Liebe Leserinnen, liebe Leser,

wir freuen uns, Ihnen nach einer längeren, Corona-bedingten Pause eine neue Ausgabe des Forschungsmagazins ‚Blick in die Wissenschaft' in der Ausgabe 44/45 präsentieren zu können.

Die Corona-Pandemie hat auch die Universität Regensburg und alle ihre Mitglieder vor große Herausforderungen gestellt. Dennoch konnten zentrale Zukunftsprojekte weitergeführt und umgesetzt werden. So stellt vor allem die Gründung unserer neuen Fakultät für Informatik und Data Science (FIDS) einen wahren Meilenstein in der Geschichte und Entwicklung der Universität Regensburg dar. Als größtes Strukturprojekt seit der Gründung der Fakultät für Medizin vor 30 Jahren ist unsere Informatikfakultät ein Zukunftsprojekt von weitreichenden Dimensionen. Mit der neuen strategischen Schwerpunktsetzung im Bereich Informatik und Data Science und vor allem auch der Querschnittsorientierung der neuen Fakultät sieht sich die Universität Regensburg sehr gut gerüstet, ihre bisherigen Stärken in diesen Bereichen zu bündeln, weiter auszubauen und zu vertiefen. Schließlich sind *Digital Transformations* als eines der vier Gestaltungsfelder und Zukunftsthemen in unserem *Hochschulentwicklungsplan 2025* fest verankert. Dieses Gestaltungsfeld adressiert die neue Fakultät ebenso wie den Bereich *Integrated Sciences in Life, Health, and Disease* als ein weiteres Schwerpunktgebiet unserer Universität.

Die Grundsatzbeschlüsse in den Gremien der Universität Regensburg im Sommer und Herbst 2019 zur Einrichtung der neuen Fakultät erfolgten nach einer vorhergehenden Phase intensiver Planungen dann letztlich fast zeitgleich mit der Regierungserklärung des Bayerischen Ministerpräsidenten Dr. Markus Söder am 10. Oktober 2019 und der Verkündigung der Hightech Agenda Bayern. Unterstützt und beschleunigt durch die Mittel der Hightech Agenda Bayern konnte der Auf- und Ausbau der Fakultät für Informatik und Data Science zügiger umgesetzt werden, nachdem die neue Fakultät im März 2020 formal gegründet und im Laufe des WS 2021/22 aus sich heraus handlungs- und funktionsfähig wurde. Im Mai 2022 konnten wir gemeinsam mit Ministerpräsident Dr. Markus Söder und Staatsminister für Wissenschaft und Kunst Markus Blume den offiziellen Kickoff für die Fakultät begehen. Dass dieser komplexe Prozess im Kontext der Herausforderungen der Corona-Pandemie vollzogen und abgeschlossen werden konnte, ist ein Zeichen für die Bedeutung dieser gesamtuniversitär-strategischen Maßnahme und für den Rückhalt für das Großprojekt in der universitären Gemeinschaft.

Im Laufe des Gründungsprozesses ist es gelungen, die verschiedenen Informatiknahen und -interessierten Kräfte der Universität an einen Tisch zu bringen und gemeinsam ein zukunftsorientiertes Konzept für die Fakultät zu entwickeln. Ein externes Gutachten mit hochrangiger Expertise skizzierte und evaluierte 2019 wesentliche inhaltliche Schwerpunkte und Strukturierungen für die neue Fakultät, an denen sich in den Jahren 2019-2021 die von Vizepräsident Prof. Dr. Nikolaus Korber geleitete Gründungskommission in der

konkreten Arbeit zum Aufbau der Fakultät orientierte. In insgesamt 15 Berufungsverfahren wurden die ersten neuen Professuren in der Fakultät zügig besetzt – ein Prozess, der in Kürze abgeschlossen sein wird. Im Besetzungsprozess hat sich vor allem auch gezeigt, wie attraktiv die Neugründung einer Fakultät und die Möglichkeiten zur Mitgestaltung und zum Aufbau neuer Strukturen für Wissenschaftlerinnen und Wissenschaftler sind und wie viel Zukunftspotential von unserer neuen Fakultät ausgeht. So konnten wir zum Wintersemester 2023/24 130 Studierende für den B.Sc. Informatik und den B.Sc. Data Science begrüßen.

Im vorliegenden Heft von ‚Blick in die Wissenschaft' möchten wir Ihnen nunmehr vor allem die Forschungsaktivitäten der Fakultät für Informatik und Data Science näher vorstellen. Dabei beglückwünsche ich die Fakultät, dass sie bereits eineinhalb Jahre nach ihrer vollständigen Handlungs- und Funktionsfähigkeit und während der weiteren Planungen zum Aufbau und der Ausarbeitung ihrer Studiengänge insbesondere im Master-Bereich ein so vielfältiges Themenheft zu ihren aktuellen Forschungsarbeiten vorlegen konnte.

