

Mathematik, Machine Learning und Artificial Intelligence

Sebastian Pokutta

*Machine Learning (ML) und Artificial Intelligence (AI)*¹ haben in den letzten Jahren enorme Aufmerksamkeit in der Wissenschaft, der Presse und der Öffentlichkeit erfahren. Dies liegt nicht nur an dem Vorstellungsvermögen des Einzelnen, was denn nun diese neue Technologie an Möglichkeiten und – beflügelt durch den letzten Terminator-Film – Gefahren mit sich bringen mag, sondern insbesondere daran, dass ML/AI bemerkenswerte Erfolge vorzuweisen hat und das Potenzial besitzt eine Revolution der Wissenschaft, im Kuhnschen Sinne², herbeizuführen: von einem Modell-getriebenen zu einem Daten-getriebenen Paradigma.

Vorbemerkung. Ich verwende die Begriffe *Machine Learning (ML)*, *Artificial Intelligence (AI)* und *Data Science (DS)* im austauschbaren Sinne. Auch wenn sich diese Begriffe je nach Fachgebiet etwas unterscheiden, dient die Unterscheidung im öffentlichen Diskurs oft eher dem Marketing als der inhaltlichen Differenzierung.

ML Erfolgsbeispiele

Beispiel 1 – Computer Vision

Eines der frühen Felder, in denen sogenannte *Deep Convolutional/Neural Networks* – eine der Kerntechniken im modernen ML – überraschende Erfolge vorzuweisen hatten, ist der Bereich *Computer Vision*. Oftmals geht es bei sogenannten diskriminativen Ansätzen um Objekterkennung (z. B. Erkennen von Verkehrszeichen), Gesichtserkennung (z. B. Identifikation beim Flugzeug-Boarden), Nachverfolgung von Objekten in Videos (z. B. räumliche Zuordnung von Fahrzeugen und Personen im Verkehrsraum). Bei generativen

Ansätzen geht es oft um die Rekonstruktion von fehlenden Inhalten (z. B. um die realistische Ersetzung fehlender Teile in Bildern³) als auch um die Erzeugung von fiktiven Inhalten (z. B. *Deep Fakes*, bei denen realistische Videos künstlich erzeugt werden⁴). In *Computer Vision* wurden in einigen Benchmarks früh Ergebnisse erzielt, die weit über die Leistungsfähigkeit des Menschen hinausgingen. Die Anwendungen dieser Methoden sind vielseitig. Kanonische Beispiele sind visuelle Systeme im Autonomen Fahren⁵ (z. B. Erkennung von Verkehrszeichen, Fußgängern und Fahrzeugen) und medizinische Diagnostik⁶ (z. B. Krebsfrüherkennung und CT/MRI Interpretation). ML in *Computer Vision* ist auch erfolgreich in die Kunst vorgedrungen⁷ und die erste erfolgreiche Auktion von AI-generierter Kunst fand 2018 bei Christie's statt; das Bild wurde für \$432 500 versteigert.⁸

Beispiel 2 – Protein Folding

Proteine sind „Bausteine aller Lebensformen“. Nobelpreisträger Christian Anfinsen postulierte, dass die energie-minimale Faltung eines Proteins im 3D-Raum durch sei-

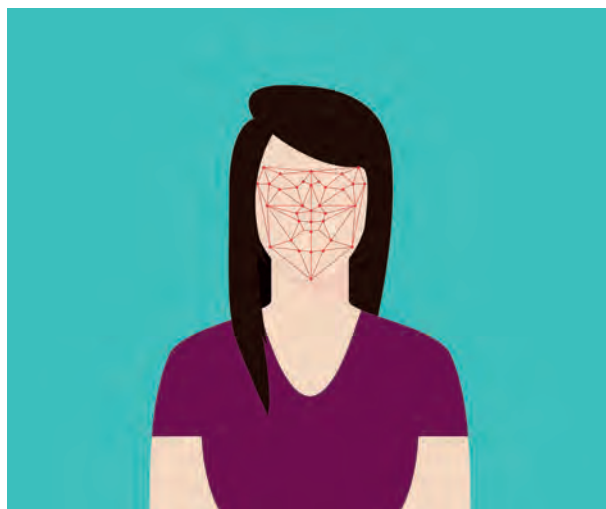
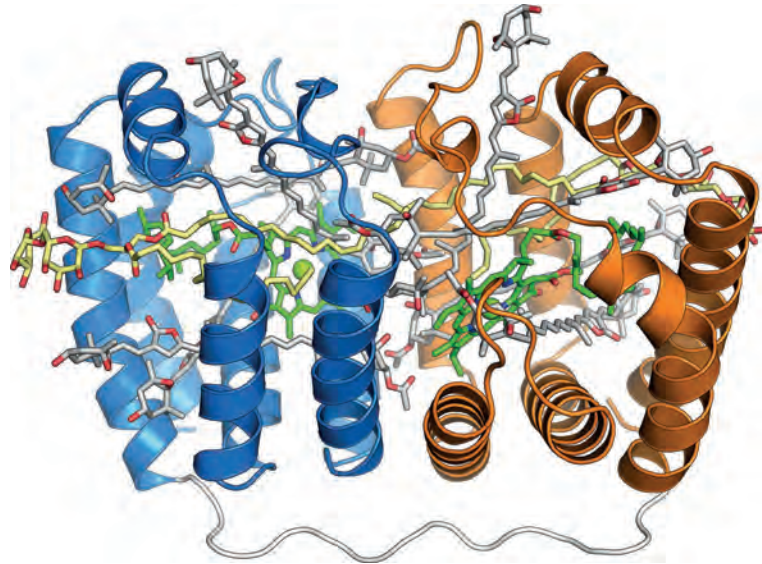


