dann existieren Koeffizienten $\lambda_i, i = 1, \dots, d_k$ mit mind. einem $\lambda_i \neq 0$, so dass $\sum_{i=1}^{d_k} \lambda_i (b_i x)^k \equiv 0 \quad \forall x \in \mathbb{R}^n$.

Für $\alpha \in \mathbb{N}_0^m, |\alpha| = k$ gilt:

$$0 = D^\alpha \left(\sum_{i=1}^{d_k} \lambda_i (b_i x)^k \right) = k! \sum_{i=1}^{d_k} \lambda_i b_i^\alpha.$$

Also:

$$\sum_{i=1}^{d_k} \lambda_i b_i^\alpha \equiv 0, \text{ mit mindestens einem } \lambda_i \neq 0.$$

$\Rightarrow \{b_i^\alpha\}_{i=1, \dots, d_k}^{|\alpha|=k}$ bildet eine $d_k \times d_k$ -Matrix, welche nach obiger Darstellung singular ist.

$\Rightarrow \exists \{\mu_\alpha\}_{|\alpha|=k} \subseteq \mathbb{R}$, so dass $\sum_{|\alpha|=k} \mu_\alpha b_i^\alpha = 0$ für $i = 1, \dots, d_k$ mit mindestens einem $\mu_\alpha \neq 0$.

Setze nun $p(x) = \sum_{|\alpha|=k} \mu_\alpha x^\alpha$.

Dann gilt $p \in H_k \setminus \{0\}$ und $p(b_i) = 0$, für $i = 1, \dots, d_k$. Widerspruch!

Also gilt $\text{span}\{(b_i x)^k\} = H_k$

$\Rightarrow H_k \subseteq \text{span}M(\Omega)$, k beliebig

$$\Rightarrow H \subseteq \text{span}M(\Omega)$$

Da in $\text{span}M(\Omega)$ der Raum der Polynome enthalten ist, ist $\text{span}M(\Omega)$ nach Weierstraß dicht in $C(\mathbb{R}^n)$. **q.e.d.**

Definition 1.7 Eine Funktion $\sigma : \mathbb{R} \rightarrow \mathbb{R}$, so dass $\lim_{t \rightarrow \infty} \sigma(t) = 1$ und $\lim_{t \rightarrow -\infty} \sigma(t) = 0$, heißt sigmoidale Funktion.

Satz 1.8 Sei σ eine stetige sigmoidale Funktion. Dann ist der Raum der Funktionen $\{t \mapsto \sigma(z \cdot t + l) : z, l \in \mathbb{Z}, t \in \mathbb{R}\}$ fundamental in $C(\mathbb{R})$.

Beweis:

Sei $f \in C(\mathbb{R})$ gegeben, und $S := \text{span}\{t \mapsto \sigma(z \cdot t + l) : z, l \in \mathbb{Z}, t \in \mathbb{R}\}$

Zu zeigen: Jede Umgebung von f schneidet S ,

d.h. $\{g \in C(\mathbb{R}) : \|f - g\|_{\mathbb{R}} < \epsilon\} \cap S \neq \emptyset$.

Angenommen, dies sei nicht der Fall.

Dann existiert $\epsilon > 0$ und $K \subseteq \mathbb{R}$ kompakt, so dass:

$$B_{K,\epsilon} := \{g \in C(\mathbb{R}^n) : \|f - g\|_K < \epsilon\} \cap S = \emptyset.$$

K liegt in einem Intervall $I = [a, b]$, mit rationalen Zahlen $a, b \in \mathbb{Q}$.

Man zeigt im folgenden, dass die Restriktion der Funktionen in S auf I dicht in $C(I)$ ist.

\Rightarrow Die Restriktion der Funktion auf K ist dicht in K .

\Rightarrow Widerspruch!, da dann $B_{K,\epsilon} \cap S \neq \emptyset$.