Das facettenreiche und vielfältige Themenspektrum dieses Sonderhefts illustriert, wie die Fakultät für Informatik und Data Science die an der Universität Regensburg bisher vorhandenen IT-Kompetenzen erfolgreich bündelt und in die Zukunft gerichtet erweitert. Sie ermöglicht die essentielle interdisziplinäre Vernetzung mit der gesamten Universität, von den Geistes- und Sozialwissenschaften bis zu den Natur- und Lebenswissenschaften. Die Beiträge verdeutlichen, wie interdisziplinäre Forschung das Fundament starker methodischer und fachlicher Grundlagen weiterentwickelt und wie die bisherigen Informatik-Schwerpunkte der Universität Regensburg (Computational Science, Informationswissenschaft, Medieninformatik, Wirtschaftsinformatik) erfolgreich in die neue Fakultät überführt werden konnten und die Querschnittsorientierung unterstützen.

Das vorliegende Heft mit seinem Schwerpunkt auf aktuellen Forschungsarbeiten begleitet den im Wintersemester 2023/24 erfolgten Start der beiden grundständigen Bachelor-Studiengänge Informatik und Data Science. Ein kurzer Überblicksbeitrag zur Lehre in der neuen Fakultät zeigt anschaulich die bereits gewachsene Vielfalt der Informatikstudiengänge und die intensive und gelebte Verbindung von Forschung und Lehre auch an dieser neuen Fakultät.

Unsere neue Fakultät leistet hervorragende Arbeit und ich bin sicher, dass Ihnen die nachfolgenden Seiten einen spannenden Einblick in die verschiedenen Facetten der FIDS geben werden. Ein ganz besonderes Dankeschön möchte ich an dieser Stelle an das gesamte Dekanat der Fakultät für Informatik und Data Science und insbesondere an Forschungsdekanin Prof. in Dr. Meike Klettke richten, die für diese Sonderausgabe die Koordinationsarbeit der vorliegenden Ausgabe federführend übernommen hat.

Prof. Dr. Udo Hebel
Präsident der Universität Regensburg

Blick in die Wissenschaft
Forschungsmagazin
der Universität Regensburg
ISSN 0942-928-X
Heft 44/45
31. Jahrgang

Herausgeber
Prof. Dr. Udo Hebel
Präsident der Universität Regensburg

Redaktionsleitung für diese Ausgabe
Prof.in Dr. Meike Klettke / Fakultät für Informatik und Data Science

Redaktionsbeirat
Prof. Dr. jur. Christoph Althammer
Prof. Dr. rer. nat. Ferdinand Evers
Prof. Dr. rer. nat. Stefan Friedl
Prof. Dr. rer. nat. Mark W. Greenlee
Prof. Dr. theol. Andreas Merkt
Prof. Dr. phil. Omar W. Nasim
Prof. Dr. rer. nat. Klaus Richter
Prof. Dr. rer. pol. Daniel Rösch
Prof. Dr. med. Ernst Tamm
Prof. Dr. paed. Oliver Tepner
Prof. Dr. phil. Christiane Heibach

Universität Regensburg
93040 Regensburg
Telefon +49 941 9432300
Telefax +49 941 9433310

Verlag
Universitätsverlag Regensburg GmbH
Leibnizstraße 13, 93055 Regensburg

Telefon +49 941 78785-0
Telefax +49 941 78785-16

info@univerlag-regensburg.de
www.univerlag-regensburg.de
Geschäftsführer: Dr. Albrecht Weiland, Felix Weiland M.A.

Abonnementservice
bestellung@univerlag-regensburg.de

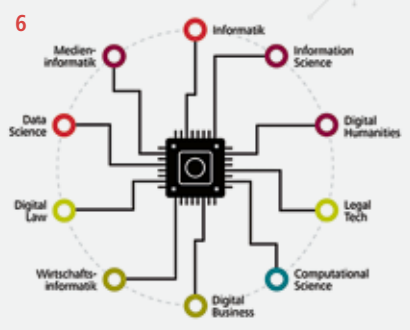
Anzeigenleitung
Larissa Nevecny
MME-Marquardt
info@mme-marquardt.de

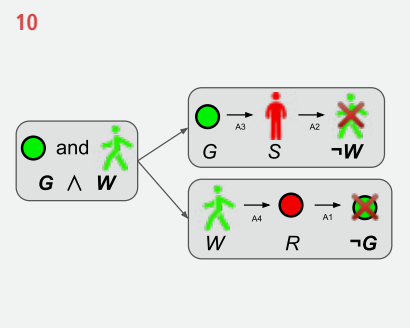
Herstellung
Universitätsverlag Regensburg GmbH
info@univerlag-regensburg.de

