



Dr. Julian Lorenz
Senior Quantitative Analyst

Auch quantitative Anlagestrategien können von Künstlicher Intelligenz und Big Data profitieren

Juni 2018

Maschinen mit menschenähnlicher Intelligenz (oder gar überlegener Superintelligenz) zu konstruieren, ist seit jeher Traum und Albtraum in Science-Fiction wie in der Wissenschaft. In den vergangenen Jahren haben die Themen »Künstliche Intelligenz« (KI) und »Maschinelles Lernen« (englisch: machine learning) einen regelrechten Hype erfahren. Alle grossen Technologie-Unternehmen arbeiten mit grossem Eifer daran. Googles KI-Projekt »DeepMind« verkündete im Dezember 2017 mit dem Sieg seines KI-Programms »AlphaZero« über die bisherigen weltweit stärksten Schachprogramme einen weiteren spektakulären Meilenstein. Bereits einige Monate zuvor hatte »AlphaZero« beim asiatischen Brettspiel »Go« gegen den amtierenden Weltmeister gewonnen. Bemerkenswert bei »AlphaZero« ist, dass es vollständig autodidaktisch lernt: Seine menschlichen Lehrer brachten ihm nur die Spielregeln bei. Es lernte und perfektionierte daraufhin das Spiel, indem es einen Tag lang gegen sich selbst spielte. Der Zeitpunkt, dass Computer den Menschen in sämtlichen Disziplinen in den Schatten stellen, steht scheinbar kurz bevor.