Abbildung: teghjatipras, Creative Commons, CC0 1.0



Proteinkomplex (Abbildung: Opabinia regalis (CC BY-SA 4.0, creativecommons.org/licenses/by-sa/4.0))

ne Aminosäuresequenz bestimmt sei.⁹ Dies führte zu dem sogenannten *Protein Folding Problem*: Gegeben eine Aminosäuresequenz, was ist die energieminimale Faltung? Das *Protein Folding Problem* ist eines der größten ungelösten Probleme in der Biologie. Projekte wie Folding@Home¹⁰ nutzen die Rechenressourcen ungenutzter Computer in aller Welt, um Tausende von Proteinen in aufwendigen Simulationen zu falten. Vor wenigen Wochen hat Deepmind, ein AI-Forschungslabor aus Großbritannien, welches seit 2014 zu Google gehört, bekanntgegeben, das *Protein Folding Problem* mit Hilfe von AI gelöst zu haben.¹¹ Stark vereinfacht gesagt werden hier Konfigurationen nicht aussimuliert, sondern die finale Konfiguration „vorhergesagt“; der Vollständigkeit halber ist zu erwähnen, dass Deepmind's Ergebnisse noch nicht *peer-reviewed* sind. Bestätigen sich die Ergebnisse jedoch, so wären die Auswirkungen in Medizin, Biologie, Genetik, etc. enorm.

Beispiel 3 – Reinforcement Learning und Lernen von komplexen Dynamiken

Reinforcement Learning ist ein maschineller Lernansatz, bei dem ein Algorithmus automatisch lernt und seine eigenen Fähigkeiten verbessert. Dieser Prozess wird als Reinforcement Learning bezeichnet und wird häufig in Bereichen wie Suche und Robotik eingesetzt.

Eine der Schlüsselideen des *Reinforcement Learning* ist, dass es ein Belohnungssignal verwendet, um das Verhalten eines Algorithmus zu steuern. Dieses Signal ist normalerweise eine Belohnung (typischerweise eine Belohnung aus der Umgebung), die der Algorithmus erhält, wenn er eine Aufgabe ausführt. Dieses Signal kann verwendet werden, um einen Algorithmus zu motivieren, die Aufgabe weiterhin auszuführen.

Die Erfolge, die mit *Reinforcement Learning* erzielt werden konnten, sind beeindruckend. So gilt es, *AlphaGo's* Sieg gegen den Go-Weltmeister Lee Sedol¹², zu dem Zeitpunkt,

ob der Komplexität von Go, für unmöglich gehalten zu nennen. Andere Beispiele sind Multi-Agenten Modelle, die selbständig komplexe Verhalten beim „Versteckenspielen“ rein durch Belohnung bei erfolgreichem „Verstecken“ bzw. „Finden“ lernen.¹³ Darüber hinaus gibt es starke Parallelen zwischen *Reinforcement Learning* und der Rolle von Dopamin beim menschlichen Lernen.¹⁴

Beispiel 4 – Natural Language Processing

ML/AI-Ansätze waren auch bei der Verarbeitung natürlicher Sprache äußerst erfolgreich und eine besondere Anwendung ist die Übersetzung. Mittlerweile gibt es sehr leistungsfähige Tools wie *DeepL*¹⁵ oder *Google Translate*¹⁶, die qualitativ hochwertige Übersetzungen zwischen verschiedenen Sprachpaaren anbieten und oft an professionelle Übersetzer herankommen oder mit ihnen konkurrieren. Selbst die komplexesten Sprachen können mit Ansätzen der künstlichen Intelligenz angegangen werden: Bereits 2016 gelang es Google¹⁷, ein Neuronales Netzwerk zu trainieren, dessen Übersetzungsgenauigkeit für Chinesisch–Englisch an das Niveau menschlicher Übersetzer herankommt; eine höchst nicht triviale Aufgabe, wenn man bedenkt, dass Chinesisch, wie die meisten asiatischen Sprachen, eine stark kontextabhängige Sprache ist.

Der obige Absatz ist eine solche Übersetzung aus dem Englischen verbatim durch DeepL; Fußnoten und Formatierung wurden nachträglich eingefügt. Das englische Original lautete:

ML/AI approaches have been also extremely successful in natural language processing and one particular application is translation. By now, there are very powerful tools out there such as DeepL or Google Translate that offer high quality translations between different language pairs, often getting close or rivaling professional translators. Even the most complex

languages can be tackled using artificial intelligence approaches: already in 2016 Google managed to train a neural network whose translation accuracy for Chinese to English approaches the level of human translators; a highly non-trivial task given that Chinese, as most Asian languages, is a highly context-dependent language.