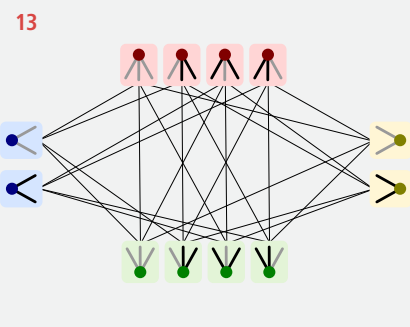
Einzelpreis € 7,00
Doppelheft € 14,00

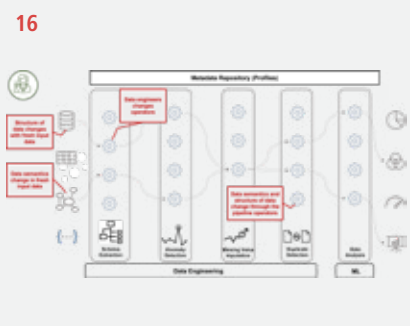
Jahresabonnement
bei zwei Ausgaben pro Jahr
€ 10,00 / ermäßigt € 9,00
Für Schüler, Studierende und Akademiker/innen im Vorbereitungsdienst (inkl. 7 % MwSt.) zzgl. Versandkostenpauschale € 1,64 je Ausgabe. Bestellung beim Verlag. Für **Mitglieder des Vereins der Ehemaligen Studierenden der Universität Regensburg e.V.**, des **Vereins der Freunde der Universität Regensburg e.V.** und des **Vereins ehemaliger Zahnmedizinstudenten Regensburg e.V.** ist der Bezug des Forschungsmagazins im Mitgliedsbeitrag enthalten.

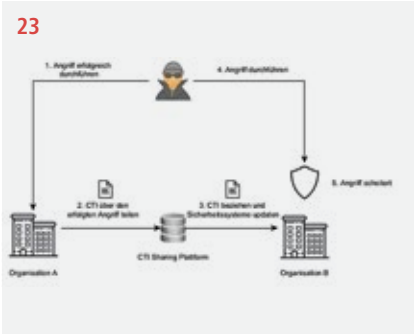
Inhalt

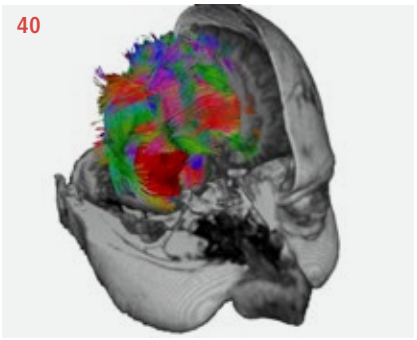


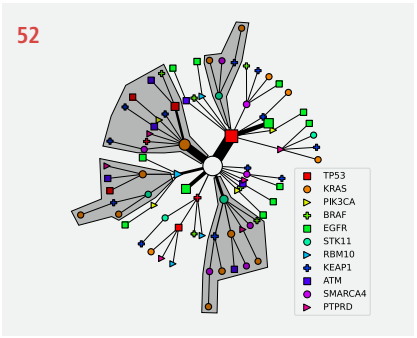












Kann man den Entscheidungen Künstlicher Intelligenz trauen?