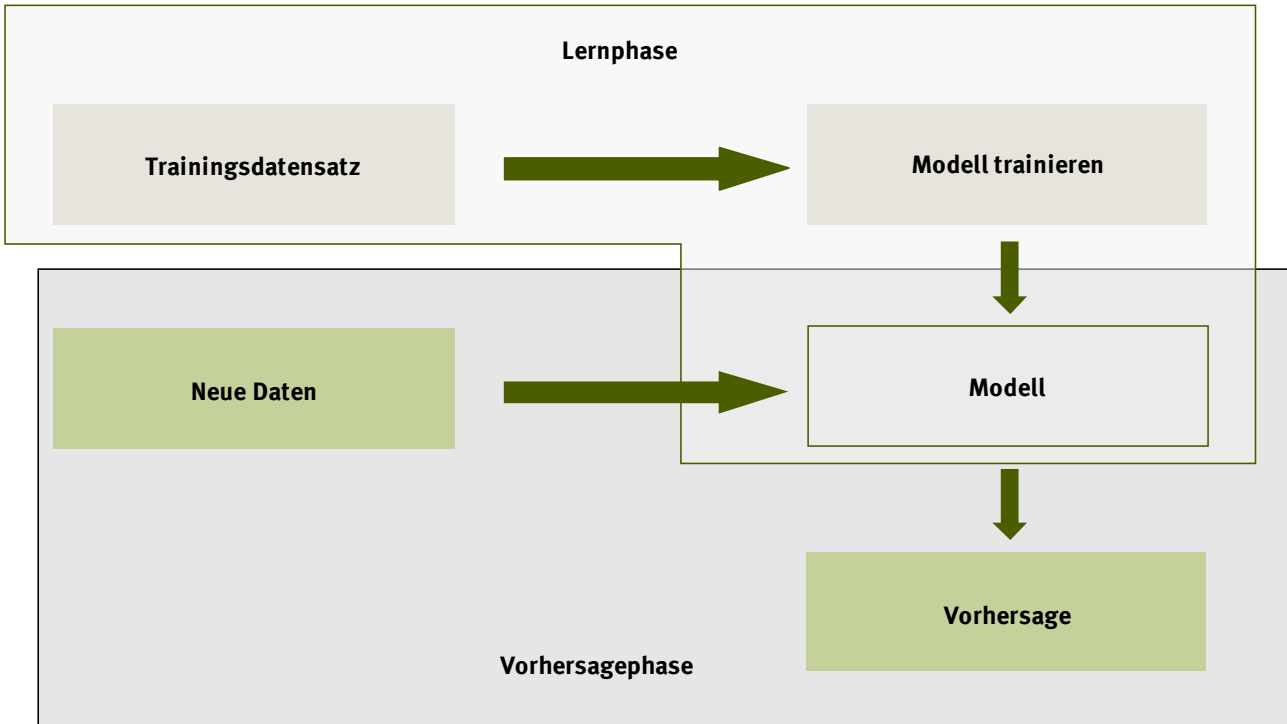
Maschinen lernen Lernen

Der Fortschritt der vergangenen Jahre ist beachtlich. Aber KI ist eigentlich kein neues Thema. Ihre Anfänge reichen mindestens bis in die fünfziger Jahre des vergangenen Jahrhunderts zurück. Es folgten Wellen der Euphorie mit futuristischen Prognosen – und der Ernüchterung (»Winter der künstlichen Intelligenz«). Im neuen Jahrtausend gab es mit Projekten von IBM (z.B. IBM Deep Blue im Schach oder IBMs Watson) und schliesslich Google (»Google's cats«) ein erneutes Aufblühen und medienwirksame Erfolge.

Ein KI-System lernt aus Beobachtungen und Daten der Vergangenheit und kann diese nach Beendigung der Lernphase verallgemeinern, das heisst, es erkennt Muster und Gesetzmässigkeiten (siehe Abbildung 1). Das Ziel des KI-Systems ist, Daten intelligent miteinander zu verknüpfen, Zusammenhänge zu erkennen und Vorhersagen zu treffen, wenn neue Daten vorgelegt werden.

Bei KI-Systemen unterscheidet man zwischen »überwachtem Lernen« (engl: supervised learning) und »unüberwachtem Lernen« (engl: unsupervised learning). Beim »überwachten Lernen« erhält das KI-System ständig Rückmeldungen. Bei jedem Datenpunkt kann die Lernsoftware ihre Vorhersage mit einem vom Programmierer vorgegebenen korrekten Wert abgleichen, um das Modell anzupassen und zu optimieren. Falls die korrekten Werte nicht vorliegen, ist ein solches Training nicht möglich und es liegt »unüberwachtes Lernen« vor. Die Lernmaschine versucht dann, in den Eingabedaten Muster zu erkennen, die vom strukturlosen Rauschen abweichen.

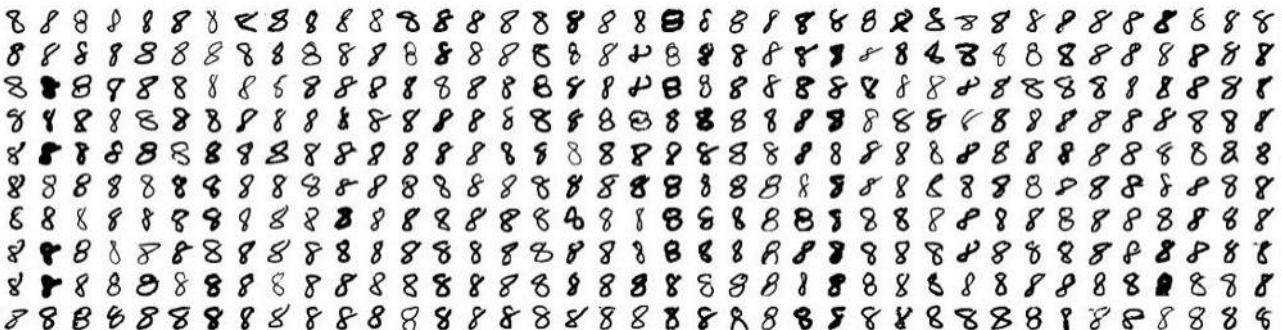
Abbildung 1: Schematik des maschinellen Lernens



