1 Einführung in die künstliche Intelligenz und das maschinelle Lernen

Ribana Roscher und Lukas Drees

1.1 Was ist künstliche Intelligenz?

Die künstliche Intelligenz (KI, engl.: Artificial Intelligence – AI) hat sich in der Öffentlichkeit zu einem wichtigen und weitverbreiteten Thema entwickelt und wird mittlerweile im Kontext einer Vielzahl von Anwendungsbereichen genannt. Die Auffassung von KI ist vielfältig und reicht von Techniken zur Datenanalyse und -interpretation, die sogenannte schwache KI (engl.: Artificial Narrow Intelligence), bis hin zu der Vorstellung einer dem Menschen ähnlichen oder sogar überlegenen Intelligenz, der sogenannten starken KI (engl.: Artificial General Intelligence) und künstlichen Superintelligenz (engl.: Artificial Super Intelligence). Den aktuellen Stand der Technik stellt die schwache KI dar, auf die auch im weiteren Verlauf des Buchs fokussiert wird. Ein Grund für diese breite Auffassung ist das Fehlen einer offiziellen Definition, die unter anderem durch die schnelle und dynamische Entwicklung des KI-Felds begründet ist (Monett et al. 2020).

Seit Mitte des letzten Jahrhunderts ist die KI ein aktives Forschungsfeld, zu dem Wissenschaftler wie John McCarty und Alan Turing einen wesentlichen Beitrag geleistet haben. Bereits in den frühen Anfängen forschten beide an der Idee, dass Maschinen die menschliche Intelligenz simulieren können, wobei John McCarty 1956 auf einer Konferenz am Dartmouth College (USA) erstmals den Begriff KI einführte. Das Feld durchging seitdem mehrere Erfolgsperioden mit Durchbrüchen, gefolgt von Rückschlägen und Ernüchterungen mit stagnierenden Forschungsfortschritten, den sogenannten „AI winters“. Stetige Weiterentwicklungen von Hardware und neuen Algorithmen geben dem Feld jedoch wiederkehrenden Aufwind. Eine bedeutende Entwicklung, die bis in die heutige Forschung Einzug hält, ist der Wechsel von reinen Expertensystemen, bei denen durch Formeln und meist simple, designte Regeln Expertenwissen zur Lösung einer Aufgabe verwendet wird, hin zu datengetriebenen Modellen, die aus Beobachtungen lernen.

KI ist mit weiteren Disziplinen wie Data Science (deutsch: Datenwissenschaften) eng verknüpft. Data Science umfasst mehrere Disziplinen, wie maschinelles Lernen, Statistik und Datenmanagement, mit dem Ziel, durch wissenschaftliche Ansätze Wissen und Erkenntnisse aus Daten zu extrahieren. Ein wesentlicher Aspekt dieses Gebiets ist der umfangreiche Einsatz von Domänenwissen. Es gibt verschiedene Arten von Wissen mit unterschiedlichem Formalitätsgrad, die von hoch formalisiertem, naturwissenschaftlichem Wissen über Wissen aus technischen oder Produktionsprozessen bis hin zu Weltwissen und schließlich individueller (Experten-)Intuition reichen. In der Geodäsie wird Wissen oft in Form von mathematischen Gleichungen, wie analytischen Ausdrücken oder Differenzialgleichungen, oder als Beziehungen zwischen Instanzen und/oder Klassen in Form von Regeln oder Einschränkungen repräsentiert. Wissen kann außerdem durch numerische Simulationsmodelle oder durch menschliche Interaktion eingebracht werden.

Maschinelles Lernen wird oft mit KI gleichgesetzt, ist jedoch vielmehr ein Teilgebiet von KI. Maschinelle Lernmethoden ermöglichen es, Muster in Daten zu erkennen und eine Vorhersagefunktion zu ermitteln, um damit eine bestimmte Aufgabe wie die Detektion von Objekten oder die Zuordnung von Objekten zu einer Klasse zu lösen. Eine spezielle Art des maschinellen Lernens, welches vor allem in den letzten Jahren viel Aufmerksamkeit erhalten hat, ist das Deep Learning (deutsch: tiefes Lernen) (Goodfellow et al. 2016, Camps-Valls et al. 2021). Es wird vorrangig im Zusammenhang mit neuronalen Netzen verwendet. Das Wort „tief" bezieht sich auf die Komplexität der maschinellen Lernmodelle, d. h., die Vorhersagefunktion umfasst nicht selten mehrere Millionen Parameter, ist häufig verschachtelt und nichtlinear. Data Science und im Speziellen maschinelles Lernen spielen eine zentrale Rolle im KI-Bereich.