Sei hierzu E die Restriktion in S auf I . Sei μ ein lineares stetiges Funktional in $(C(I))^*$, d.h. $\mu : C(I) \rightarrow \mathbb{R}$, so dass $\mu(e) = 0, \forall e \in E$.

Nach Hahn-Banach ist zu zeigen: $\mu(C(I)) \equiv 0$, denn dann ist E dicht in $C(I)$.

Identifiziere nach Riesz $(C(I))^*$ mit $C(I)$.

Dann gilt: $\int_a^b \sigma(zt + l) d\mu(t) = 0, \quad \forall z, l \in \mathbb{Z}$

Wähle $l, p, q \in \mathbb{Z}$ mit $q > 0$ und $\frac{p}{q} \in I$.

Definiere $r : I \rightarrow \mathbb{R}$ durch:

$$r(t) = \begin{cases} 0 & , a \leq t < \frac{p}{q} \\ \sigma(l) & , t = \frac{p}{q} \\ 1 & , \frac{p}{q} < t \leq b \end{cases}$$

Betrachte: $\sigma\left(nq\left(t - \frac{p}{q}\right) + l\right)$ für $t \in I$.

Für $n \rightarrow \infty$ konvergiert σ punktweise gegen r .

$$\begin{aligned} \Rightarrow 0 &= \lim_{n \rightarrow \infty} \int_a^b \overbrace{\sigma\left(nq\left(t - \frac{p}{q}\right) + l\right)}^{\in E} d\mu(t) \\ &\stackrel{\text{Lebesgue}}{=} \int_a^b r(t) d\mu(t) \\ &= \mu\left(t \in I : t > \frac{p}{q}\right) + \sigma(l)\mu\left(\frac{p}{q}\right) \end{aligned}$$

Falls $l \rightarrow -\infty$, geht $\sigma(l) \rightarrow 0$.

$$\Rightarrow \mu\left(t \in I : t > \frac{p}{q}\right) = 0, \quad \forall \frac{p}{q} \in I$$

Daraus folgt mit $l \rightarrow \infty$ ($\sigma(l) \rightarrow 1$): $\mu\left(\frac{p}{q}\right) = 0, \forall \frac{p}{q} \in I$

$\Rightarrow \mu \equiv 0$ auf I .

$\Rightarrow E$ ist dicht in $C(I)$, Widerspruch! **q.e.d.**

Satz 1.9 Sei σ eine stetige sigmoidale Funktion. Dann ist der Raum der Funktionen $H := \{t \mapsto \sigma(z \cdot t + k) : z \in \mathbb{Z}^n, k \in \mathbb{Z}, t \in \mathbb{R}^n\}$ fundamental in $C(\mathbb{R}^n)$.

Beweis: Wende Satz 1.5 an.

1. $G := \{t \mapsto \sigma(kt + l) : k, l \in \mathbb{Z}, t \in \mathbb{R}\}$ ist fundamental in $C(\mathbb{R})$.

2. Identifiziere den \mathbb{R}^n mit $(\mathbb{R}^n)^*$, $\Phi := \mathbb{Z}^n$ ist Untermenge des dualen Banachraumes \mathbb{R}^n .

3. $\Phi_0 := \left\{\frac{\phi}{\|\phi\|} : \phi \in \Phi \setminus \{0\}\right\}$ ist dicht in $S(\mathbb{R}^n)$.

$\Rightarrow A(G, \Phi)$ ist fundamental in $C(\mathbb{R}^n)$.

Ein Element von $A(G, \Phi)$ hat die Form:

$$\begin{aligned} (g \circ \phi)(t) &= g(\phi(t)) = g(zt) \quad \text{mit } z \in \mathbb{Z}^n \\ &= \sigma\left(\underbrace{kzt}_{\in \mathbb{Z}^n} + \underbrace{l}_{\in \mathbb{Z}}\right) \end{aligned}$$

d.h. $A(G, \Phi) \subseteq H$ **q.e.d.**