Quelle: BANTLEON

Muster und Gesetzmässigkeiten zu erkennen, bedeutet typischerweise ein System von unscharfen Grenzen anstelle einfacher Regeln. Ein klassisches Beispiel und eines der frühen Anwendungsgebiete neuronaler Netzwerke ist die Handschrifterkennung. Was für uns Menschen selbstverständlich erscheint, ist für die maschinelle Umsetzung nicht leicht – beispielsweise das Erkennen einer handgeschriebenen Acht, vorgelegt in einem Pixel-Bild. Wäre jede handgeschriebene Acht exakt gleich, könnte eine Maschine sie durch eine einfache Regel identifizieren, nämlich durch bloßen Vergleich. Doch keine handgeschriebene Acht gleicht einer anderen (Abbildung 2), und auch die Position sowie die Orientierung im vorgelegten Pixel-Bild werden variieren. Ein neuronales Netz zur Schrifterkennung wird nun mit einer Vielzahl von verschiedenen handgeschriebenen Achten trainiert. Es lernt daraus, sogar Achten zu erkennen, die den gesehenen Achten nur *ähnlich* sind, aber mit keiner identisch. Die Treffsicherheit des Erkennens in der Vorhersagephase steigt mit der Menge der Achten in der Trainingsphase.

Abbildung 2: Beispiele von Trainingsdaten für ein neuronales Netzwerk zur Handschrifterkennung



Quelle: MNIST Datenbank, <http://yann.lecun.com/exdb/mnist>

Wie hier in diesem Beispiel der Handschrifterkennung sind aktuelle KI-Systeme stets spezialisierte Systeme, das heisst für eine bestimmte Anwendung und Aufgabenstellung konzipiert und implementiert. In ihren jeweiligen Aufgabenstellungen können solche KI Systeme bereits herausragende Ergebnisse erzielen. Von »Artificial General Intelligence«-KI-Systemen, die die volle Breite menschlicher kognitiver Fähigkeiten in beliebigen (auch untrainierten) Situationen erreichen oder sogar übertreffen, ist die KI-Forschung allerdings noch sehr weit entfernt.

Die »Big Data«-Revolution

Eine zentrale Rolle spielen für KI-Systeme die Daten, die als Input verwendet werden. Dass in vielen Bereichen täglich wachsende Datensätze aus einer zunehmenden Zahl an Datenquellen zur Verfügung stehen, ist deshalb ein treibender Faktor im Fortschritt der KI. Man spricht hier oft von »Big Data«.

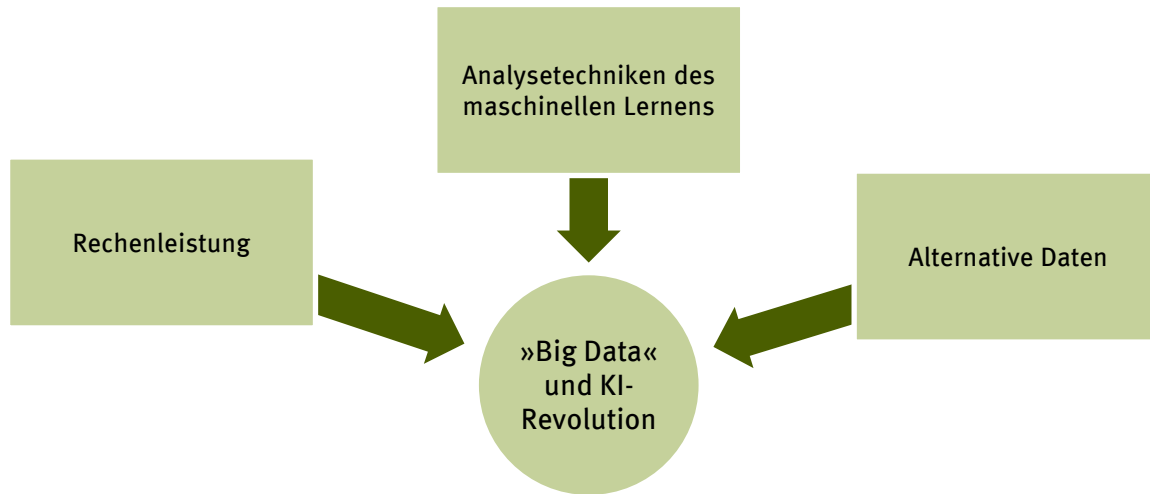
**Abbildung 3: Klassifikation von »Big Data«
nach »United Nations Global Working Group on Big Data for Official Statistics«**