1.2 Maschinelles Lernen

1.2.1 Das generelle Framework

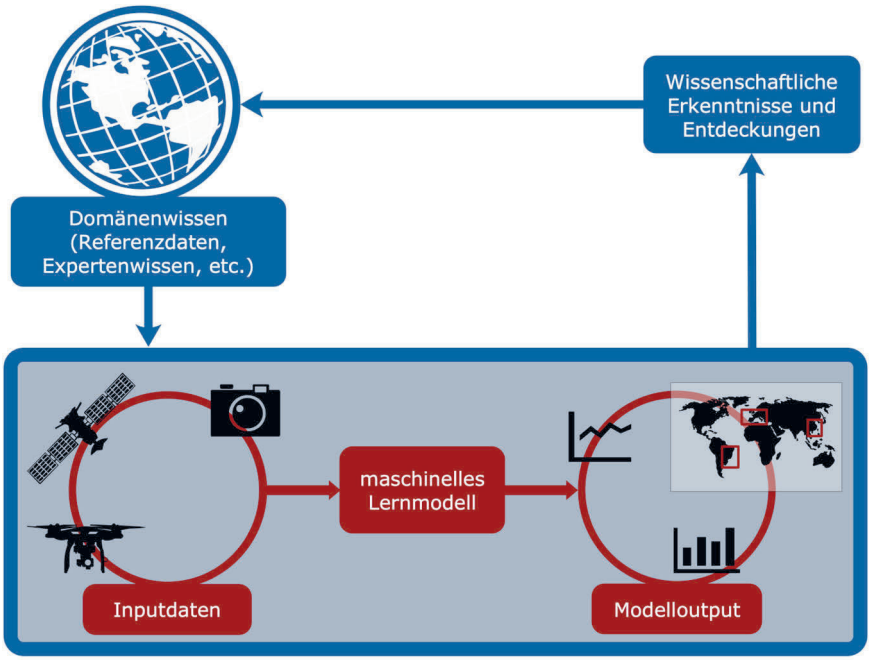

Abb. 1.1: Allgemeines Framework des maschinellen Lernens (eigene Darstellung)

Wie in Abbildung 1.1 dargestellt, sind im Allgemeinen Framework Inputdaten gegeben, beispielsweise durch geodätische Beobachtungen unterschiedlicher Dimension:

- 1D: Messwerte (Tachymeter, Nivellier, Spektrometer),
- 2D: Bilder (Roboter, UAV, Satellit),
- 3D: Punktwolken (Laserscanner).

Zusätzlich zu den Inputdaten können Referenzoutputs gegeben sein, die sogenannten Zielvariablen, die mit dem Input Trainingsdatenpaare bilden. In der Trainingsphase wird aus den Trainingsdatenpaaren oder alleinig aus den Inputs ein Modell gelernt, dessen Kern eine Vorhersagefunktion ist. Die Parameter der Vorhersagefunktion werden durch Optimierung geschätzt, sodass die Funktion den Input auf den Referenzoutput abbildet. Handelt es sich um vorher ungesehene Inputdaten, ist dies die Testphase. Falls der Output kategorisch ist, handelt es sich um eine Klassifikation und der Output wird als Label oder Annotation bezeichnet. Falls der Output kontinuierlich ist, handelt es sich um eine Regressionsanalyse. Eine oft verwendete Definition des maschinellen Lernens ist nach Mitchell (1997) wie folgt gegeben:

„Ein Computerprogramm lernt durch Erfahrung E in Bezug auf eine Klasse von Aufgaben T und ein Leistungsmaß P, wenn sich seine Leistung bei Aufgaben in T, gemessen durch P, mit Erfahrung E verbessert."

Basierend auf dieser Definition sind im Laufe der Zeit verschiedene Arten des maschinellen Lernens in Abhängigkeit von der Natur der Aufgaben T entstanden, die unterschiedliche Leistungsmaße P verwenden, um zu bewerten, wie gut die Aufgaben erfüllt sind. Die Aufgaben greifen auf unterschiedliche Arten der Erfahrung E zurück, welche im einfachsten Fall durch die Trainingsdatenpaare gegeben sind.