Zu den Auswirkungen unsicherer Daten auf die Entscheidungen Neuronaler Netze

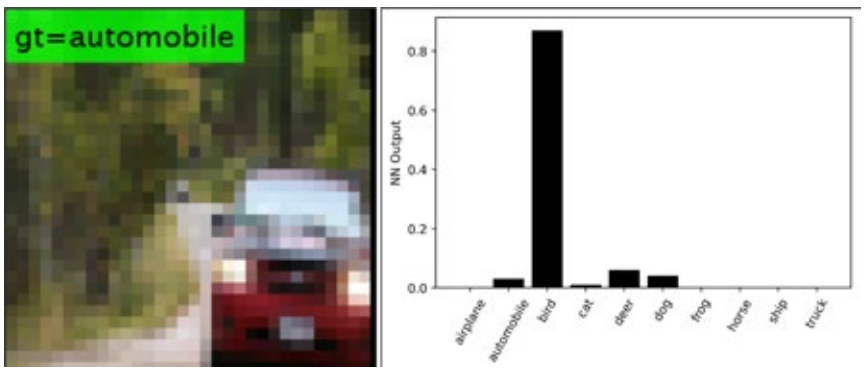
Thomas Krapf, Prof. Dr. Bernd Heinrich

Im Laufe der letzten Jahre haben sich maschinelle Lernverfahren stetig weiterentwickelt und sind zunehmend in der Lage, immer anspruchsvollere und komplexere, reale Aufgaben zu bewältigen. So basieren beispielsweise neue Verfahren beim automatisierten Fahren, in Industrieprozessen, in der medizinischen Diagnostik oder in der maschinellen Textverarbeitung wesentlich auf Neuronalen Netzen. Ein geläufiges Beispiel mag hier das Sprachmodell GPT und darauf basierend ChatGPT sein, das mit interaktiv generierten Texten von menschen-ähnlicher Qualität auf sich aufmerksam machte. Insgesamt sind Neuronale Netze nicht nur ein zentrales Thema für Forschung und Industrie geworden, sondern erhalten zunehmend Einzug in unseren persönlichen Alltag.

Soll ein Neuronales Netz für eine Klassifikationsaufgabe trainiert werden, so lernt es auf Basis einer größeren Datenmenge Muster und Zusammenhänge zwischen den Attributwerten der Dateninstanzen und dem Auftreten der jeweiligen tatsächlichen Klasse. Die für das Training und die Nutzung maschineller Lernverfahren verwendeten Daten liegen in der Realität jedoch oft mit Datenqualitätsdefekten vor. Beispielsweise sind Daten durch Mess-, Übertragungs- oder Erfassungsfehler inkorrekt, sie veralten im Zeitverlauf, sie sind unvollständig oder inkonsistent. Schlechte Datenqualität kann sich auf die Güte der Klassifikation oder Regression maschineller Lernverfahren stark auswirken (»garbage in, garbage out«), wie in experimentellen Analysen bereits mehrfach gezeigt wurde (z. B. Budach et al. 2022). Die Folge sind inkorrekte und wenig robuste Entscheidungen der Lernverfahren. Eine Illustration dieser Problematik aus dem Kontext der Bilderkennung (z. B. beim automatisierten Fahren) wird in Abbildung 1 dargestellt, welche die Klassifikation eines Objektes auf der unscharfen Aufnahme einer Fahrzeugkamera durch ein klassisches Neuronales Netz veranschaulicht.

Die Dateninstanz (hier die Kameraaufnahme) ist aufgrund schlechter Licht- oder Wetterverhältnisse verpixelt und damit von schlechter Datenqualität. Dadurch klassifiziert das Neuronale Netz (z. B. LeNet) das Objekt in der Aufnahme (eindeutig) als »Vogel« anstatt als »Automobil«. Diese Prognose beeinflusst unmittelbar die nächsten Aktionen des Fahrzeugs und kann damit schwere Unfälle mit Personenschäden zur Folge haben (vgl. auch den tödlichen Unfall mit einem Tesla-Fahrzeug (NHTSA 2017) und die mangelhafte Bilddatenqualität). Die Problematik schlechter Datenqualität und ihr Einfluss auf die Güte und Robustheit von Entscheidungen Neuronaler Netze betrifft jedoch nicht nur Bilddaten, sondern auch Daten anderer Form. So sind beispielsweise in strukturierten Daten (wie medizinischen Patientendaten oder Sensordaten in Industrieprozessen) sehr oft fehlende oder inkorrekte Werte enthalten.

Um schlechte Datenqualität strukturiert zu berücksichtigen, kann die resultierende (sogenannte aleatorische) Unsicherheit über die wahre (aber unbekannte) Ausprägung eines Wertes modelliert und in den Lernprozess oder die Entscheidungsfindung einbezogen werden. Diese Modellierung erfolgt durch Wahrscheinlichkeitsverteilungen, welche Informationen über mögliche Ausprägungen des unsicheren Wertes sowie über deren jeweiligen Wahrscheinlichkeiten enthält. Klassische Neuronale Netze können allerdings mit wahrscheinlichkeitsbasierten Dateninstanzen nicht umgehen. Sie verarbeiten sichere Werte im Input zu einer punktweisen Entscheidung (z. B. für eine Klasse), meist basierend auf Werten einer Softmax-Funktion. D. h. für jede Klasse wird ein Wert ausgegeben (vgl. Y-