Soziale Medien »Daten generiert von Menschen«	Geschäftsprozesse »Daten generiert durch Geschäftsaktivitäten«	Internet-der-Dinge »Daten generiert von Maschinen und Sensoren«
<ul style="list-style-type: none"> • Nachrichten • Soziale Netzwerke: Facebook, Twitter ... • Blogs • Bilder: Instagram, Flickr ... • Videos: Youtube ... • Internet-Suchdaten • Mobile Textmessages: SMS • E-Mail • ... 	<ul style="list-style-type: none"> • Kommerzielle Transaktionen • Finanzdaten • Unternehmensdaten • E-Commerce • Kreditkarten • Regierungsbehörden • ... 	<ul style="list-style-type: none"> • Daten von Sensoren • Wetterdaten/Umwelt • Verkehr • Satelliten • Mobile Sensoren (Lokalisationsdaten) • Daten von Computersystemen (Logs) • ...

Quelle: <http://unstats.un.org/bigdata>

Derart grosse Datensätze und ihre Verarbeitung sind auch eine Herausforderung für die Hardware. Nicht zuletzt wegen damals noch mangelnder Computerleistung blieb den Anfängen der KI-Forschung im 20. Jahrhundert zunächst der Erfolg versperrt. Die beständig wachsende Rechenleistung ist somit ein weiterer wesentlicher Pfeiler der KI-Revolution.

Abbildung 4: Eckpfeiler der »Big Data« und KI-Revolution



Quellen: BANTLEON

Der dritte Faktor in dieser Entwicklung ist die rege Forschung bei Algorithmen. Hier kommen als Erstes künstliche neuronale Netzwerke in den Sinn. In der Tat basieren viele der jüngsten Erfolge auf Weiterentwicklungen von künstlichen neuronalen Netzwerken, oft in Kombination mit anderen Techniken (beispielsweise »Bestärkendes Lernen«; englisch: reinforcement learning). Neuronale Netzwerke sind allerdings nur eine von vielen verschiedenen Techniken des maschinellen Lernens im Schnittbereich aus Statistik und Informatik. Abbildung 5 zeigt eine Auflistung der verfügbaren Algorithmen.

Abbildung 5: Lernalgorithmen

Maschinelles Lernen / Künstliche Intelligenz					
Überwachtes Lernen		Unüberwachtes Lernen		Deep Learning	Andere Verfahren
Regression	Klassifikation	Clustering	Faktoranalyse	Zeitreihen, unstrukturierte Daten	Bestärkendes Lernen, teilüberwachtes Lernen, aktives Lernen
Lasso, Ridge, Elastic Net ...	SVM, Decision Trees, Random Forest, HMM ...	K-means, Spektrales Clustering ...	PCA, ICA ...	Deep Boltzmann, Convolutional NN, Multilayer Perceptron ...	

Quellen: BANTLEON

KI zur Finanzmarktprognose – der »perfekte Portfolio Manager«?

Wie sieht es mit KI-Systemen für die Prognose von Finanzmarktentwicklungen aus? An »Big Data« mangelt es an den Finanzmärkten kaum. Neben den klassischen Daten (Marktdaten, fundamentale Unternehmenskennzahlen, makroökonomische Daten etc.) stehen immer mehr alternative Datensätze zur Verfügung. Die Datenauflösung und Datenrate (Updatefrequenz) nimmt zudem weiter zu (z.B. Marktdaten in »Tick-Auflösung«).

Ähnlich wie bei unserem Beispiel der Handschrifterkennung sind in Finanzmärkten die Preisbewegungen und Marktdynamiken nie identisch – aber scheinbar wiederholen sich bestimmte Muster und Ähnlichkeiten. KI-Systeme sollen genau diese komplexen, unscharfen Muster lernen, anstelle starrer Regeln wie etwa in der Charttechnik.