Eine rein datengetriebene Bestimmung der Vorhersagefunktion durch maschinelles Lernen kann zu Modelloutputs führen, die inkonsistent oder unplausibel sind. Maschinelle Lernansätze, die sich im Speziellen mit der Schätzung wissenschaftlich konsistenter und plausibler Ergebnisse beschäftigen oder Vorwissen über den funktionalen Zusammenhang zwischen Daten und Modelloutput ausnutzen, werden dem Theory-guided Machine Learning und dem (Physics-)Informed Machine Learning zugeordnet (Von Rueden et al. 2019, Karniadakis et al. 2021, Karpatne et al. 2017). Darüber hinaus können aus dem Modelloutputs und dem Modell selber weitere wissenschaftliche Erkenntnisse abgeleitet werden, indem der Schätz- und Entscheidungsprozess analysiert wird. Mit solchen Verfahren beschäftigt sich das Explainable Machine Learning (Roscher et al. 2020).

Klassische maschinelle Lernmethoden wie logistische Regression (Berkson 1944), Random Forests (Breiman 2001) oder Support Vector Machines (Steinwart & Christmann 2008) verwenden häufig statt der originalen Inputdaten eine daraus abgeleitete Repräsentation, die aus manuell extrahierten Merkmalen besteht. Dabei ist das Ziel, eine Repräsentation zu schaffen, die besser für die zu lösende Aufgabe geeignet ist als die originale Datenrepräsentation. Es werden beispielsweise Merkmale extrahiert, die eine hohe Diskriminierungsfähigkeit zur Klassifikation besitzen, eine Verringerung der Komplexität des Modells darstellen oder eine bessere Interpretierbarkeit oder Visualisierbarkeit haben. Für Bilder werden beispielsweise oft spektrale Merkmale wie Textureigenschaften verwendet und für Punktwolken räumliche Merkmale, wie die mittlere Abweichung von Punkten von einer Ebene in einer kleinen Umgebung. Im Zuge des Vormarschs von neuronalen Netzen wurde die manuelle Wahl einer geeigneten Merkmalsrepräsentation größtenteils abgelöst. Grund hierfür ist, dass bei neuronalen Netzen die Merkmalsextraktion und das Lernen der Vorhersagefunktion gemeinsam während des Lernprozesses passiert, sodass die Repräsentation entsprechend der zu lösenden Aufgabe optimiert wird. Dadurch entfällt die oftmals schwierige und suboptimale manuelle Wahl der Merkmale, allerdings ist die gelernte Repräsentation oft schwieriger zu interpretieren.

1.2.2 Taxonomie des maschinellen Lernens

Die rasche Entwicklung im maschinellen Lernen bringt konstant neue Herangehensweisen hervor, die eine umfassende Taxonomie zu einer Herausforderung machen. Im Folgenden sollen die wichtigsten Gruppen genannt werden, die jedoch nicht als distinkte Bereiche betrachtet werden sollen, sondern vielmehr als Dimensionen im Raum der vielfältigen Ansätze betrachtet werden können.

1.2.3 Überwachtes und unüberwachtes Lernen

Diese Einteilung ist die vermutlich bekannteste und lässt sich über die Verfügbarkeit von Zielvariablen und dem Lernziel definieren. Beim unüberwachten Lernen (engl.: Unsupervised Learning) stehen keine Zielvariablen zur Verfügung und das allgemeine Ziel besteht im Erkennen von Strukturen in den Daten, die sogenannte Mustererkennung. Typische unüberwachte Lernaufgaben sind:

- Clustering: Ziel ist die Gruppierung von Daten, wobei der Erfolg des Clusterings anhand von Clustereigenschaften wie Kompaktheit und Separierbarkeit beurteilt wird.
- Analyse der gelernten Repräsentation: Das Repräsentationslernen wird meist in Kombination mit einer anderen Lernaufgabe verwendet, soll allerdings an dieser Stelle separat genannt werden, da eine Analyse der gelernten Repräsentation den Fokus darstellen kann. Die Güte der Repräsentation ist hier durch den Grad der Interpretierbarkeit gegeben. Eine hohe Interpretierbarkeit ist beispielsweise gegeben, wenn Merkmalsdimensionen einer Semantik zugeordnet werden können.
- Dimensionsreduktion: Diese Aufgabe kann als Sonderfall des Repräsentationslernens angesehen werden, wobei die primäre Anforderung an die Repräsentation die Reduktion der Komplexität der Inputdaten ist. Der Erfolg des Ergebnisses wird oftmals an der verlorenen Information zwischen Originaldaten und den reduzierten Daten gemessen und daran, ob die reduzierten Daten eine bessere Interpretierbarkeit gewährleisten.