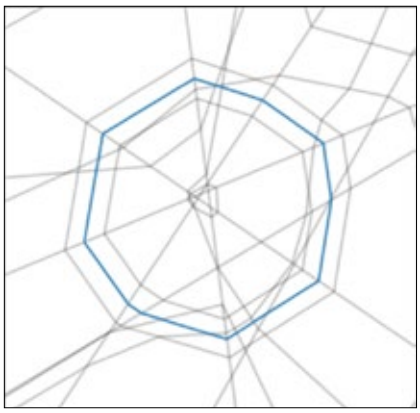


1 Prognose eines Neuronalen Netzes für ein unscharfes Bild (in Anlehnung an Gast & Roth 2018)

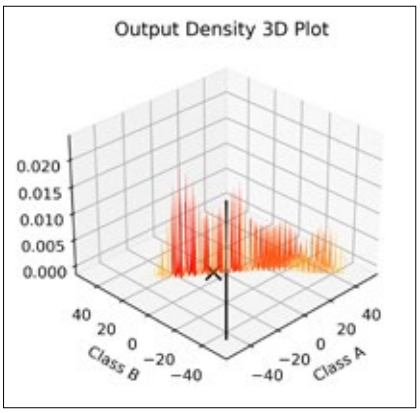
Achse in Abbildung 1) und die Klasse mit dem höchsten Softmax-Wert wird als Entscheidung gewählt. Die Softmax-Funktion neigt jedoch bekannterweise zur Über-Konfidenz (vgl. auch Abbildung 1). Im Gegensatz dazu erzeugt die Berücksichtigung von datenqualitäts-induzierter Unsicherheit in den Inputdaten eines Neuronalen Netzes ebenso eine Wahrscheinlichkeitsverteilung im Outputraum des Netzes. Mit Hilfe dieser können für die jeweiligen Klassen nunmehr fundierte Wahrscheinlichkeiten ausgegeben werden. Dies hat Vorteile: Wird vereinfacht ein Neuronales Netz z. B. zur Unterstützung der Diagnose einer Krankheit herangezogen, so wäre es bei einer Verteilung von 40% für die Klasse »Patient besitzt Krankheit« und 60% für die Klasse »Patient besitzt Krankheit nicht« dringend geboten, nicht einfach die Klasse mit der höheren Wahrscheinlichkeit als korrekt anzunehmen. Vielmehr sind aufgrund der nicht unerheblichen Wahrscheinlichkeit von 40% einer Krankheit weitere Untersuchungen anzuraten. Dies zeigt insbesondere auch die Wichtigkeit einer fundierten Konfidenz von Prognosen, vor allem in risikoreichen Bereichen wie der medizinischen Diagnostik oder dem automatisierten Fahren, in denen die Gesundheit von Personen betroffen ist. Daher sind Methoden nötig, um die Unsicherheit in den Inputdaten eines Neuronalen Netzes möglichst exakt auf dessen Outputraum abzubilden.

Die exakte Propagation von Unsicherheit in Inputdaten, repräsentiert mittels Wahrscheinlichkeitsdichten durch ein Neuronales Netz, welches mathematisch eine Verkettung von linearen Transformationen und nicht-linearen Aktivierungsfunktionen ist, bringt einige methodische Herausforderungen mit sich. Für probabilistischen oder teilweise probabilistischen Input ist eine solche Propagation ungleich schwieriger und sogar analytisch nicht möglich, da für die Abbildung beliebiger Dichten durch nicht-lineare Funktionen im Allgemeinen keine geschlossene Lösung existiert. Daher müssen andere Methoden entwickelt werden, um Wahrscheinlichkeitsdichten exakt durch Neuronale Netze zu propagieren.

Ein Ansatz mit der komplexen, nicht-linearen Struktur von Neuronalen Netzen umzugehen, ist eine alternative Betrachtungsweise des Netzes als mathematische Funktion. So kann ein Neuronales Netz, welches grundsätzlich als iterative Verkettung von einzelnen Schichten (jeweils bestehend aus linearen Transformationen und nicht-linearen Aktivierungsfunktionen) definiert ist, auch als eine stückweise lineare Funktion auf dessen Inputraum aufgefasst werden (vgl. Sattelberg et al. 2020). Dies bedeutet, dass das Netz auf polytop-förmigen Teilmengen des Inputraums durch eine (affin) lineare Abbildung gegeben ist. Diese sog. linearen Regionen teilen somit den Inputraum disjunkt auf, wie in Abbildung 2 schematisch zu sehen ist.



2 Lineare Regionen im Inputraum eines Neuronalen Netzes (schematische Abbildung, Sattelberg et al. 2020)



3 Dichte der Output-Verteilung

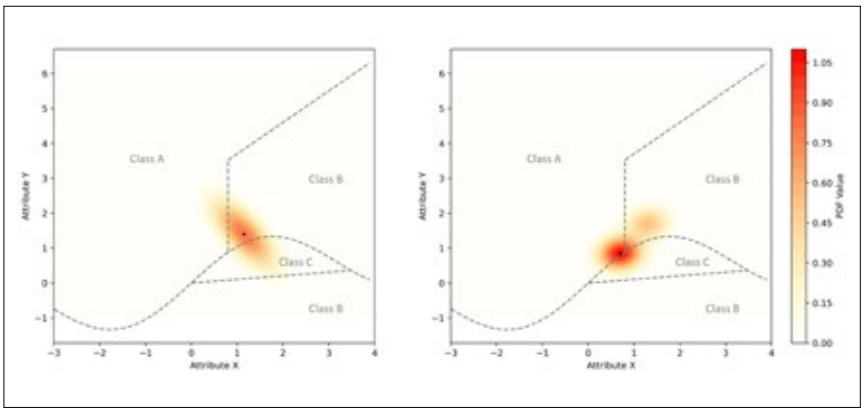
Bildnachweis: Lehrstuhl Wirtschaftsinformatik II