Je nach verwendeter Datenquelle aus der Big-Data-Datenflut sind die möglichen Anwendungsgebiete höchst vielfältig. Einige Beispiele:

- Hochfrequenz-Marktdaten für kurzfristige Trendprognosen
- Automatische News-Klassifizierung in Echtzeit zur Bestimmung des Markt-Sentiments
- Unternehmenskennzahlen aller Art und Detailtiefe zur Klassifizierung eines Aktien-Universums in Over- und Underperformer
- Internet-Suchdaten (Suchmaschinen) und Daten aus sozialen Netzwerken (Twitter, Facebook etc.) zur Prognose von Trends im Konsumentenverhalten
- Satellitendaten, beispielsweise zur Beobachtung von Verkehrsaufkommen oder zur Beobachtung des globalen Handels (z.B. Schifffahrtsrouten, Häfen etc.)
- Kundenaufkommen in Supermärkten anhand der Parkplatzbelegung
- Wetterdaten zur Prognose von Rohstoffpreisen oder Konsum
- Daten von Immobilienportalen zur Beobachtung des Zustands und der Entwicklung des Immobilienmarktes
- Produktpreise in Onlineshops zur Inflationsprognose (z.B. »MIT Billion Prices Project«)
- Daten aus Karriere- und Stellenportalen zur Vorhersage der Beschäftigungsentwicklung

Zur konkreten Veranschaulichung betrachten wir ein einfaches Beispiel. Unser Ziel ist die Prognose des DAX auf monatlicher Basis. Als Datengrundlage zur Vorhersage verwenden wir monatliche Preisdaten verschiedener Finanzmärkte (DAX, Eurostoxx50, S&P500, MSCI World, EUR/USD) sowie fünf »BANTLEON Frühindikatoren« (»BANTLEON Economic Indicator A – E«), die ebenfalls monatlich errechnet werden. Unsere Aufgabenstellung lautet also, ein quantitatives System zu entwickeln, mit dem wir jeweils am Monatsanfang unter Verwendung der bis zu diesem Zeitpunkt vorliegenden Daten die DAX-Rendite bis zum Monatsende prognostizieren.

Um unser Modell zu trainieren, verfügen wir über historische Monatsdaten der vergangenen 25 Jahre (1992–2017) aller Zeitreihen. Für jede Zeitreihe generieren wir deren 3-Monats-, 6-Monats- und 9-Monats-Veränderung (bezeichnet mit 3M, 6M und 9M) und verwenden diese als unsere »Features« (in der Statistik: »erklärende Variablen«). Zusätzlich zu den Features sind die realisierten Monatsrenditen des DAX bekannt. In der Sprache des maschinellen Lernens sind dies »Labels«, und es liegt ein Problem aus dem »überwachten Lernen« vor. Abbildung 6 veranschaulicht die Situation.

Abbildung 6: DAX-Monatsprognose als »überwachtes Lernen«

		FEATURES				LABELS	
		Feature A	Feature B	Feature C	Feature D	DAX	1-Monatsrendite
TRAININGS- PHASE	02.01.2017	0,13	19,51	-0,88	3,03	11'481	+ 54
	01.02.2017	0,18	19,69	0,00	2,98	11'535	+ 299
	01.03.2017	0,12	19,67	0,00	2,94	11'834	+ 478
	03.04.2017	0,07	19,75	0,30	2,91	11'312	+ 126
	01.05.2017	-0,02	19,83	0,54	2,92	12'438	+ 177
	01.06.2017	0,09	19,81	0,19	2,92	12'615	- 290
	03.07.2017	0,07	19,82	-0,15	2,94	12'325	- 207
	01.08.2017	0,20	19,80	0,19	2,97	12'118	- 63
	01.09.2017	0,17	19,45	-1,99	2,87	12'055	+ 773
	02.10.2017	0,15	19,38	0,75	2,74	12'828	+ 401
VOR- HERSAGE- PHASE	01.11.2017	0,17	19,42	-0,33	2,66	13,229	?