Beim überwachten Lernen (engl.: Supervised Learning) stehen gelabelte Informationen zur Verfügung. Typische überwachte Lernaufgaben sind:

- Klassifikation: Als Klassifikation werden alle Aufgaben bezeichnet, bei denen die Inputdaten zu einer vorher definierten Klasse zugeordnet werden. Der Erfolg der Klassifikation wird an der Übereinstimmung von Schätzung und Referenz gemessen.
- Regression: Bei Regressionsaufgaben wird ein kontinuierlicher Output basierend auf den Inputdaten geschätzt. Wie bei der Klassifikation wird der Erfolg der Regression daran gemessen, wie sehr die Schätzung und die Referenz übereinstimmen.
- Zukunftsvorhersage: Ähnlich zur Regression wird in dem Fall eine Schätzung vorgenommen, die in der Zukunft liegt, was eine Berücksichtigung der Zeitinformation beinhaltet. Anders als bei der Regression, bei der eine Approximation das Ziel ist, ist bei der Vorhersage eine Extrapolation notwendig.
- Szenenklassifikation: Als Spezialfall der Klassifikation wird bei der Szenenklassifikation eine komplette Bildszene, in der Fernerkundung meist eine Luftbild- oder Satellitenbildszene, zu einer oder mehreren Klassen zugeordnet.
- Objektdetektion: Diese meist im Bereich Bildinterpretation angegangene Aufgabe hat zum Ziel, Objektinstanzen zu erkennen, denen eine bestimmte Semantik zugeordnet werden kann. Die Instanzen werden in den meisten Anwendungen durch achsparallele Bounding Boxen repräsentiert.

- Semantische Segmentierung und Instanzsegmentierung: Bei der semantischen Segmentierung wird jedem Datenpunkt (z. B. Pixel, Punkt in einer Punktwolke) eine vorher definierte Klasse zugeordnet, wobei benachbarte Datenpunkte der gleichen Klasse ein semantisches Segment bilden. Bei der Instanzsegmentierung wird darüber hinaus noch zwischen Segmenten unterschieden, wenn sie eigenständige Instanzen bilden, was von Bedeutung ist, wenn Instanzen mit der gleichen Semantik benachbart sind.

Des Weiteren gibt es Ansätze, die überwachtes und unüberwachtes Lernen verbinden, indem sie zum Beispiel sowohl Daten mit dazugehörenden Zielvariablen als auch ohne verwenden. Im Bereich der Geodäsie und Geoinformation sind diese Ansätze sehr vielversprechend, da durch den Einsatz diverser Sensoren meist große Mengen an Daten vorhanden sind; zusätzliche Zielvariablen sind jedoch oft zeitaufwendig und teuer zu beschaffen. Etablierte Lernparadigmen sind zum Beispiel:

- Semi-supervised Learning: Bei diesem meist in der Klassifikation verwendeten Lernparadigma werden gemeinsam ungelabelte Daten und gelabelte Daten verwendet, allerdings müssen die ungelabelten Daten zu einer der vorher definierten Klassen gehören. Es ist lediglich unbekannt, zu welchen Klassen sie gehören.
- Self-taught Learning: Ähnlich zum Semi-supervised Learning werden auch hier gelabelte und ungelabelte Daten verschiedener Klassen verwendet, allerdings gibt es bei den ungelabelten Daten keinerlei Restriktionen, außer, dass sie von der gleichen Datenart (z. B. RGB-Bilder) sein müssen.
- Self-supervised Learning: Bei dieser Art des Lernens wird alleinig auf den Inputdaten gelernt, jedoch in einem überwachten Setting. Ein Beispiel ist das Entrauschen von Bildern. Dazu wird ein Rekonstruktionsmodell gelernt, wobei Input und Output identisch sind. Die Vorhersagefunktion wird so gewählt, dass Rauschen und unwichtige Muster unterdrückt werden, indem beispielsweise eine Dimensionsreduktion der Daten enthalten ist. Der Erfolg der Rekonstruktion wird anhand des Unterschieds zwischen Input und Output gemessen. Self-supervised Learning wird häufig bei neuronalen Netzen eingesetzt, um eine gute Initialisierung der Parameter mit ungelabelten Daten zu bestimmen.