Dann gibt es noch generative Sprachmodelle wie *GPT-3*¹⁸, die Texte aus einem einfachen „Seed“ Satz generieren, mit dem das Modell „angestoßen“ wird. Dies klingt erstmal unglaublich. Zur Demonstration, wie weit diese Technik fortgeschritten ist, habe ich die ersten beiden Absätze des obigen Beispiels 3 mittels einem deutlich einfacheren, aber frei verfügbaren *GPT-2* Modell¹⁹ erzeugt, welches mit dem Satzanfang „Reinforcement Learning is a machine learning approach“ gespeist wurde und dann ausgab:

Reinforcement Learning is a machine learning approach that uses an algorithm to automatically learn and improve its own capabilities. This process is called reinforcement learning, and it is widely used in areas like search and robotics.

One of the key ideas of Reinforcement Learning is that it uses a reward signal to help guide an algorithm's behavior. This signal is typically a reward (typically a reward from the environment) that the algorithm receives when it performs a task. This signal can be used to motivate an algorithm to continue to perform the task, and it

Ich habe dann den letzten Satz beendet durch Entfernen von „ , and it“ und dann mit DeepL zurück ins Deutsche übersetzt. Die Lesenden mögen das Resultat selbst beurteilen. Das Ganze lässt sich auf die Spitze treiben: *The Guardian* hatte einen vollständigen Artikel²⁰ mit einem *GPT-3* Modell generiert und veröffentlicht; das Ergebnis ist beeindruckend.

Die Rolle der Mathematik

Was hat dies alles mit Mathematik zu tun? *Machine Learning* kann als ein Teilgebiet der Mathematik betrachtet werden. Oder anders herum gesagt: die Mathematik ist (neben der Informatik und der Statistik) ein zentraler Bestandteil von *Machine Learning*. Oftmals wird ML auch als *Hype*, fern der Mathematik, abgetan. Dabei wird jedoch vergessen, dass die Ursprünge des Machine Learning in der Mathematik liegen (Stichwort: Radon Transformation) und von Titanen wie Kolmogorov, von Neumann, Turing oder Wiener vorangetrieben wurden. ML Forschung in Deutschland findet (überwiegend) an der Schnittstelle zwischen Mathematik, Informatik und Statistik (sofern sie isoliert betrachtet wird) sowohl an Universitäten als auch außeruniversitären Forschungseinrichtungen statt.

Mathematik als rigoroses Fundament

Was kann also die Mathematik beitragen? Im Folgenden betrachten wir exemplarisch ein Standardproblem in ML, das

sogenannte *Empirical Risk Minimization Problem*, welches in der einen oder anderen Form in allen oben genannten Beispielen auftaucht. Die Darstellung wurde hier bewusst vereinfacht ohne Referenzen zu einzelnen akademischen Artikeln gegeben, um Verzerrungen jeglicher Art zu vermeiden. Für weitere Probleme und Berührungspunkte seien die Lesenden an einen Übersichtsartikel²¹ in *SIAM News* verwiesen.

Empirical Risk Minimization (ERM) Problem

Gegeben endlich viele Trainingsdaten I bestehend aus Paaren (x, y) , wobei aus $x \in \mathbb{R}^m$ jeweils $y \in \mathbb{R}$ vorhergesagt werden soll, suchen wir ein Modell f parametrisiert durch einen Vektor θ , welches die „beste“ Vorhersage liefert:

$$\min_{\theta} \frac{1}{|I|} \sum_{(x,y) \in I} |y - f(x, \theta)|^2. \quad (\text{ERM})$$

Die Funktion f in (ERM) kann durchaus einfach (z. B. eine lineare Funktion) oder aber sehr komplex (z. B. ein Neuronales Netzwerk) sein. Ein konkretes, einfaches Beispiel für ein (ERM) Problem ist die lineare Regression, bei der wir *abhängige* Variablen y durch *unabhängige* Variablen x mit Hilfe eines linearen Modells erklären möchten:

$$\min_{\theta} \frac{1}{|I|} \sum_{(x,y) \in I} |y - \theta^T x|^2. \quad (\text{linReg})$$

Hier sind θ einfach die Koeffizienten des Normalenvektors der Hyperebene.

Das (ERM)-Problem ist zentral in ML und bringt viele mathematische Fragen mit sich, welche Gegenstand aktiver Forschung sind. Im Folgenden taucht mehrfach der Begriff „Neuronale Netze“ auf; die Lesenden mögen sich hier einfach eine kontinuierliche, relativ komplexe aber gutartige Funktion $f(\cdot, \theta) : \mathbb{R}^n \rightarrow \mathbb{R}$ vorstellen, die durch den Vektor θ parametrisiert ist und die als Funktion von θ ebenfalls gutartig und fast überall differenzierbar ist.

Wie kann (ERM) gelöst werden?