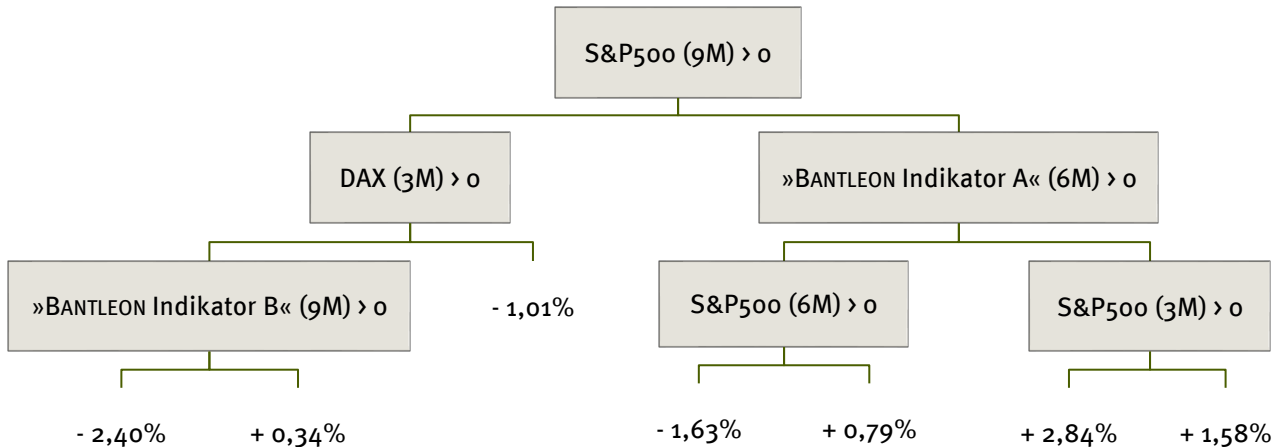
Quelle: BANTLEON

Für ein solches Problem kommen viele Modelle (Lernalgorithmen) in Frage. Wir betrachten im Folgenden sogenannte Entscheidungsbäume (englisch: decision trees, auch CART-classification and regression trees). Dabei wird aus den vorgelegten Daten eine hierarchische Struktur von möglichst wenigen Entscheidungswegen gelernt, und zwar in Form eines binären Baumes. An jeder Verzweigung des Baumes wird ein Feature mit einem Schwellenwert verglichen. Die Blätter des Baumes enthalten den jeweiligen Prognosewert. Der Vorteil von Entscheidungsbäumen als Modell besteht in ihrer Einfachheit, der intuitiven Interpretierbarkeit und grafischen Visualisierbarkeit. Sie sind sowohl für Klassifikations- als auch für Regressionsprobleme verwendbar.

Zu einem vorgelegten Trainingsdatensatz kann mit effizienten Verfahren der zugehörige optimale Entscheidungsbaum ermittelt (gelernt) werden. »Optimal« bedeutet, dass die Struktur des Entscheidungsbaumes (welche Features und Schwellenwerte als Verzweigungsregeln und welche Prognosewerte an den Baumblättern) so gewählt ist, dass er in Bezug auf die vorgelegten Daten am besten passt – etwa wie bei einer linearen Regression die Koeffizienten der erklärenden Variablen so bestimmt werden, dass der quadratische Fehler der zu erklärenden Variable minimiert wird.

Abbildung 7 zeigt den optimal gelernten Entscheidungsbaum für unseren Datensatz mit einer maximalen Baumtiefe von drei. An den Verzweigungspunkten sind die Verzweigungsbedingungen der Features aufgetragen. Falls eine Verweisungsbedingung erfüllt ist, folgt man dem rechten Ast – anderenfalls dem linken. Jeweils am Ende eines jeden Verzweigungspfades ist die prognostizierte DAX-Monatsrendite angegeben. Beispielsweise prognostiziert das Modell eine positive Monatsrendite von 1,58%, falls die aktuelle 9-Monats-Veränderung des S&P500 positiv ausfällt, der »BANTLEON Frühindikator A« (6-Monats-Veränderung) im positiven Bereich liegt und auch die aktuelle 3-Monats-Veränderung des S&P500 positiv ist (Ast ganz rechts am Baum). Die prognostizierte Monatsrendite von 1,58% ist dabei der Mittelwert aller Monatsrenditen in den Trainingsdaten, die in diesen Ast fallen. Ein Entscheidungsbaum wie in Abbildung 7 kann nun in der Vorhersagephase zur Prognose verwendet werden.