Diese Repräsentation des Netzes ist in unserer Methode für eine exakte Propagation nutzbar, indem zunächst die linearen Regionen des Netzes identifiziert werden und darauf basierend die Inputdichte (in Teilmassen) zerlegt wird. Die linearen Regionen und die darin enthaltenen Teilmassen der Inputdichte können nun exakt durch das Netz propagiert werden, da auf jeder linearen Region das Netz durch eine lineare Abbildung gegeben ist. Da sich die propagierten linearen Regionen im Output überschneiden können, müssen zuletzt noch deren Schnittmengen identifiziert und die propagierten Teilmassen der Inputdichte geeignet aggregiert werden. Als Ergebnis unserer Methode ergibt sich damit eine exakt propagierte Dichte im Outputraum des Netzes, welche als fundierte Basis zur Einschätzung der Unsicherheit einer Entscheidung des Neuronalen Netzes dient.

Neben der exakten Propagation von Wahrscheinlichkeitsdichten durch ein gegebenes Netz kann in diesem Kontext auch die Frage nach dem Training eines Neuronalen Netzes bei unsicheren Daten untersucht werden. Auch diese Fragestellung umfasst mehrere methodische Herausforderungen, insbesondere, dass sich Dichten oft über mehrere Bereiche des Inputraums erstrecken, die mit verschiedenen, auch inkorrekten Klassen im Output korrespondieren. Um zu verhindern, dass dadurch falsche Zusammenhänge erlernt werden, kann auch hier eine Aufteilung des Inputraumes und des Netzes in sogenannte Mixture of Experts-Netze, welche ausschließlich auf Teilmengen des Inputraums trainiert werden, hilfreich sein.

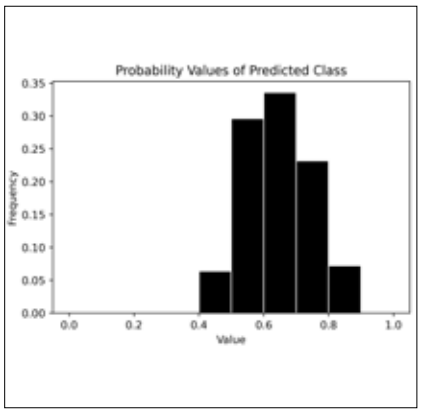
Aufgrund der exakten Propagation unsicherer Inputdaten mittels unserer Methode ergibt sich eine Reihe von Vorteilen, die hier auszugsweise diskutiert werden. Zur Illustration wird ein einfach gehaltenes Klassifikationsproblem zugrunde gelegt, das basierend auf nur zwei stetigen Datenattributen X und Y eine Zuordnung von Dateninstanzen durch ein Neuronales Netz in drei Klassen A, B und C betrachtet (z. B. die Klassifikation einer Patientin in »Hat Diabetesrisiko«, »Hat kein Diabetesrisiko« und »Hat Diabetes«, auf Basis der Attribute »Alter« und »Körpergewicht in Kg«).

1) Transparenz über die tatsächlichen Klassenwahrscheinlichkeiten: Die durch die exakte Propagation erhaltene Dichte im Output des Netzes ermöglicht es, die Wahrscheinlichkeiten für alle Klassen fundiert abzuleiten. In Abbildung 3 ist die propagierte Dichte für eine einzelne Dateninstanz unseres einfachen Klassifikationsproblems dargestellt. Die Dichte erstreckt sich über die Bereiche der beiden Klassen A und B im Outputraum (die Klasse C hat für die Instanz eine Wahrscheinlichkeit von 0 und bleibt daher aus Darstellungsgründen unberücksichtigt). Zudem ist die Entscheidungsgrenze (schwarze Linie) eingezeichnet, welche diese Bereiche trennt. Es ist zu erkennen, dass der Peak der Verteilung zwar in Klasse B ist, jedoch die größere, dichtere Masse den Bereich von Klasse A einnimmt. Folglich besitzt die Klasse B ca. 38 % Wahrscheinlichkeit, die Klasse A ca.