Für spezifische Problemformulierungen, wie die der linearen Regression, können oft geschlossene Lösungen (Pseudo-Inverse im Falle von linearer Regression) hergeleitet werden. Im Allgemeinen jedoch ist (ERM) ein nicht-konvexes Optimierungsproblem, welches z. B. mit Gradientenabstiegsverfahren lokal optimal gelöst werden kann. Dies ist z. B. der Fall, wenn f ein durch θ parametrisiertes Neuronales Netz ist. Viele Arbeiten in diesem Themenkomplex zielen darauf ab, neue Methoden zu finden, die effizienter Informationen aus den Daten in das Netzwerk übertragen, da oftmals in Anwendungen in der realen Welt Trainingsdaten nur kostenintensiv generiert werden können: z. B. um MRT Daten zum Training von medizinischen Diagnostik-Systemen in einer gewünschten Weise zu sammeln, müssen teure MRT Aufnahmen von Patienten gemacht werden. Andere Fragen drehen sich um die Effizienz der Trainingsmethoden (zumeist Varianten vom Gradientenabstieg) bei extremen Datengrößen und verteiltem Training. Hier zeigt sich zum Beispiel der Vorsprung von Tesla, wo Daten der gesamten Fahrzeugflotte aufgenommen und zum Training von Fahrerassistenz-

systemen benutzt werden, welche dann wiederum per *over-the-air* Update in der gesamten Fahrzeugflotte aktualisiert werden können: so ist ein dezentrales ML-System massiver Größe entstanden; ein globaler Supercomputer auf Rädern mit mehr als 800 000 Fahrzeugen²² (2019 und 2020 Zulassungen), jedes mit AI-optimierter Hardware ausgestattet.²³ Mathematisch gesehen ergeben sich spannende Fragen, was z. B. die Konvergenz und Effizienz eines asynchronen, dezentralen, stochastischen Gradientenabstiegs dieser Größe angeht. Tesla konnte das Problem der Trainingsdatengewinnung über *Scale* (sprich: schiere Größe) lösen. Diesen Luxus hat aber nicht jeder. Andere natürliche Fragen hier drehen sich um die Erforschung von Algorithmen mit einer höheren *sample-efficiency*, um mit deutlich weniger Trainingsdaten ML-Systeme trainieren zu können.

Wie gut ist eine Lösung?

Wenn (ERM) nicht-konvex ist, kann es viele global minimale Lösungen geben. Jedoch sind nicht alle Lösungen gleich „gut“. Was wir suchen, sind Lösungen θ , die gut auf neuen, ungesehenen Daten Vorhersagen treffen können. Dies bezeichnet man als Generalisierungsverhalten, und es ist bisweilen nur ansatzweise verstanden, welche Lösungen gut generalisieren. Der Mensch ist z. B. exzellent, was das Generalisierungsverhalten angeht: es reichen wenige geschriebene Beispiele des Buchstabens „A“ aus, damit der Mensch im Folgenden zuverlässig den Buchstaben „A“ erkennen kann, auch wenn er noch so unsauber geschrieben ist. Typische ML Systeme hingegen brauchen Millionen von solchen Trainingsbeispielen und moderate Abweichungen vom „Standard-A“ können zu einer Verwirrung des Systems führen. Viele Fragen in diesem Bereich führen zu Fragen in der Stochastik (wie gut approximiert die empirische Verteilung der Lösungen die unbekannte echte?), der Statistik (welchen *Bias*, sprich welche impliziten Annahmen, induziert das Modell?) und Topologie (wie sieht die Mannigfaltigkeit der Lösungen aus und welche Invarianten charakterisieren solche mit guter Generalisierung?).

Wie sieht die Struktur der Lösungen aus?

Komplexe Neuronale Netze können Milliarden von Parametern haben; damit ist θ in diesen Fällen extrem hochdimensional. Abgesehen von der Herausforderung, ein solches (ERM) Problem zu lösen, wenn man eine Lösung berechnet hat um z. B. Krebs in Röntgenbildern zu diagnostizieren, wie erhält der Arzt eine „Erklärung“ für die Diagnose? Im Gegensatz zu einfacheren Modellen wie linearer Regression können nicht einfach die Koeffizienten der Regression herangezogen werden. Eng damit verbunden sind Fragen der „Transparenz“ und „Verifikation“: wie stelle ich sicher, dass z. B. ein Neuronales Netzwerk das tut was es tun soll und nicht „irgendwo abguckt“? Dieses Problem lässt sich anhand des berühmten *Klugen Hans*²⁴ verdeutlichen: Anfang des 20. Jahrhunderts gab es ein Pferd „Hans“, trainiert durch den Mathematiklehrer Wilhelm von Osten, das anscheinend Rechenaufgaben durch Klopfen mit dem Huf lösen konnte. Es stellte sich jedoch heraus, dass das Pferd gelernt hatte, die Mimik der Zuschauer bzw. des Trainers zu „lesen“ und dann die richtige Antwort durch Mimik

Veränderung der Zuschauer erkennen konnte. Gelernt zu rechnen hatte Hans jedoch nicht; es sei angemerkt, dass eine 13-köpfige wissenschaftliche Kommission aus Experten eingesetzt wurde, um das Phänomen aufzulösen, es aber letztlich Kommissionsleiter Stumpf's Studenten Pfungst gelang, das Phänomen zu erklären. Phänomene dieser Art stellen im Machine Learning ein reales Problem dar, da z. B. Neuronale Netze mit erstaunlicher Effizienz Seitenkanäle wie Rauschen von verschiedenen Kameratypen in Bilddaten lernen können. Hat nun eine Person nur Hunde mit einem iPhone fotografiert und eine andere nur Katzen mit einem Android Phone, so kann ein Neuronales Netz augenscheinlich Katzen von Hunden unterscheiden gelernt haben; in Realität hat es aber vielleicht nur gelernt, die beiden Kameras anhand ihres Rauschmusters zu unterscheiden. Ebenso sind Fragen der „Fairness“ von Lösungen hier zu finden: Wie stellt man sicher, dass z. B. ein Neuronales Netzwerk in seinen Prognosen und Entscheidungen gewisse Personengruppen nicht systematisch diskriminiert oder bevorzugt?