Abbildung 7: Entscheidungsbaum für DAX-Monatsprognose



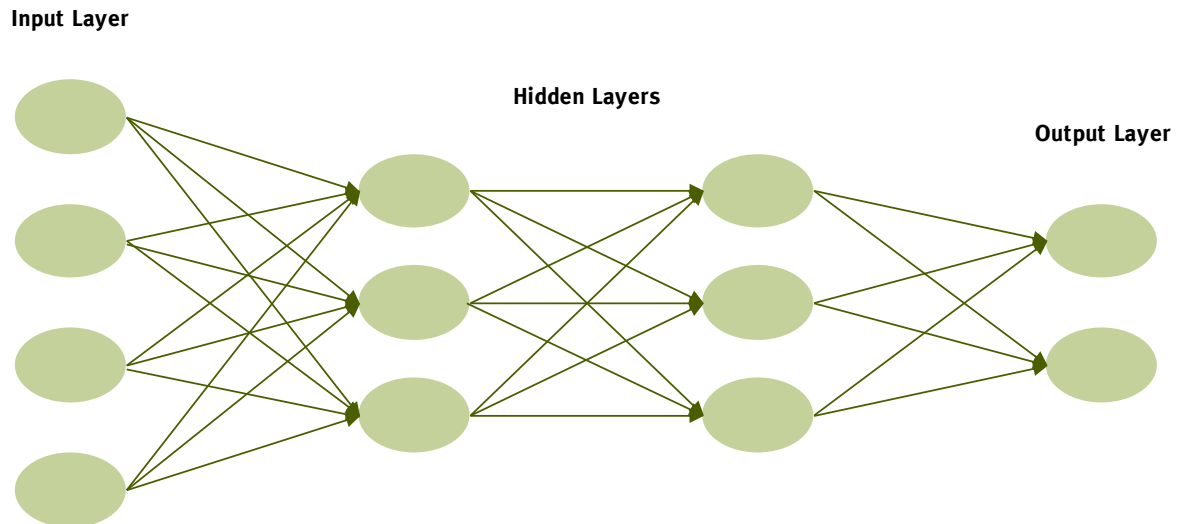
Quelle: BANTLEON

Ein zentrales Problem im »überwachten Lernen« ist die Wahl der Modellkomplexität. Prinzipiell möchte man ein Modell haben, das möglichst genau zu den Trainingsdaten passt. Bei Entscheidungsbäumen wird die Modellkomplexität im Wesentlichen durch die zugelassene maximale Baumtiefe gesteuert. Je tiefer sich der Baum verzweigt, desto besser kann er die Trainingsdaten erfassen. Allerdings nimmt mit höherer Modellkomplexität die Gefahr der Überanpassung (Overfitting) zu. Das Modell passt dann zwar sehr gut zu den Trainingsdaten, liefert aber in der Vorhersagephase auf bisher ungesehenen Daten schlechtere Ergebnisse. Dies ist das Verzerrung-Varianz-Dilemma (englisch: bias-variance tradeoff). Dabei versteht man unter Verzerrung den Fehler aufgrund falscher (zu einfacher) Annahmen im Lernalgorithmus (Unteranpassung) und unter Varianz den Fehler aufgrund Empfindlichkeit gegenüber Rauschen in den Trainingsdaten durch Überanpassung. Mithilfe verschiedener Techniken, beispielsweise Kreuzvalidierung (englisch: cross-validation), lässt sich diese Fragestellung in den Griff bekommen und eine »optimale« Modellkomplexität auswählen.

Es gibt zahlreiche Weiterentwicklungen der hier vorgestellten einfachen Entscheidungsbäume, die verschiedene Verbesserungen liefern, etwa die Klassifikationsgüte erhöhen oder die Gefahr der Überanpassung verringern. Eine Methode ist der Einsatz von Mengen von Entscheidungsbäumen (Entscheidungswälder; englisch: decision forests) anstelle einzelner Bäume. Man spricht hier von Ensemble-Techniken: Ein einzelner Entscheidungsbaum kann zwar keine optimale Klassifikation liefern, aber die Mehrheitsentscheidung einer Menge geeigneter Bäume kann eine deutlich bessere Klassifikation erzielen. Verbreitete Methoden sind z.B. »Random Forest« und »Boosting«.

