



FOTO: SVEN DÖRING FÜR MPG

Die untereinander verknüpften Nervenzellen des Gehirns dienen als Vorbild für künstliche neuronale Netzwerke, die ein Team am Max-Planck-Institut für Mathematik in den Naturwissenschaften erforscht.

# INTELLIGENZ MIT PLAN

TEXT: THOMAS BRANDSTETTER

Selbstlernende Algorithmen sind dabei, unsere Gesellschaft gehörig umzukrempeln. Doch allzu oft verstehen ihre Entwickler selbst nicht genau, wie sie funktionieren. Mit grundlegenden Theorien zum maschinellen Lernen wollen Forschende des Max-Planck-Instituts für Mathematik in den Naturwissenschaften nun Abhilfe schaffen.

FOTO: REED HUTCHINSON / UCLA



Guido Montúfar forscht am Max-Planck-Institut für Mathematik in den Naturwissenschaften und an der University of California in Los Angeles.

Lernfähige Algorithmen übernehmen immer mehr Aufgaben, denen bis vor Kurzem nur Menschen gewachsen schienen. Kaum ein Onlinedienst kommt ohne sie aus, und sei es nur, um Werbung zu schalten. Künstliche Intelligenz, kurz KI, übersetzt Texte oder schreibt sie gleich selbst, so wie

der Chatbot ChatGPT, der aus unzähligen im Internet verfügbaren Texten verblüffend sinnvolle Antworten auf beinahe beliebige Fragen erstellt. Künstliche Intelligenz versucht sich aber auch als Künstlerin und macht manches Fahrerassistenzsystem zu einem umsichtigeren Verkehrsteilnehmer als die Person am Steuer. Der aktuelle Höhenflug von KI lässt sich vor allem mit der Weiterentwicklung künstlicher neuronaler Netze erklären, die in vieler Hinsicht dem menschlichen Gehirn nachempfunden sind. Sie werden am Computer als tief verzweigte Netzwerke künstlicher Neuronen simuliert, deren Verbindungen sich dynamisch an neue Erfahrungen anpassen, um so Muster in Daten zu erkennen oder neue Verhaltensweisen zu erlernen. Und ähnlich wie im Gehirn ist auch bei seinen elektronischen Pendanten oftmals nur schwer nachvollziehbar, was in deren Tiefen genau vor sich geht. Zwar gibt ihnen der Erfolg recht, und für viele Zwecke mag es auch vollkommen ausreichen, sie schlicht als nützliche Blackbox zu betrachten. Um aber das enorme Potenzial der Technik noch weiter auszuschöpfen, halten Fachleute es für nötig, KI endlich auf das Fundament einer soliden Theorie zu stellen. So könnten sie im Detail verstehen, wie die Algorithmen lernen. Auch Guido

Montúfar und sein Team vom Max-Planck-Institut für Mathematik in den Naturwissenschaften und von der University of California, Los Angeles, arbeiten auf dieses Ziel hin.

59

„Wir untersuchen die mathematische Seite künstlicher neuronaler Netze“, erklärt der Forscher. Vieles sei bisher hauptsächlich auf einer praktischen Ebene entwickelt worden, während es an Theorien dazu oftmals gefehlt habe. „Da ist vieles einfach ausprobiert worden“, sagt Montúfar. „Und zum Teil waren die Leute dann selbst überrascht, dass es funktioniert hat.“ Gemeinsam mit seinen Kolleginnen und Kollegen arbeitet der Mathematiker jetzt an einer Theorie neuronaler Netze.

## Große Netze auch für wenige Daten

Im Wesentlichen sind es drei Aspekte, die für den aktuellen Siegeszug der künstlichen Intelligenz verantwortlich zeichnen. So können die Entwickler auf immer mehr Rechenleistung in Form verbesserter Hardware in immer größeren Rechenzentren zurückgreifen. Daher ließen sich immer größere neuronale Netze ins