Welche Funktionen können überhaupt gelernt werden?

Bei genauerem Hinschauen sieht man, dass es sich bei (ERM) um ein *Funktionsapproximationsproblem* handelt. Man versucht eine unbekannte Funktion zu approximieren, von der nur einzelne Funktionsauswertungen, unsere Daten I , bekannt sind. Damit führt die Reise schnell zu fundamentalen Problemen wie dem *Bias-Variance tradeoff*²⁵, eine Art „Unschärferelation“ für das Lernen: Gegeben eine Punktmenge I , so gibt es immer ein Polynom, welches die Daten perfekt erklärt bzw. approximiert mittels polynomieller Interpolation.²⁶ Aber ein solches Polynom wird ungesehene Daten nicht gut approximieren, da es nicht gut generalisiert (*overfitting*). Auf der anderen Seite kann man einfach eine Hyperebene durch die Daten legen. Ein solches Modell wird die Daten nur mäßig erklären (*underfitting*), ist aber robust bezüglich ungesehener Daten. Wie aber findet man ein Modell, was die Daten gut approximiert *und* ungesehene Daten gut erklärt? Und welche Daten (bzw. unbekannte Funktionen) können mit welchen Funktionsklassen gut approximiert werden? Hier sind viele der mathematischen Fragen ungeklärt und bisher gewonnene Einsichten zumeist empirischer Natur.

Mathematik im Dialog

Eine andere wichtige Aufgabe der Mathematik ist der Dialog mit benachbarten Feldern in der ML Forschung. Viele Ergebnisse und Erfolge basieren auf ad-hoc Ansätzen mit unzureichender mathematischer Grundlage. Dies führt dann zu der unbefriedigenden Situation, in der eine gewisse „Technik“ eine Aufgabe lösen kann, es aber unklar ist, warum dem so ist oder wie diese Technik in andere Bereiche übertragen werden kann.

Die Mathematik kann in diesem Dialog viel leisten und mitgestalten. Andernorts wird dies bereits gelebt. Die amerikanische National Science Foundation (NSF) hat vor Jahren die *Transdisciplinary Research In Principles Of Data Science (TRIPODS)* Initiative²⁷ ins Leben gerufen, um transdisziplinäre Forschungszentren zwischen Informatik, Statistik und Mathematik zu etablieren, und in der aktuellen Aus-

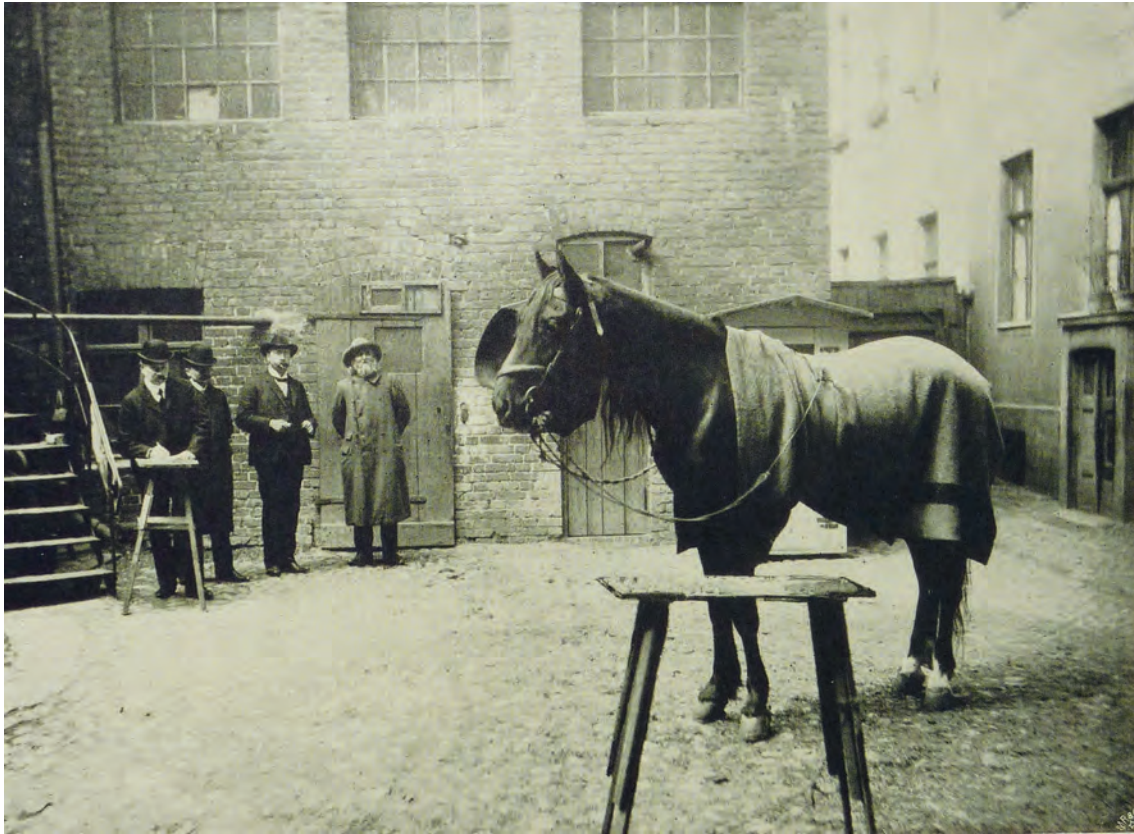


Abbildung: Wikimedia Commons

Der Kluge Hans wird geprüft. Im Hintergrund rechts Wilhelm von Osten (Karl Krall, *Denkende Tiere*. Leipzig 1912, Tafel 1)

schreibung für NSF AI Zentren²⁸ ist die Mathematik zentral verankert.