Als ganz andere Klasse von Verfahren käme für unser Beispiel der DAX-Monatsprognose die Verwendung von künstlichen neuronalen Netzwerken in Frage. Hier gibt es sehr viele Varianten. Die Topologie (Struktur) des Netzwerkes definiert, wie die künstlichen Neuronen miteinander verbunden sind (siehe das Beispiel in Abbildung 8). Typischerweise sind diese in Schichten angeordnet. Die vorderste Schicht ist die Eingabe-Schicht (englisch: input layer), in die die Features der Daten eingespeist werden. Die hinterste Schicht ist die Ausgabe-Schicht (englisch: output layer). Dazwischenliegende Schichten werden als verdeckte Schichten (englisch: hidden layers) bezeichnet. Jedes Neuron sammelt den Input entlang aller seiner eingehenden Verbindungen. Jede Verbindung hat ein Gewicht. Falls die aggregierte Inputstärke über dessen spezifischem Schwellenwert liegt (bzw. in manchen Varianten komplexere Aktivierungsfunktionen erfüllt), »feuert« das Neuron auf seinen ausgehenden Verbindungen in die nächste Schicht.

Abbildung 8: Beispieltopologie eines künstlichen neuronalen Netzwerkes



Quelle: BANTLEON

Nach der Konstruktion einer solchen Topologie folgt die Trainingsphase, in der das Netz lernt und zwar indem es sich gemäss einer fest vorgegebenen Vorschrift selbst modifiziert. Die Modifikation des Netzwerkes erfolgt typischerweise hauptsächlich durch Modifikation der Gewichte entlang der Verbindungen zwischen den Neuronen. Das Ziel der iterativen Modifikationen ist es, jeweils für eine vorgegebene Eingabe (Features) eine gewünschte Ausgabe zu produzieren.

Durch die Wahl der Topologie sind künstliche neuronale Netzwerke ein sehr vielseitiges und mächtiges Werkzeug. Mit der Hinzunahme weiterer Schichten steigt die Komplexität massiv. Dies sind die in den vergangenen Jahren sehr populär gewordenen »Deep Learning«-Methoden – künstliche neuronale Netzwerke mit zahlreichen Zwischenschichten. »AlphaZero« etwa verwendet mehr als zehn verdeckte Schichten. Die grosse Herausforderung bei solchen komplexen Netzwerken besteht darin, dass der Lernprozess sehr viel aufwendiger wird, was sie erst in den vergangenen Jahren wegen der gestiegenen Verarbeitungsgeschwindigkeit der Hardware praktikabel machte. Weiter steigende Rechenleistung wird die Leistungsfähigkeit von künstlichen neuronalen Netzwerken kontinuierlich verbessern. Der nächste revolutionäre Entwicklungsschritt bahnt sich zudem in Form von neuromorphen Chips an. Dies sind Mikrochips, die nach dem Beispiel von natürlichen Nervennetzen versuchen, auf Hardwareebene die Funktionalität eines künstlichen neuronalen Netzwerkes zu implementieren. Gegenüber der aktuellen Technik der Simulation von künstlichen neuronalen Netzwerken auf Softwareebene erhofft man sich eine massive Erhöhung der Effizienz und Geschwindigkeit. Damit würden noch grössere und mächtigere künstliche neuronale Netzwerke realisierbar werden.

Fazit

Techniken des maschinellen Lernens und die Erschliessung alternativer Datenquellen etablieren sich bei quantitativen Investments. Der Übergang von bekannten Techniken aus der Statistik zu maschinellem Lernen und künstlicher Intelligenz ist fließend. Finanzmärkte sind allerdings aufgrund ihrer Komplexität – sehr viele Einflussfaktoren und Akteure – schwer vorherzusagen. Von einem selbstlernenden »AlphaZero für die Finanzmärkte« ist man deshalb noch weit entfernt. Aber gleichermassen wie KI die Gesellschaft nachhaltig verändern wird, dürfte sie auch die Finanzbranche verändern.