Statistische Aspekte des Deep Learnings

Prof. Dr. Michael Kohler
Fachbereich Mathematik
Technische Universität Darmstadt

kohler@mathematik.tu-darmstadt.de

Darmstadt, 18.09.2019

Vorbemerkung

Wenn ich mich an das Ende meiner Schulzeit (vor 30 Jahren) zurückerinnere, so fällt mir sofort ein, dass diese Zeit im nachhinein vor allem geprägt war durch den *Zusammenbruch der Sowjetunion* und den *Fall der Berliner Mauer*.

Wenn Sie später einmal auf die heutige Zeit zurückblicken werden, was vermuten Sie, wie Sie diese dann charakterisieren werden ?

Meine Antwort dazu ist:

Durch eine **industrielle Revolution**, bei der im Rahmen der **Digitalisierung** die *Computer immer mehr Aufgaben übernehmen*.

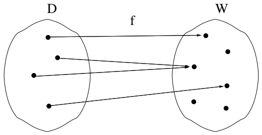
Was da dahinter steckt, erfahren Sie in den nächsten 15 Minuten ...

Überblick

1. *Lernen von Funktionen*
2. *Künstliche Neuronale Netze*
3. *Konvergenzgeschwindigkeit*

1. Lernen von Funktionen (1/4)

Eine *Funktion* $f : D \rightarrow W$ beschreibt eine Zuordnung von den Werten einer Definitionsmenge D zu Werten in einer Wertemenge W, wobei jedem Element der Definitionsmenge genau ein Element der Wertemenge zugeordnet wird.



Viele praktische Anwendungen lassen sich lösen, wenn man mit dem Computer geeignete Funktionen berechnen kann, z.B.

- Menge von Kamerabildern → Menge von Verkehrsschildern
- Menge von Motorzuständen → Menge von Schadstoffwerten am Auspuff

1. Lernen von Funktionen (2/4)

Problem: In vielen Anwendungen weiß man gar nicht, wie man die Funktionswerte der gesuchten Funktion berechnet.

Z.B. Vorhersage des Stickoxidsausstoßes am Auspuff bei einem LKW mit funktionierendem Katalysator in Abhängigkeit des Motorzustandes

Kennt man hier den funktionalen Zusammenhang, so kann man durch Messen des Stickoxidsausstoßes während des Fahrens erkennen, ob der Katalysator noch richtig funktioniert oder nicht.

Aber: Der funktionale Zusammenhang ist hier unbekannt.

Ausweg: Lernen aus beobachteten Werten ...

1. Lernen von Funktionen (3/4)

Beobachtete Werte bei Probefahrten mit einem LKW mit neuem Katalysator (die ersten 5 Datenpunkte von 30.000, gemessen wurden: Temperatur im Katalysator, Abgasmassenstrom, Stickoxidkonzentration vor Katalysator, AdBlue-Menge im Katalysator, Stickoxidkonzentration nach Katalysator):

	T-SCR	Exh-MF	NOx-EO	DEF-Mf	NOx-TP
1	258.46	597.75	856	524.8	3
2	257.96	561.75	646	331.5	15
3	257.86	522.00	599	269.8	15
4	257.36	508.75	584	233.9	14
5	257.56	571.25	628	242.7	10

1. Lernen von Funktionen (4/4)

Damit Aufgabenstellung:

Ausgehend von fehlerbehafteten Beobachtungen von Funktionswerten einer Funktion (mit Werten in der Menge der reellen Zahlen) **an zufälligen Stellen wollen wir diese Funktion rekonstruieren.**

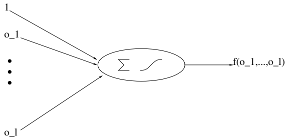
Problem dabei: Definitionsbereich ist hochdimensional und die Bauart der Funktion komplett unbekannt.

Im Folgenden modellieren wir die unbekannte Funktion durch tiefe neuronale Netze (und das Anpassen dieser tiefen neuronalen Netze an Daten nennt man dann *Deep Learning*).