Machine Learning in Deutschland

Es gibt viele kluge Köpfe in Deutschland und ohne Frage hat Deutschland seine international renommierten ML-Leuchttürme in Personen und Institutionen. Darüber hinaus haben die Verantwortlichen und die Politik (im weiteren Sinne) verstanden, dass es sich bei ML um ein Feld handelt, welches für die technologische Entwicklung und den Wohlstand Deutschlands von zentraler Bedeutung ist. Es gibt verschiedenste Initiativen auf allen Ebenen, um Deutschland im Bereich ML stark zu machen und besser zu positionieren; allen voran die KI-Strategie der Bundesregierung. Die DFG hat kürzlich ein Schwerpunktprogramm (SPP 2298)²⁹ zum Thema „Theoretical Foundations of Deep Learning“ eingerichtet, um ML mit einem starken theoretischen Fundament zu untersetzen und es gibt verschiedenste ML-Initiativen u. a. vom BMBF³⁰ und BMWi³¹. Komplementär dazu wird die Initiative für Nationales Hochleistungsrechnen und der damit einhergehende NHR Verbund³² die Forschung (im Allgemeinen) mit massiven Compute Ressourcen unterstützen. Was den Austausch zwischen den Universitäten und der Industrie betrifft, werden Marktkräfte ihr Übriges tun. Es gibt viel Potenzial in Deutschland, welches es zu realisieren gilt. Insbesondere müssen die oben genannten Initiativen, mit entsprechenden flexiblen und

leichtgewichtigen Strukturen untersetzt werden. Eines der größten Unterschiede zwischen Deutschland und z. B. den USA ist der unfassbare administrative Aufwand³³ und die stark mangelnde Digitalisierung, dessen Implikationen im besonderen Maße in der Corona Pandemie sichtbar wurden. Wenn es Deutschland gelänge, sein Potenzial zu entfalten, hätten wir die Chance, in ML federführend zu werden.

Wie steht es denn um *Machine Learning* in Deutschland heute? Wenn man über den Ist-Zustand spricht, so ist die Antwort: Nicht gut. Deutschland hängt im internationalen Vergleich viele Jahre hinterher. Es fehlt ein breiter, nachhaltiger Austausch zwischen Industrie und Universitäten. Des Weiteren gibt es nur wenige Studiengänge, die wissenschaftlichen Nachwuchs relevant für ML ausbilden. So können Universitäten und Hochschulen ihre drei Kernaufgaben (1) Bildung, (2) Forschung und Innovation und (3) Transfer in Wirtschaft und Gesellschaft im Bereich ML (und damit letztlich auch Mathematik) nur bedingt erfüllen. Während (2) grundsätzlich gut aufgestellt ist, sind insbesondere (1) und (3) Schwachstellen. Darüber hinaus findet ML populärwissenschaftlich nicht die gleiche Beachtung wie andersorts, weshalb die meisten Referenzen hier aus dem englischsprachigen Raum stammen. Der populärwissenschaftliche Diskurs ist jedoch wichtig, um Jugendliche und Nachwuchs für ML und damit ein Stück mehr für die Mathematik zu begeistern.