Bildnachweis: Lehrstuhl Wirtschaftsinformatik II

4 Vergleich zweier Dateninstanzen mit identischen Klassenwahrscheinlichkeiten



Bildnachweis: Lehrstuhl Wirtschaftsinformatik II

5 Wahrscheinlichkeiten für die Entscheidungen der Testinstanzen

62 %. Die Entscheidung muss demnach offensichtlich als unsicher eingestuft werden. Diese Unsicherheit bliebe durch klassische, nicht probabilistische Analysen verborgen bzw. es können sogar falsche Entscheidungen resultieren: Würde nämlich – wie üblich – nur ein als sicher angenommener einzelner Wert (vgl. in Abbildung 3 z. B. den Modalwert als schwarzes Kreuz) als Input für das Netz verwendet werden, so würde dieser zur falschen Entscheidung »Klasse B« führen (die Instanz ist tatsächlich Klasse A). Erschwerend signalisiert das klassische Netz eine eindeutige Entscheidung, da die sich ergebenden Softmax-Werte der Klassen (0, 1, 0) lauten. D. h. es wird eine hohe Konfidenz angezeigt. Da keine weitere Information über die Unsicherheit dieser Entscheidung vorliegt, würde infolgedessen z. B. das tatsächliche Diabetesrisiko der Patientin nach dieser Analyse unerkannt bleiben.

2) Lage der Wahrscheinlichkeitsdichte und robuste Entscheidungen: Neben den reinen Klassenwahrscheinlichkeiten können mehrere inhärente Eigenschaften der Dichte, wie die generelle Verteilung der Wahrscheinlichkeitsmasse, deren Streuung oder deren Entfernung zu Entscheidungsgrenzen, analysiert werden. Die Abbildung 4 zeigt dies für zwei weitere Dateninstanzen unseres einfachen Klassifikationsproblems. Da wir nun den Inputraum betrachten, sind an den Achsen nicht mehr die Klassen (wie in Abbildung 3) abgetragen, sondern die Datenattribute X und Y, deren Unsicherheit im Inputraum sowie die Entscheidungsgrenzen zwischen den Klassen A, B, und C (grau gestrichelte Linien). Beide Dateninstanzen weisen gerundet mit (0.28, 0.40, 0.32) identische Klassenwahrscheinlichkeiten auf und wären damit nach Punkt 1) als vergleichbar zu interpretieren. Allerdings liegen die Wahrscheinlichkeitsdichten der beiden Dateninstanzen sehr unterschiedlich im Raum. Während die gesamte Wahrscheinlichkeitsmasse der linken Dateninstanz sehr nahe bei den oder sogar auf den Entscheidungsgrenzen zwischen den drei Klassen liegt, befindet sich ein größerer Teil der Wahrscheinlichkeitsmasse der rechten Dateninstanz tiefer im Bereich der Klasse B. Eine Erhöhung oder Reduzierung der Werte der beiden Datenattribute in diesem Bereich führen demnach zu keiner Änderung der Klassenentscheidung. Die rechte Instanz weist dahingehend eine höhere Robustheit auf als die linke Dateninstanz. Robustheit stellt in diesem Zusammenhang eine zentrale Güteeigenschaft Neuronaler Netze dar. Daneben wird eine weitere potenzielle Problematik evident: Klassische Neuronale Netze entscheiden i. d. R. nach dem Prinzip der relativen Mehrheit, d. h. die Klasse mit dem höchsten Softmax-Wert wird gewählt (s. o.). Hier sind die berechneten Softmax-Werte auf Basis der Modalwerte der beiden Dateninstanzen mit gerundet (0, 1, 0) eindeutig. Dies signalisiert eine hohe Robustheit der Entscheidung, welche sowohl durch die tatsächlichen Klassenwahrscheinlichkeiten der beiden Instanzen von (0.28, 0.40, 0.32) als auch durch die Lage der Wahrscheinlichkeitsdichte nicht gerechtfertigt ist. Allein schon durch die Betrachtung der Klassenwahrscheinlichkeiten wird klar, dass die beiden nicht-gewählten Klassen einen Großteil der Wahrscheinlichkeitsmasse (ca. 60%) ausmachen, d. h. ihr Eintreten ist insgesamt wahrscheinlicher als der von Klasse B.