2. Künstliche neuronale Netze (1/5)

Für künstliche neuronale Netze verwenden wir eine sehr einfache Modellierung von

a) Nervenzellen:



$$f(o_1, \ldots, o_l) = \sigma\left(w_0 + w_1 \cdot o_1 + \cdots + w_l \cdot o_l\right) \quad (o_1, \ldots, o_l \in \mathbb{R})$$

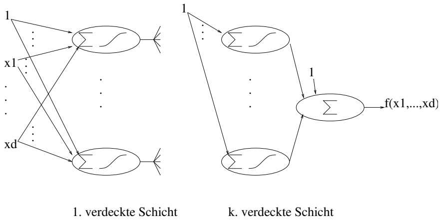
wobei

- $w_0, \ldots, w_l \in \mathbb{R}$ die sogenannten *Gewichte* sind,
- $\sigma : \mathbb{R} \to \mathbb{R}$ die sogenannte *Aktivierungsfunktion* ist, die eine Art Schwellenwertbildung macht, z.B. die sogennante *logistische Aktivierungsfunktion*

$$\sigma(x) = \frac{1}{1 + e^{-x}}.$$

2. Künstliche neuronale Netze (2/5)

b) Netzwerken von Nervenzellen:



forwärtsgerichtetes neuronales Netz mit k verdeckten Schichten

Definiert Funktion $f : \mathbb{R}^d \to \mathbb{R}$, die abhängt von

- Topologie des Netzwerkes,
- Aktivierungsfunktion,
- Gewichten $w_{i,j}^{(l)} \in \mathbb{R}$.

Berechnung dieser Funktion:

$$f(x) = \sum_{i=1}^{k_L} w_{1,i}^{(L)} \cdot f_i^{(L)}(x) + w_{1,0}^{(L)},$$

für $w_{1,0}^{(L)}, \dots, w_{1,k_L}^{(L)} \in \mathbb{R}$ und für $f_i^{(L)}$'s rekursiv definiert durch

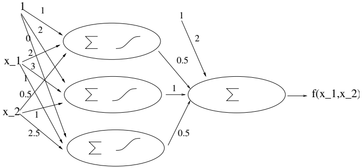
$$f_i^{(s)}(x) = \sigma\left(\sum_{j=1}^{k_{s-1}} w_{i,j}^{(s-1)} \cdot f_j^{(s-1)}(x) + w_{i,0}^{(s-1)}\right)$$

für $w_{i,0}^{(s-1)}, \dots, w_{i,k_{s-1}}^{(s-1)} \in \mathbb{R}$ $(s = 2, \dots, L)$ und

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

für $w_{i,0}^{(0)}, \dots, w_{i,d}^{(0)} \in \mathbb{R}$.

Beispiel:



Wir berechnen sukzessive:

$$
\begin{aligned}
y_1 &= \sigma\left(1 \cdot 1 + 2 \cdot x_1 + 0.5 \cdot x_2\right) \\
y_2 &= \sigma\left(2 \cdot 1 + 3 \cdot x_1 + 1 \cdot x_2\right) \\
y_3 &= \sigma\left(0 \cdot 1 + 1 \cdot x_1 + 2.5 \cdot x_2\right) \\
f(x_1, x_2) &= 2 \cdot 1 + 0.5 \cdot y_1 + 1 \cdot y_2 + 0.5 \cdot y_3
\end{aligned}
$$

2. Künstliche neuronale Netze (5/5)

Was ist besonders an neuronalen Netzen ?

Meine Lieblingsantwort (in Anlehnung an den früheren Wahlkampfslogan von Bill Clinton) ist:

It is the network, stupid!

Denn aufgrund der Netzwerkstruktur sind neuronale Netze besonders gut geeignet, um hierarchisch definierte Zusammenhänge zu modellieren.

3. Konvergenzgeschwindigkeit (1/2)

Wir passen ein neuronales Netz an eine gegebene Menge von n Datenpunkten an, indem wir bei geeignet gewählter fester Topologie des Netzes die Gewichte durch *Minimierung des durchschnittlichen quadratischen Fehlers* der Vorhersage der Funktionswerte bei den Datenpunkten durch das neuronale Netz minimieren.

Dann gilt:

3. Konvergenzgeschwindigkeit (2/2)

Satz (Kohler und Langer (2019)). Ist die zu schätzende Funktion so aufgebaut, dass

- rekursiv immer wieder Funktionen von d^* Variablen berechnet werden,
- die berechneten Funktionen alle p-fach stetig differenzierbar sind,

so ist der mittlere quadratische Fehler des obigen Schätzers nach oben beschränkt durch eine Konstante mal

$$n^{-\frac{2p}{2p+d^*}},$$

wobei n die Anzahl der Datenpunkte ist.