Werk setzen, in denen heute unzählige künstliche Neuronen vernetzt sind. So wie es im Gehirn auf die Zahl und die Stärke der Synapsen, der Verbindungen zwischen den Nervenzellen, ankommt, sind auch bei einem künstlichen neuronalen Netz die Verknüpfungen entscheidend. So sollen es etwa bei GPT-3, einem Vorläufer von ChatGPT, 175 Milliarden Verknüpfungen sein – im menschlichen Gehirn sind es 100 Billionen. Auftrieb hat KI auch bekommen, weil aufgrund der voranschreitenden Di-

nigen Trainingsdaten sinnvolle Ergebnisse liefern können“, sagt Montúfar. Früher habe man geglaubt, die hohe Flexibilität großer Netze sei nur bei entsprechend großen Datensätzen zielführend. Denn naturgemäß sind große neuronale Netze auch sehr komplex. Und dies könnte, so die lange Zeit vorherrschende Überzeugung, leicht zu falschen Entscheidungen führen, wenn sie mit nur wenigen Daten trainiert werden. „Das war eine Art Mantra, das eigentlich auch mathematisch fundiert ist“, erklärt Guido Montúfar. „Doch die Theorie, auf der dieser Glauben beruhte, war unvollständig.“ In der Praxis habe man dann selbst ohne Theorie schon sehr bald beobachtet, dass neuronale Netze umso besser funktionieren, je größer sie sind.

So könnte man also durchaus auch argumentieren, dass es eigentlich keine so große Rolle spielt, warum genau ein Netz funktioniert – Hauptsache, es funktioniert. „Vor zehn Jahren waren tatsächlich recht viele Leute davon überzeugt, dass man keine Theorie brauche“, sagt Montúfar. Das mag daran gelegen haben, dass ein theoretisches Verständnis in manchen Fällen tatsächlich nicht unbedingt nötig ist. Eine mathematische Theorie zu entwickeln liegt nun mal nicht jedem, selbst wenn er erfolgreich KI-Algorithmen programmiert. Doch wer keine Theorie zur Hand hat, auf der er aufbauen kann, muss stattdessen eben vieles einfach ausprobieren. Das ist umständlich und kostet Zeit und Ressourcen. Man stelle sich nur einen Chemiker vor, der keine Ahnung hat von Elementen und Molekülen und auf gut Glück verschiedene Substanzen mischt. Außerdem könnte KI zunehmend auch in sensiblen Bereichen eingesetzt werden und etwa medizinische Diagnosen unterstützen oder aber Fahrzeuge autonom lenken. Da wüsste man schon gerne, wie genau sie ihre Entscheidungen trifft.

Darüber hinaus gibt es trotz all der beeindruckenden Erfolge künstlicher Intelligenz wahrlich noch genug Verbesserungsbedarf. Besonders beunruhigend sind zum Beispiel die Gefahren durch böswillige Angriffe etwa beim autonomen Fahren. Sie

können das KI-System, das Verkehrsschilder erkennt und interpretiert, austricksen und damit in der Folge Unfälle verursachen. Schon eine minimale, aber gezielte Veränderung an einem Schild kann unter Umständen dazu führen, dass das System anstelle einer tatsächlichen Beschränkung auf 80 Kilometer pro Stunde eine Beschränkung auf 200 Kilometer zu erkennen glaubt. Für das menschliche Auge hingegen wäre die Veränderung an dem Schild kaum zu bemerken. Doch warum kann ein künstliches neuronales Netz auf eine so fatal falsche Spur gebracht werden? „Um diese Frage zu beantworten und letztendlich solche Attacken zu verhin-

---

## AUF DEN PUNKT GEBRACHT

Wie künstliche neuronale Netze lernen, eine Aufgabe zu lösen, und welche Kriterien sie dabei anwenden, ist meist unklar. Das ist unter anderem bei Anwendungen von künstlicher Intelligenz in der Medizin oder bei der Bilderkennung für autonomes Fahren problematisch.

60 Eine Theorie künstlicher neuronaler Netze würde helfen, deren Entscheidungsfindung nachvollziehbar zu machen. Sie könnte außerdem die Suche nach geeigneten Algorithmen beschleunigen.