3) Globale Robustheit der Entscheidungen eines Neuronalen Netzes: Die bisher auszugsweise vorgestellten Vorteile beziehen sich auf die Analyse von Entscheidungen für einzelne Instanzen. Die Wahrscheinlichkeiten für die Klassen lassen sich aber auch auswerten, um die Entscheidungen des Neuronalen Netzes für alle Dateninstanzen (»global«) hinsichtlich ihrer Robustheit zu beurteilen. Ordnet man für alle Testinstanzen im betrachteten Klassifikationsproblem die Wahrscheinlichkeiten der getroffenen Entscheidungen (d. h., das Maximum der Klassenwahrscheinlichkeiten) Intervallen zu, so ergibt sich Abbildung 5. Hier sieht man, dass z. B. ca. ein Drittel aller Entscheidungen nur mit einer Klassenwahrscheinlichkeit knapp über 50 % gefällt wurden. Insgesamt ist die Robustheit der Entscheidungen des Neuronalen Netzes basierend auf der gegebenen hohen Datenunsicherheit (schlechte Datenqualität) wenig überzeugend. Hier werden Schwächen sichtbar, die infolge der reinen Betrachtung der Softmax-Werte (hier werden über 95 % aller Entscheidungen mit einem Softmax-Wert nahe eins getroffen) nicht transparent werden.

In den letzten Jahren hat die Frage nach der Robustheit und Nachvollziehbarkeit KI-basierter Prognosen stark an Bedeutung gewonnen. Da KI immer mehr auch bei risikoreichen Entscheidungen zur Unterstützung oder als Entscheidungsinstanz herangezogen wird, muss vor allem in diesen Einsatzgebieten neben der Prognosegüte auch ein Maß für die Robustheit berücksichtigt und in ein verantwortungsvolles Handeln einbezogen werden. Ein solches Maß ist Grundvoraussetzung, dass das Vertrauen in die Prognosen maschineller Lernverfahren wächst und damit ihr Potenzial voll ausgeschöpft werden kann. So können Personen (z. B. Ärzte) Neuronale Netze zur Ergänzung und Entlastung auch in kritischen Domänen unterstützend heranziehen, aber auch die Unsicherheit der

Entscheidung berücksichtigen, um individuell auf dieser Basis das weitere Vorgehen zu bestimmen.

Die grundsätzliche Bedeutung dieser Thematik für Individuen, Gesellschaft und Industrie wurde auch bereits erkannt. So berät die EU im Rahmen des »AI-Act«-Gesetzes eine gemeinsame Position und Strategie zur Regulierung des Einsatzes von KI-Technologien. Dieser Rechtsrahmen soll Regeln festlegen, um sicherzustellen, dass KI-Systeme transparent und verantwortungsvoll eingesetzt werden. Herausforderungen wie Datenqualität und maschinelles Lernen basierend auf unsicheren Daten sind explizit genannt. Gleichzeitig fördert der AI-Act weitere innovative Entwicklungen, da Leitlinien für die Entwicklung und Nutzung von KI-Systemen bereitgestellt werden sollen, um eine positive Integration von KI in verschiedene Lebensbereiche grundsätzlich zu ermöglichen.

Referenzen

Budach, L.; Feuerpfeil, M.; Ihde, N.; Nathansen, A. Noack, N.; Patzlaff, H., Naumann, F.; and Harmouch, H. 2022. The Effects of Data Quality on Machine Learning Performance. arXiv preprint arXiv:2207.14529.

Gast, J. and Roth, S. 2018. Lightweight Probabilistic Deep Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*: 3369–3378. IEEE Computer Society.

NHTSA. 2017. PE 16-007. Tesla Crash Preliminary Evaluation Report. U.S. Department of Transportation, National Highway Traffic Safety Administration. Technical report.

Sattelberg, B.; Cavalieri, R.; Kirby, M.; Peterson, C.; and Beveridge, R. 2020. Locally Linear Attributes of ReLU Neural Networks. arXiv preprint arXiv:2012.01940.

Foto © privat

Prof. Dr. Bernd Heinrich
ist seit 2011 Inhaber des Lehrstuhls für Wirtschaftsinformatik II an der Universität Regensburg. Er promovierte an der Universität St. Gallen/Schweiz und habilitierte sich an der Universität Augsburg. Seine Forschungsschwerpunkte liegen in den Bereichen Data Science, Data Quality & Data Uncertainty, Process Planning & Process Science sowie Service Systems

Foto © privat

Thomas Krapf, M. Sc.,
Thomas Krapf, geboren 1997 in Amberg. 2015–2019 Studium der Mathematik mit Nebenfach Betriebswirtschaftslehre und Aktuarwissenschaften an der Universität Regensburg mit Vertiefung in den Bereichen der partiellen Differentialgleichungen und angewandten Analysis. Seit Januar 2021 Tätigkeit als wissenschaftlicher Mitarbeiter und Doktorand am Lehrstuhl für Wirtschaftsinformatik II der Universität Regensburg. Seine Forschungsschwerpunkte umfassen die Bereiche Data Quality und Data Uncertainty in Machine Learning.