



Machine learning

voor
dummies[®]

John Paul Mueller
Luca Massaron



BBNC
uitgevers

Amersfoort, 2019

Inhoud in vogelvlucht

Inleiding	1
Deel 1: Inleiding in machine learning	7
HOOFDSTUK 1: De waarheid over AI	9
HOOFDSTUK 2: Leren in het tijdperk van big data	23
HOOFDSTUK 3: Een blik op de toekomst	37
Deel 2: Gereedschap voor machine learning	49
HOOFDSTUK 4: Een R-distributie installeren	51
HOOFDSTUK 5: Met RStudio in R coderen	67
HOOFDSTUK 6: Een Python-distributie installeren	93
HOOFDSTUK 7: Coderen in Python met Anaconda	111
HOOFDSTUK 8: Andere hulpmiddelen voor machine learning	139
Deel 3: Aan de slag met wiskundige beginselen	147
HOOFDSTUK 9: De wiskunde achter machine learning doorgronden	149
HOOFDSTUK 10: De juiste curve zoeken	169
HOOFDSTUK 11: Machine learning valideren	183
HOOFDSTUK 12: Eenvoudige algoritmen	203
Deel 4: Slimme gegevens en big data inzetten ..	223
HOOFDSTUK 13: Gegevens voorbereiden	225
HOOFDSTUK 14: Gebruikmaken van gelijkenissen	245
HOOFDSTUK 15: Op eenvoudige wijze met lineaire modellen werken	267
HOOFDSTUK 16: De complexiteit van neurale netwerken	291
HOOFDSTUK 17: Een stap verder met Support Vector Machines	309
HOOFDSTUK 18: Ensembles van zelflerende algoritmen	327
Deel 5: Machine learning op echte problemen toepassen	345
HOOFDSTUK 19: Afbeeldingen classificeren	347
HOOFDSTUK 20: Mening en gevoelens classificeren	363
HOOFDSTUK 21: Producten en films aanbevelen	385
Deel 6: Het deel van de tientallen	401
HOOFDSTUK 22: Tien pakketten voor machine learning	403
HOOFDSTUK 23: Tien manieren om modellen voor machine learning te verbeteren	409
Index	419

1

Introduction in machine learning

IN DIT DEEL . . .

Ontdek je hoe AI echt werkt en wat het voor je kan betekenen.

Leer je de betekenis kennen van de term *big data*.

Lees je over de rol van statistiek bij machine learning.

Ontdek je wat machine learning in de toekomst gaat brengen.

IN DIT HOOFDSTUK

De AI-hype met een korrel zout nemen

De droom van AI definiëren

AI en machine learning met elkaar vergelijken

Beseffen waar de techniek stopt en creativiteit begint

Hoofdstuk 1

De waarheid over AI

Kunstmatige intelligentie, oftewel Artificial Intelligence (AI), speelt in het dagelijks leven een steeds belangrijkere rol door het succes van technologieën als Siri (<https://www.apple.com/ios/siri>). Zoek je in de stad naar het beste sushirestaurant of de kortste route naar het concertgebouw, dan is het zowel leuk als nuttig om tegen je smartphone te praten. Terwijl jij aan het praten bent, leert je smartphone meer over jouw unieke manier van spreken en begrijpt hij je verzoeken steeds beter. Het vermogen van een smartphone om jouw manier van spreken te leren kennen en te interpreteren is een voorbeeld van AI. Machine learning vormt een deel van de technologie die wordt gebruikt om dit mogelijk te maken. Je maakt tegenwoordig waarschijnlijk al beperkt gebruik van machine learning en AI zonder erover na te denken of het zelfs in de gaten te hebben. De mogelijkheid om met apparaten te communiceren en ervoor te zorgen dat ze daadwerkelijk doen wat je van plan bent, is een praktijkvoorbeeld van machine learning. Aanbevelingssystemen, bijvoorbeeld die van Amazon, tonen producten aan de hand van criteria zoals eerdere aankopen of producten die bij je huidige keuze aansluiten. Het gebruik van zowel AI als machine learning neemt in de toekomst alleen maar toe.

In dit hoofdstuk verdiep je je in AI en ontdek je wat het betekent vanuit verschillende perspectieven, inclusief hoe het je beïnvloedt als consument, wetenschapper of technicus. Je leest ook dat AI niet hetzelfde is als

machine learning, al worden die twee in de media vaak door elkaar gehaald. Machine learning is echt wat anders dan AI, ook al zijn beide verwant.