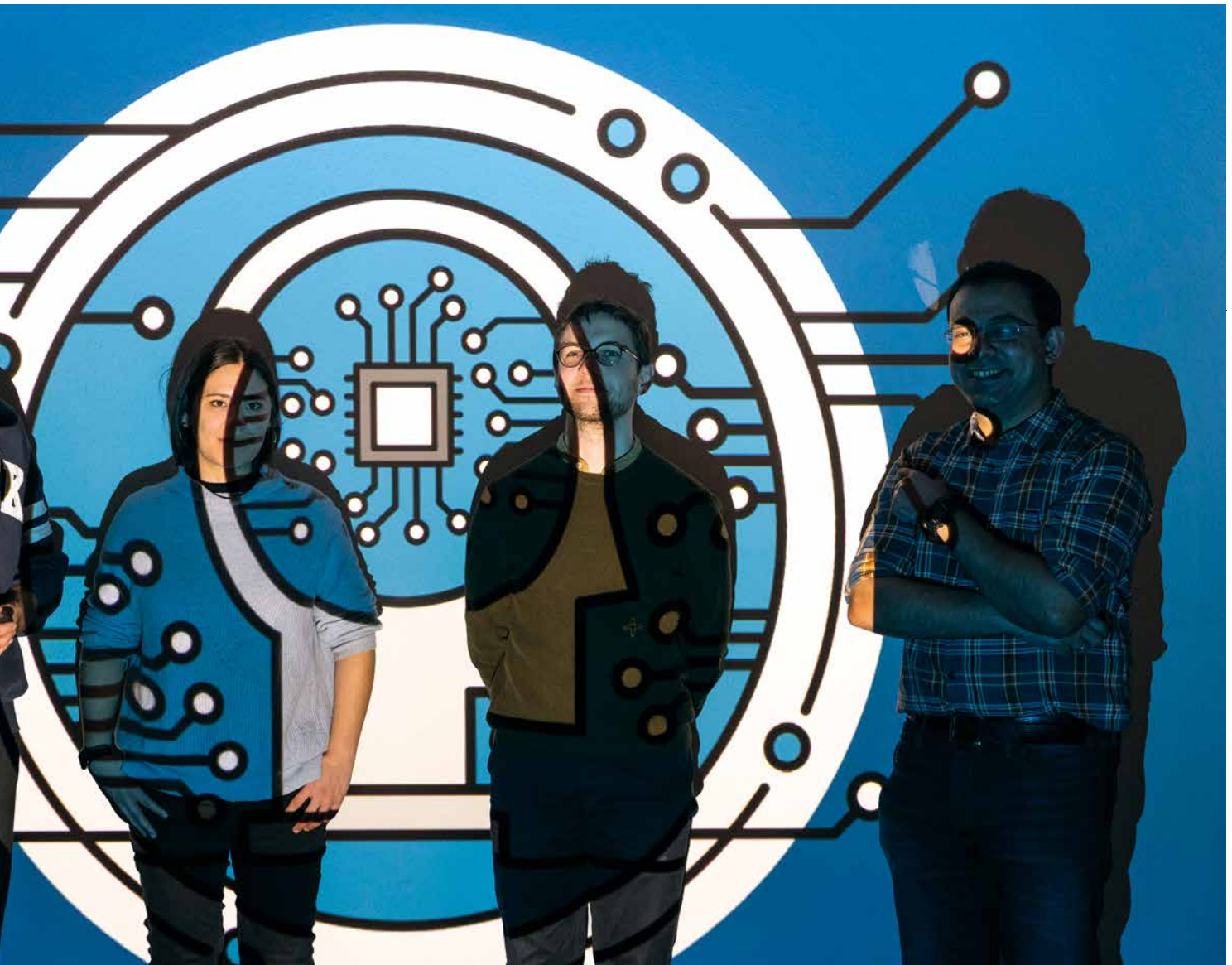
Beim Training neuronaler Netze werden mathematische Funktionen identifiziert, die ein Problem lösen. Im Normalfall sind dafür viele Funktionen geeignet, jedoch nicht alle gleich gut. Damit die optimale Funktion gefunden wird, sollte eine Theorie auch andere Eigenschaften der Funktionen berücksichtigen als nur ihre Fähigkeit, das Trainingsproblem zu lösen.