1.2.4 Generatives und diskriminatives Lernen

Eine weitere gängige Einteilung von Lernverfahren ist in generativ und diskriminativ gelernte Modelle. Im Kontext der Klassifikation wird bei generativen Modellen zunächst die Verteilung der Daten gelernt und daraus die Entscheidungsgrenze zwischen den Klassen abgeleitet, während diskriminative Modelle direkt auf die Schätzung der Entscheidungsgrenze abzielen. Mit generativen Modellen besteht die Möglichkeit, neue Daten zu generieren, die zu der gelernten Verteilung gehören. Dies kann beispielsweise explizit geschehen, indem man die Verteilung parametrisch definiert (z. B. wie eine Gauß-Verteilung) und dann Datenpunkte aus dieser Verteilung zieht, oder es kann implizit geschehen, indem ein Modell gelernt wird, welches aus der implizit definierten Verteilung abtasten kann, ohne dass diese explizit definiert werden.

1.3 Künstliche neuronale Netze

1.3.1 Aufbau und Funktionsweise

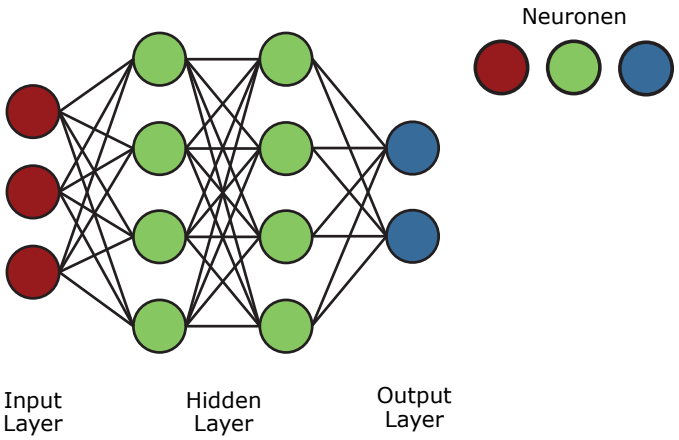

Abb. 1.2: Schematische Darstellung eines neuronalen Netzes (eigene Darstellung)

Bei einem künstlichen neuronalen Netz handelt es sich um ein leistungsstarkes Modell des maschinellen Lernens, welches eine Kette von mathematischen Operationen darstellt. Die Grundbausteine eines neuronalen Netzes sind die Neuronen, die als Knotenpunkte in einem Graph oder Netz visualisiert werden können (Abb. 1.2). Man unterscheidet im Wesentlichen zwischen Neuronen im Input Layer, die die Daten repräsentieren, Neuronen im Output Layer, die den Modelloutput darstellen, und Neuronen in den Hidden Layern (deutsch: verborgene Schichten). Jedes Neuron nach dem Input Layer kann als Funktion angesehen werden, dessen Funktionswert vom Input, der Gewichtung des Inputs und dem Bias abhängt. Der Funktionswert wird Aktivierung genannt und im Neuron gespeichert. Der Input zu jedem Neuron ist jeweils die Aktivierungen aus dem vorherigen Layer und in der Abbildung durch eine schwarze Verbindungslinie gekennzeichnet. Die Inputs werden gewichtet, wobei die Gewichte kontinuierliche Werte annehmen können und mit den Werten der jeweiligen Inputneuronen multipliziert werden. Somit hat jeder Input einen unterschiedlichen Einfluss auf jedes Neuron im folgenden Layer. Zusätzlich besitzt jedes Neuron optional ein Bias, der auf den Funktionswert addiert wird. Um die Approximationsfähigkeit des neuronalen Netzes zu erhöhen, wird die Aktivierung, d. h. der Funktionswert eines Neurons, in der Regel durch eine nichtlineare Aktivierungsfunktion angepasst, wodurch sich auch der Wertebereich der Aktivierungen beeinflussen lässt. Klassische Aktivierungsfunktionen sind eine Sigmoid-Funktion mit einem Ausgabewertebereich von [0 1], eine tanh-Funktion mit [-1 1] und eine ReLu-(Rectified Linear Unit-)Funktion mit [0 1]. Der Modelloutput ist somit das Ergebnis aus einer Verkettung von Funktionen, deren Parameter eine geeignete Repräsentation des Inputs lernen, um die gestellte Aufgabe zu lösen.

Die Komplexität eines künstlichen neuronalen Netzes bemisst sich im Wesentlichen anhand der Anzahl und Art der Neuronen und der Tiefe, d. h. der Anzahl an Layer, in denen diese verschachtelt bzw. verkettet sind. Das erste neuronale Netz, entwickelt von Frank Rosenblatt