Zwei Konsequenzen daraus sind: (a) Der Mangel des fachrelevanten Nachwuchses wird durch fachfremden kompensiert, was dann dazu führt, dass mathematisch-informatische Fragen unter anderem als ingenieurtechnische Fragen fehlinterpretiert werden: wenn man (nur) einen Hammer hat, muss alles ein Nagel sein. So lässt sich dann auch (aber nicht ausschließlich) erklären, dass Tesla beim Autonomen Fahren der traditionellen Automobilindustrie viele Jahre voraus ist;³⁴ wer einmal Einblick bei Tesla erhalten hat, weiß, dass es sich nicht nur um Jahre handelt. (b) Deutschland hat keinerlei Internet-Giganten wie Google, Facebook, Apple, Amazon, Microsoft, Alibaba, Huawei, Tencent vorzuweisen. Diese kooperieren aber eng mit Forschungseinrichtungen und tragen signifikant zur öffentlich frei verfügbaren (!) Grundlagenforschung im Bereich ML bei; Google Research³⁵ ist in einigen Statistiken³⁶ mit Abstand weltweit führend, was die Anzahl an Veröffentlichungen bei den Top ML Konferenzen „International Conference on Machine Learning (ICML)“ und „Conference on Neural Information Processing Systems (NeurIPS)“ angeht. Darüber hinaus versorgen diese Giganten mit ihren akademischen Programmen Universitäten und Forschungseinrichtungen mit (oft kostenlosen) Compute Ressourcen, die in der Forschung benötigt werden; dieses „giving back“ wird hierzulande kritisch gesehen, zugleich höre ich aber, dass Mathematiker-Kollegen ihre Modelle mangels Ressourcen nicht oder nur beschränkt rechnen können: die Forschung wird ausgebremst. In Zahlen basierend auf einem Index angelehnt an den Nature Index³⁷ und berechnet für Veröffentlichungen bei ICML und NeurIPS kommt Deutschland mit einem Wert von 91,5 auf Platz 6 in einer kürzlich veröffentlichten Statistik.³⁸ Der Spitzenreiter USA erreicht 1677,8 gefolgt von China (281,2) und UK (161,0); Google alleine erreichte 220,1 gefolgt von der Stanford University (106,1), MIT (99,6) und UC Berkeley (86,7). Damit liegt Deutschland als Ganzes gleichauf mit einzelnen Top US Universitäten; die erste deutsche Einrichtung auf der Liste ist die Universität Tübingen auf Platz 57 mit einem Wert von 11,9. Dies ist nur eine von vielen Statistiken, und ob man Veröffentlichungen zählen sollte, ist fraglich. Allerdings führen ähnliche Vergleiche auch mit anderen Statistiken und Maßen zu ähnlichen Ergebnissen. Der Kontrast wird nochmals deutlicher, wenn mit *Gesamtausgaben für Forschung und Entwicklung*³⁹ – im internationalen Vergleich mit *Total Research Expenditures (TRE)* bezeichnet – diesen obigen Ausgangsgrößen eine Eingangsgröße gegenüber gestellt wird.

Wie kommen wir also weiter? Pauschalantworten helfen nicht, vielleicht jedoch eines: Diskussionen hierzulande sind oft durch ein *Mindset of Scarcity* (Gier, Neid und Angst) geprägt, in denen alles als ein Nullsummenspiel erscheint. Eine Nahrung eines *Mindset of Abundance*, weg von einem Gegeneinander hin zu einem Miteinander, würde für viele der Probleme einen Lösungsweg bereiten. Denn die Mathematik in Deutschland hat durchaus das Potenzial, wie auch die historische Grundlage, dieses Miteinander aktiv zu gestalten.

Schluss

Abschließend lässt sich sagen, dass die deutsche Mathematik einen wesentlichen Beitrag in der ML Forschung leisten kann, wenn sie denn möchte. Etwas schärfer formuliert könnte man die Mathematik auch in der Verantwortung sehen, hier ihren Beitrag zu leisten und die für sie vorge-sehene Rolle einzunehmen, um ML mit einem rigorosen Fundament zu untersetzen. Deutschland hat, wenn es denn seine Karten richtig spielt, die Chance in der ML Forschung international wettbewerbsfähig zu sein. Es liegt aber auch noch einiges an Arbeit vor uns.

Danksagung

Ich bedanke mich sehr bei Heike Balluneit, Omid Nohadani, Marc Pfetsch und Tu-Lan Vu-Han für die hilfreichen Kommentare zu einer frühen Version dieses Artikels.

Anmerkungen

1. Ich benutze bewusst das Englische. Die deutschen Begriffe „Maschinelles Lernen“ und „Künstliche Intelligenz“ sind nicht nur Übersetzungen, sondern Reduktionismen, die nur teilweise mit dem globalen Verständnis von ML und AI kongruent sind.
2. Thomas Kuhn. *The Structure of Scientific Revolutions*. (1962)
3. NVIDIA Inpainting Demo, www.nvidia.com/research/inpainting/; Adobe Video Inpainting. /medium.com/syncedreview/smooth-exclusion-new-adobe-algorithm-aces-video-inpainting-3131a64c25ed.
4. MIT Technology Review. *The year deepfakes went mainstream*, www.technologyreview.com/2020/12/24/1015380/best-ai-deepfakes-of-2020/; *The Verge*. *All of these faces are fake celebrities spawned by AI*, www.theverge.com/2017/10/30/16569402/ai-generate-fake-faces-celebs-nvidia-gan
5. *Computer Vision applications in Self-Driving Cars, becominghuman.ai/computer-vision-applications-in-self-driving-cars-610561e14118*; *Computer Vision makes Autonomous Vehicles intelligent and reliable*, www.analyticsinsight.net/computer-vision-makes-autonomous-vehicles-intelligent-and-reliable/
6. *Nature Outlook, How AI is improving cancer diagnostics*, www.nature.com/articles/d41586-020-00847-2; *Siemens Health Solutions, AI-Rad Companion*, www.siemens-healthineers.com/digital-health-solutions/digital-solutions-overview/clinical-decision-support/ai-rad-companion
7. *Tensorflow Neural Style Transfer Tutorials*, www.tensorflow.org/tutorials/generative/style_transfer
8. *Is artificial intelligence set to become art's next medium?* www.christies.com/features/A-collaboration-between-two-artists-one-human-one-a-machine-9332-1.aspx
9. Christian B. Anfinsen, *Studies on the Principles that govern the Folding of Protein Chains*, www.nobelprize.org/uploads/2018/06/anfinsen-lecture.pdf
10. *Folding@Home Project* foldingathome.org
11. *DeepMind AlphaFold* deepmind.com/blog/article/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology
12. *Wikipedia, AlphaGo versus Lee Sedol*, en.wikipedia.org/wiki/AlphaGo_versus_Lee_Sedol; *Deepmind, AlphaGo Case Studies*, deepmind.com/research/case-studies/alphago-the-story-so-far
13. *OpenAI, Emergent tool use*, openai.com/blog/emergent-tool-use/
14. Dabney et al., *A distributional code for value in dopamine-based reinforcement learning*, *Nature* (2020), www.nature.com/articles/s41586-019-1924-6
15. *DeepL*, www.deepl.com
16. *Google Translate*, translate.google.com
17. *Google AI Blog, A Neural Network for Machine Translation, at Production Scale*, ai.googleblog.com/2016/09/a-neural-network-for-machine.html; *Wired, An Infusion of AI Makes Google Translate More Powerful Than Ever*, www.wired.com/2016/09/google-claims-ai-breakthrough-machine-translation/
18. *Forbes, What Is GPT-3 And Why Is It Revolutionizing Artificial*