---

gitalisierung viel mehr Daten für das Training der künstlichen neuronalen Netze zur Verfügung stehen. Und schließlich bringt die weltweite Zusammenarbeit bei der Entwicklung neuer Algorithmen auch immer ausgefeiltere Software hervor. „Eine der Überraschungen in der Entwicklung von KI war zum Beispiel, dass große neuronale Netze auch mit relativ we-

FOTO: SVEN DÖRING FÜR MPG





dern, braucht es ein tiefer gehendes Verständnis über die Funktionsweise der Netze“, sagt Montúfar.

## Verständnis für mehr Datenschutz

Und auch beim Schutz der Privatsphäre liegt aufgrund unserer Unwissenheit noch vieles im Argen. So werden etwa Maschinen, die vorhersagen sollen, ob und wie schnell sich ein Patient in einem Krankenhaus erholen wird, vorher mit den Daten real existierender Patienten trainiert. „Diese Daten stecken also irgendwie in dem System

drin“, gibt Montúfar zu bedenken. „Und der Kunde, der dieses System gekauft hat, könnte versuchen, es aufzumachen und diese privaten Daten herauszulesen.“ Um garantieren zu können, dass so etwas nicht möglich ist, müsse man schon sehr genau wissen, was im Inneren eines solchen Systems vor sich geht.

Aktuell beschäftigt sich die Gruppe für Mathematical Machine Learning mit dem Phänomen, dass neuronale Netze bevorzugt Lösungen mit gewissen mathematischen Eigenschaften finden. „Auch wenn es mehrere Möglichkeiten gibt, wird sich das Netz letztlich meist für Lösungen mit

Erhellung für die künstliche Intelligenz: Johannes Müller, Marie-Charlotte Brandenburg, Pierre Bréchet und Pradeep Kumar Banerjee (von links) arbeiten an einer Theorie künstlicher neuronaler Netze, die helfen soll, die Funktionsweise von KI besser zu verstehen.

gewissen Eigenschaften entscheiden“, sagt Montúfar. „Und wir versuchen zu charakterisieren, wie diese Präferenzen genau aussehen.“ Den Zusammenhang veranschaulicht der Forscher, indem er die Suche nach einer geeigneten Lösung mit einer Wanderung durch eine Landschaft vergleicht. Das Netz beziehungsweise der Algorithmus, mit dem es arbeitet, beginnt seine Suche an einem vom





Musterlösungen: Die einzelnen Grafiken entsprechen unterschiedlichen Konfigurationen eines neuronalen Netzes, das etwa auf Bildern Tische identifizieren soll. Je größer die Zahl von Farbflächen in einer Grafik, desto komplexer ist die zugehörige Funktion, mit der das Netz die Aufgabe löst. Die Leipziger Mathematiker erforschen, von welchen Bedingungen die Komplexität der ermittelten Funktion abhängt.

den Ausgangspunkt der Suche nicht mehr länger dem Zufall zu überlassen, sondern dafür systematisch geeignete Bereiche zu wählen, wollen die Forschenden wissen, wie die Landschaft der Lösungen aussieht. „Wir haben in dem Bereich bereits viele Fortschritte gemacht“, sagt Montúfar. „Jetzt geht es darum, diese Erkenntnisse in eine präzise mathematische Theorie umzusetzen, die konkrete Vorhersagen über das Verhalten eines neuronalen Netzes unter verschiedenen Bedingungen ermöglicht.“

## Eine Theorie für schnelleres Training

Eine Rolle spielt dabei auch, dass Theorien zu komplexen Phänomenen oft zu stark vereinfacht werden und dabei wichtige Aspekte des eigentlichen Sachverhaltes verloren gehen. So sind mathematische Theorien zum Beispiel oft weniger kompliziert, wenn die künstlichen neuronalen Netze, die sie beschreiben, der Einfachheit halber als unendlich groß betrachtet werden. Doch auch wenn der Trend tatsächlich in Richtung immer größerer Netze geht, sind sie in Wirklichkeit natürlich trotzdem endlich. „Solche Vereinfachungen können große Probleme verursachen“, sagt Montúfar. Denn in den zwar riesigen, aber doch endlichen Netzen treten dann unvorhergesehene Effekte auf, die sehr signifikant sein können. „Wenn wir die verstehen und erklären könnten, könnten wir diese KI-Systeme in der Praxis viel schneller trainieren“, meint Montúfar. Und dies würde nicht nur sehr viel Strom und Geld sparen, sondern auch Zeit. Schließlich arbeiten die Supercomputer, auf denen diese Algorithmen laufen, oft wochen- oder gar monatelang.