Verder kijken dan de hype

Als een technologie belangrijker wordt, neemt ook de hype toe en AI kan zeker niet klagen over een tekort aan media-aandacht. Sommige mensen kijken niet alleen naar de wetenschappelijk aspecten, maar spelen ook op angstgevoelens in. Killerrobots, zoals in de film *The Terminator*, worden binnenkort echt niet populair. Je eerste echte ontmoeting met een AI-robot komt waarschijnlijker in de gezondheidszorg (<https://magazine.good.is/articles/robots-elder-care-pepper-exoskeletons-japan>) of als collega (<https://www.computerworld.com/article/2990849/robotics/meet-the-virtual-woman-who-may-take-your-job.html>). In werkelijkheid heb je al op een zeer alledaagse manier met kunstmatige intelligentie en machine learning te maken. Een van de redenen om dit hoofdstuk te lezen, is de hype als zodanig te herkennen en te ontdekken wat AI voor jou kan betekenen.



BELANGRIJK

Misschien is het je weleens opgevallen dat de termen machine learning en AI, oftewel kunstmatige intelligentie, door elkaar worden gebruikt. Machine learning is een onderdeel van kunstmatige intelligentie, maar geen synoniem ervan. In dit hoofdstuk lees je wat de relatie tussen machine learning en AI is, zodat je beter begrijpt hoe je met dit boek deel gaat uitmaken van een technologie die vroeger alleen in sciencefiction voorkwam.

Zowel machine learning als AI hebben belangrijke technische componenten. Dit betekent dat je beide technologieën op basis van *theorie* (onderbouwde en geteste uitleg) kunt kwantificeren in plaats van alleen maar aan de hand van een *hypothese* (een voorgestelde verklaring voor een fenomeen). Bovendien hebben beide sterke wetenschappelijke componenten, die mensen in staat stellen concepten te testen en nieuwe ideeën te creëren over de manier waarop het denkproces tot uitdrukking kan worden gebracht. Ten slotte heeft machine learning ook een artistieke component, die een getalenteerde wetenschapper de gelegenheid biedt om uit te blinken. In sommige gevallen lijken AI en machine learning alle logica te tartten, waardoor alleen een echte artiest ze zoals verwacht laat functioneren.



BELANGRIJK

INDERDAAD, VOLLEDIG AUTONOME WAPENS BESTAAN

We haasten ons om te melden dat er door duistere personen inderdaad volledig autonome wapens worden ontwikkeld, voordat lezers op het idee komen ons hier dissertaties over te sturen. Er staan uitweidingen over de ethische kant van AI in dit boek, maar we richten ons voornamelijk op de positieve, nuttige aspecten van AI die mensen helpen in plaats van ze te doden. Het meeste onderzoek naar kunstmatige intelligentie richt zich op positieve aspecten. Je vindt online meerdere bronnen over de voor- en nadelen van AI, zoals het volgende artikel in de Guardian: <https://www.theguardian.com/technology/2015/jul/27/musk-wozniak-hawking-ban-ai-autonomous-weapons>. Besef wel dat deze mensen ook niet zeker weten wat de toekomst van AI is.

Ook als je jezelf bang wilt maken, is er meer dan genoeg te lezen, bijvoorbeeld op de site <http://www.reachingcriticalwill.org/resources/fact-sheets/critical-issues/7972-fully-autonomous-weapons>, waar diepgaand op autonome wapens wordt ingegaan. Een website als Campaign to Stop Killer Robots (<https://www.stopkillerrobots.org/>) biedt extra informatie. We raden je aan de petitie tegen autonome wapens te tekenen op de site <http://futureoflife.org/open-letter-autonomous-weapons/>, want we hebben zulke wapens echt niet nodig.

Het is wel belangrijk om te beseffen dat er een verbod bestaat op ruimtewapens, chemische wapens en sommige laserwapens. De meeste landen beseffen dat ze niets oplossen. De kans is groot dat overheden volledig autonome wapens gaan verbieden omdat de bevolking moordende robots niet accepteert. Het komt erop neer dat in dit boek vooral aandacht wordt besteed aan de positieve aspecten van machine learning.

Dromen van elektrische schapen

Androïden (speciale robots die er menselijk uitzien en zich menselijk gedragen, zoals Data in *Star Trek*) en sommige soorten *humanoïde robots* (robots met menselijke eigenschappen die makkelijk van een mens te onderscheiden zijn, zoals C-3PO in *Star Wars*) zijn uithangborden voor AI geworden. Ze presenteren computers in een antropomorfe vorm die mensen kunnen begrijpen. Het is zelfs heel goed mogelijk dat je het verschil tussen menselijk en kunstmatig leven op een dag niet makkelijk meer ziet. Sciencefictionschrijvers als Philip K. Dick hebben dit lang geleden voorspeld en het lijkt nu mogelijk te worden. Het verhaal ‘*Dromen androïden van elektrische schapen?*’ gaat over het concept echter dan echt, een

idee dat gebruikt wordt in de film *Blade Runner* (<https://www.warnerbros.com/blade-runner>). In de volgende paragraaf lees je hoe dicht moderne technologie in de buurt komt van de idealen van sciencefictionschrijvers en speelfilms.