Intelligence? www.forbes.com/sites/bernardmarr/2020/10/05/what-is-gpt-3-and-why-is-it-revolutionizing-artificial-intelligence/?sh=1ee94124481a; OpenAI, API, openai.com/blog/openai-api/; MIT Technology Review, OpenAI's new language generator GPT-3 is shockingly good – and completely mindless, www.technologyreview.com/2020/07/20/1005454/openai-machine-learning-language-generator-gpt-3-nlp/

19. Max Woolf. Train a GPT-2 Text-Generating Model w/ GPU (Colab Notebook). colab.research.google.com/drive/1VLG8e7YSEwypxU-noRNhsv5dW4NfTGce

20. The Guardian, A robot wrote this entire article. Are you scared yet, human? www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3

21. SIAM News, The Connection Between Applied Mathematics and Deep Learning (2020), sinews.siam.org/Details-Page/the-connection-between-applied-mathematics-and-deep-learning.

22. Statista, Number of Tesla vehicles delivered worldwide from 4th quarter 2015 to 4th quarter 2020, www.statista.com/statistics/502208/tesla-quarterly-vehicle-deliveries/

23. Zum Vergleich, der Supercomputer des ZIBs hat ca. 150,000 Kerne.

24. Wikipedia, Clever Hans, simple.wikipedia.org/wiki/Clever_Hans

25. Wikipedia, Bias-Variance tradeoff, en.wikipedia.org/wiki/Bias%20%80%93%20variance_tradeoff

26. Wikipedia, Polynomial Interpolation, en.wikipedia.org/wiki/Polynomial_interpolation

27. NSF. Transdisciplinary Research In Principles Of Data Science (TRIPDS).

28. NSF, National Artificial Intelligence (AI) Research Institutes, www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf2060429.

29. DFG SPP 2298, www.dfg.de/foerderung/info_wissenschaft/2020/info_wissenschaft_20_36/index.html

30. BMBF, Künstliche Intelligenz, www.bmbf.de/de/kuenstliche-intelligenz-5965.html

31. BMWi, Künstliche Intelligenz, www.bmw.de/Redaktion/DE/Artikel/Technologie/kuenstliche-intelligenz.html

32. NHR Geschäftsstelle, www.nhr-gs.de

33. Aus eigener Erfahrung schätze ich den Faktor bei ca. 5 ein, je nach Aufgabe.

34. Forbes, Tesla Is Years Ahead Of Competitors With No Signs Of Stopping, www.forbes.com/sites/moorinsights/2020/03/05/tesla-is-years-ahead-of-competitors-with-no-signs-of-stopping/?sh=7a7396759f45

35. Google Research. research.google

36. AI Research Rankings 2020, Can the United States Stay Ahead of China? chuvpilo.medium.com/ai-research-rankings-2020-can-the-united-states-stay-ahead-of-china-61cf14b1216

37. Es gibt einen Punkt pro Artikel, der entsprechend Koautoren etc. verteilt wird. Siehe [1] und [2] für genaue Definition.

38. Insights from the International Conference on Machine Learning (ICML 2020), chuvpilo.medium.com/whos-ahead-in-ai-research-in-2020-2009da5cd799; AI Research Rankings 2020, Can the United States Stay Ahead of China? chuvpilo.medium.com/ai-research-rankings-2020-can-the-united-states-stay-ahead-of-china-61cf14b1216

39. Wikipedia, List of countries by research and development spending, en.wikipedia.org/wiki/List_of_countries_by_research_and_development_spending

*Prof. Dr. Sebastian Pokutta, Zuse-Institut Berlin,
Takustraße 7, 14195 Berlin
pokutta@zib.de*

Sebastian Pokutta ist Professor für Mathematik mit Forschungsgebiet Optimierung und Machine Learning an der TU Berlin und Vizepräsident des Zuse-Instituts Berlin. Zuvor war er der David M. McKenney Family Associate Professor am Georgia Institute of Technology und Founding Associate Director des Center for Machine Learning at Georgia Tech (ML@GT). Sebastian Pokutta hat sein Studium der Mathematik und anschließende Promotion in Mathematik an der Universität Duisburg-Essen absolviert.



Drei Exemplare der Nanokakerlake – weitgehend unerforscht

(Foto: Christoph Eyrych)