Entwickler mehr oder weniger zufällig gewählten Ausgangspunkt. Deshalb neigt es dazu, Lösungen aufzuspüren, die sich in der Nähe dieses Punktes befinden. Das neuronale Netz ist dabei gewissermaßen in der direkten Umgebung der nahe liegenden Lösung gefangen. „Wir wollen nun verstehen, wie die Nachbarschaft um diese Lösung aussieht und warum das neuronale Netz da nicht herauskommt“, erklärt Montúfar.

Die Aufgabe der künstlichen Intelligenz könnte beispielsweise darin bestehen, einen Zusammenhang zwischen Daten zu Wohnflächen und Mietpreisen herzustellen und dafür eine geeignete

mathematische Funktion zu finden. Eine vernünftige Lösung würde nicht nur einfach aus einer Linie, genauer gesagt einer mathematischen Funktion, bestehen, die die einzelnen Datenpunkte möglichst gut miteinander verbindet. Um eine plausible Abhängigkeit eines mit zunehmender Wohnungsgröße kontinuierlich steigenden Mietpreises darzustellen, sollte diese Linie auch möglichst glatt sein, also keine Knicke aufweisen. „Ein geeigneter Bereich, um die Suche in der Landschaft der Lösungen zu starten, wären also alle Funktionen, die sehr glatt sind“, erklärt Montúfar. „Umgekehrt wäre es verrückt, nach Linien zu suchen, die total zackig sind.“ Um

Ein weiterer wichtiger Fokus der Arbeit von Montúfars Team liegt auf den Daten selbst, mit denen die Maschinen trainiert werden. Ähnlich wie bei lebendigen Gehirnen hat ihre Qualität erheblichen Einfluss auf die Entwicklung eines künstlichen neuronalen Netzes. So stellte sich zu Anfang der 1960er-Jahre in einer Studie mit jungen Katzen, die wegen eines Augenfehlers von Geburt an nur verschwommen sehen konnten, heraus, dass der Schaden irreparabel war. Denn selbst wenn das Augenproblem später behoben wurde, konnten die Tiere nicht scharf sehen. „Wer das Sehen mit verschwommenen Bildern gelernt hat, dessen Gehirn kann auch im Alter nichts mehr mit scharfen Bildern anfangen“, fasst Montúfar das Ergebnis zusammen. Und auch in der Informatik wirkt sich die Art und Weise, wie Trainingsdaten gestaltet sind, auf den Lernprozess aus. „Wir wollen wissen, welche Effekte das sind“, sagt der Forscher, „also welchen Einfluss die Auswahl der Trainingsdaten auf die Entwicklung und später dann die Leis-

tungsfähigkeit eines Netzes hat.“ Spinnt man die aktuelle Entwicklung künstlicher Intelligenz weiter, könnte sich irgendwann auch die Frage nach einem künstlichen Bewusstsein stellen. „Wenn man davon ausgeht, dass Bewusstsein existiert, sehe ich keinen Grund, warum es dieses nicht auch in künstlichen neuronalen Netzen geben sollte“, meint Montúfar pragmatisch. Viele Forscher träumen schon jetzt davon, künstliche Intelligenz so weit zu entwickeln, dass sie nicht mehr bloß für einzelne Spezialaufgaben taugt, sondern ein umfassendes Verständnis der Welt entwickelt. Eine solche künstliche allgemeine Intelligenz wäre dann einer Vielzahl unterschiedlicher intellektueller Aufgaben gewachsen und würde den Menschen über kurz oder lang wohl auch in jeder davon überflügeln.