Vorteil: Konvergenzrate oben hängt nicht von der Dimension der Eingangsgröße ab!

Zusammenfasung

1. In Anwendungen ist die Schätzung von funktionalen Zusammenhängen nützlich.
2. Bei neuronalen Netzen handelt es sich um eine spezielle Funktionenklasse, die vom Menschen nur schlecht, am Rechner aber leicht rekursiv berechnet werden können.
3. Deep Learning passt tiefe neuronale Netze an beobachtete Daten an, um Funktionen zu schätzen.
4. Neuronale Netze sind besonders gut geeignet zur Schätzung von hierarchischen Zusammenhängen.

Schlusswort

Ein Aspekt der Mathematik, der bisher noch nicht erwähnt wurde:

Mathematik ist auch ein **sehr interessantes Studienfach!**

Warum soll man Mathematik studieren?

1. Extrem **vielseitiges, spannendes** und auch **herausforderndes Studienfach**
2. Seit Jahrzehnten (und auch in Wirtschaftskrisen) haben MathematikerInnen nach dem Studium **sehr gute Berufsaussichten**

   z.B. bei Banken, Versicherungen, Unternehmensberatungen, Forschungsabteilungen großer Unternehmen, ...

Wer kann Mathematik studieren?

Wichtig:

- Interesse
- Freude an abstrakten Überlegungen
- Neugier
- Ausdauer

Weniger wichtig:

- Vorwissen aus der Schule

Ganz unwichtig:

- Geschlecht

Was macht man im Mathematik-Studium an einer Universität?

- Vorlesungen in Mathematik
- Übungen zu den Vorlesungen
- Nebenfach
- Seminare
- Abschlussarbeit

Dauer: Bachelor-Studium: 3 Jahre, Master-Studium: 2 Jahre.

Was ist besonders am Mathematik-Studium an der TU Darmstadt?

- **Vielseitiges Studienangebot** in Mathematik (Mathematik, Wirtschaftsmathematik, Lehramtstudiengang Mathematik, zum Teil auch bilingual)
- sehr großer Fachbereich mit **vielen verschiedenen Vertiefungsmöglichkeiten** innerhalb der Mathematik
- **Intensive Betreuung der Studierenden** (FB Mathematik der TU Darmstadt schneidet seit Jahren bei den CHE-Rankings sehr gut ab)
- **Vielzahl von Stellenangebote** für studentische Hilfskräfte
- **Sehr gute Promotionsmöglichkeiten** nach dem Studium (z.B. im Rahmen einer Vielzahl von Drittmittelprojekten)

Wird man problemlos zum Mathematik-Studium zugelassen?

An vielen Universitäten ja, seit 2011 an der TU Darmstadt aber nicht mehr ganz so leicht:

Angesichts *hoher Abbrecherquoten* zu Beginn des Studiums (die nicht nur in der Mathematik, sondern in fast allen Fächern auftreten) und dem *großen Andrang* im Studiengang Mathematik in der Vergangenheit setzt die Zulassung zum Mathematik-Studium an der TU Darmstadt das erfolgreiche Durchlaufen eines *Eignungsfeststellungsverfahrens* voraus.

BewerberInnen mit *Abiturnote* von *2.3 oder besser* werden sofort zugelassen, alle anderen werden zu einem persönlichen Gespräch eingeladen und zu ihrer *Motivation* für das Mathematik-Studium befragt.

Weitere Informationen:

- Informationsmaterial, u.a. zum Mathematik-Studium, liegt an den beiden Ausgängen bereit.
- Bei Interesse an Aufnahme in Emailverteiler des Fachbereichs Mathematik bitte entsprechendes Formular ausfüllen und in Urne werfen.
- Frei zugängliche Aufzeichnungen von Mathematik-Vorlesungen an der TUD finden Sie unter

  www.openlearnware.de

  Dort auf *Mathematik* gehen, für SchülerInnen am ehesten geeignet ist dort die Vorlesung "Analysis I", die sich an Studierende im 1. Semester richtet.
- Onlinehilfen zur Studienwahl bietet die TU Darmstadt unter

  www.tu-osa.de

  an.