Allerdings bleibt die Frage, wann und wie wir denn entscheiden werden, ob eine allgemeine KI erreicht wurde. „Früher hätte man vielleicht einen Spamfilter schon als intelligent be-

## GLOSSAR

### KÜNSTLICHES NEURONALES NETZ

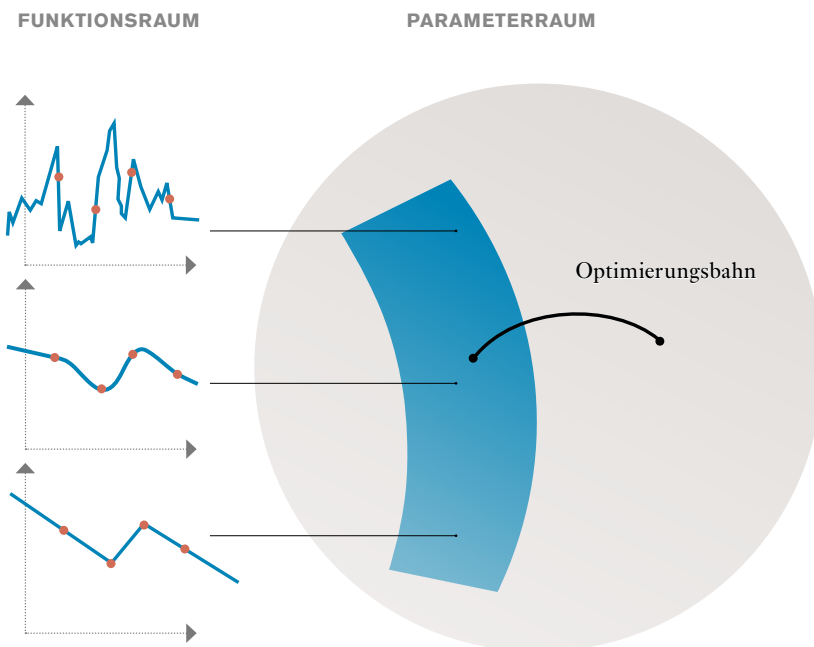
heißt ein am Computer simuliertes System, in dem zahlreiche künstliche Neuronen so vernetzt sind wie die Nervenzellen des Gehirns. Künstliche Neuronen erzeugen aus einem Eingabewert einen Ausgabewert und geben diesen an ein anderes Neuron weiter. Beim Training neuronaler Netze werden die Verbindungen zwischen den Neuronen so geknüpft und verstärkt, dass eine Aufgabe optimal gelöst wird. Die Verknüpfungen entsprechen dabei Kriterien, die bei der Suche nach Lösungen für ein Problem zu berücksichtigen sind und nach ihrer Relevanz gewichtet werden.

zeichnet, heute beeindruckt das niemanden mehr“, sagt Montúfar. Die Grenze verschiebt sich also ständig, und tatsächlich ist allgemeine KI nur schwer zu definieren. Wie viele Aufgaben soll dieses System lösen können, damit es die Bezeichnung allgemeine KI verdient? „Manche Leute haben ja bereits Tests formuliert und Benchmarks definiert, wie man das entscheiden soll“, sagt Montúfar. „Ich denke aber, es wird sich letztlich in der Praxis zeigen, ob Leute ein System als allgemeine KI wahrnehmen oder nicht.“

„Womöglich läuft es mit Computern aber auch ähnlich wie mit dem Zeugen von Kindern“, meint Montúfar. „Das sind auch intelligente Wesen, und um sie zu erschaffen, müssen wir eigentlich gar nichts verstehen.“ So könnten wir womöglich irgendwann eine allgemeine KI erschaffen, ohne genau zu wissen, was wir da eigentlich tun. „Die Frage ist, was das dann für ein Wesen sein wird“, gibt Montúfar zu bedenken. „Und wenn es irgendwie mit unserer Gesellschaft und unseren Begriffen von Zivilisation kompatibel sein soll, wäre es schon sehr gut, wenn wir es auch verstehen würden.“ Doch dazu müssen eben erst einmal die theoretischen Grundlagen von KI bekannt sein.

63

GRAFIK: GCO NACH GUIDO MONTÚFAR / MPI FÜR MATHEMATIK IN DEN NATURWISSENSCHAFTEN



Beim Lernen passt ein künstliches neuronales Netz seine Parameter, quasi seine synaptischen Verbindungen, entlang der Optimierungsbahn an und sucht den blau gefärbten Bereich der Funktionen, die alle Datenpunkte (rot) abdecken. Die mittlere Funktion mit möglichst wenigen Maxima und Minima sowie ohne Knicke löst die Aufgabe sehr gut.