Momenteel zijn robots levensecht, al heb je snel door dat je met een androïde te maken hebt. Video's op internet laten zien dat het nog wel even zal duren voordat androïden niet meer van echte mensen te onderscheiden zijn. Kijk maar eens naar de Japanse robots op <https://www.youtube.com/watch?v=MaTfzYDZG8c> en <https://www.nbcnews.com/tech/innovation/humanoid-robot-starts-work-japanese-department-store-n345526>. Een van de overtuigendste voorbeelden is Amelia (<https://vimeo.com/154631312>). Haar verhaal wordt door ComputerWorld verteld op <http://www.computerworld.com/article/2990849/robotics/meet-the-virtual-woman-who-may-take-your-job.html>. Het komt erop neer dat het waarschijnlijk technisch mogelijk is levensechte robots en androïden te maken, maar dat die nu nog niet bestaan.

De geschiedenis van AI en machine learning

Er zijn niet alleen antropomorfe redenen dat mensen de ultieme AI als een soort androïde zien. Als sinds de oude Grieken hebben mensen het over de mogelijkheid om een brein in een mechanisch lichaam te plaatsen. Een bekende mythe is de mechanische man Talos (<http://www.ancient-wisdom.com/greekautomata.htm>). Het feit dat de oude Grieken gecompliceerde mechanische apparaten hadden, waarvan er nog maar één bestaat (het Antikythera-mechanisme: <http://www.ancient-wisdom.com/antikythera.htm>), wijst erop dat hun dromen waarschijnlijk op meer dan alleen fantasie berustten. In de loop van de eeuwen hebben verschillende mensen het gehad over mechanische personen die tot denken in staat waren (zoals de Golem van Rabbi Judah Loew, <https://www.nytimes.com/2009/05/11/world/europe/11golem.html>).

AI is ontstaan uit de hypothese dat het mogelijk is om het denken te mechaniseren. Tijdens het eerste millennium werkten Griekse, Indiase en Chinese filosofen allemaal aan manieren om deze taak uit te voeren. Al in de zeventiende eeuw bespraken Gottfried Leibniz, Thomas Hobbes en René Descartes het potentieel om alle gedachten te rationaliseren als eenvoudige wiskundige symbolen. Ze kregen natuurlijk geen vat op de complexiteit van het probleem (en dit ontglipt ons tegenwoordig nog steeds, ondanks de vooruitgang waarover je in deel 3 van dit boek leest). Het idee van AI bestaat al heel lang, maar de implementatie ervan is relatief nieuw.

De moderne vorm van AI begon met Alan Turings publicatie van ‘Computing Machinery and Intelligence’ uit 1950. In dit artikel onderzoekt Turing hoe je kunt bepalen of machines in staat zijn tot denken. Dit artikel leidde tot het imitatiespel met drie spelers. Speler A is een computer en speler B is een mens. Beide proberen Speler C (een mens die Speler A noch Speler B kan zien) ervan te overtuigen dat ze een mens zijn. Kan Speler C niet consistent bepalen wie menselijk is en wie niet, dan wint de computer.

Een van de problemen van AI is te veel optimisme. Het probleem dat wetenschappers met AI proberen op te lossen, is waanzinnig ingewikkeld. In de jaren vijftig en zestig van de vorige eeuw waren wetenschappers er bijvoorbeeld van overtuigd dat er binnen 20 jaar intelligente apparaten zouden zijn. Machines deden immers allerlei ongelofelijke dingen, zoals het spelen van gecompliceerde spellen. AI boekt momenteel de grootste successen op gebieden als logistiek, datamining en medische diagnoses.

Verkennen wat machine learning voor AI kan betekenen

Machine learning maakt gebruik van algoritmen om enorme datasets te analyseren. Maar machine learning is nu nog niet in staat om het soort AI te bieden dat je in speelfilms ziet. Zelfs de beste algoritmen kunnen niet nadenken, zelfinzicht tonen of een eigen mening uiten. Machine learning is wel veel sneller in het uitvoeren van voorspellende analyses, waardoor mensen een stuk efficiënter kunnen werken. Momenteel voert AI dus analyses uit, maar moeten mensen hierna nog morele en ethische beslissingen nemen. In de paragraaf ‘De relatie tussen AI en machine learning’ verderop in dit hoofdstuk lees je meer over de manier waarop machine learning aan AI bijdraagt. Het komt erop neer dat machine learning het leerproces van AI voor zijn rekening neemt en dat komt natuurlijk niet in de buurt van de kunstmatige intelligentie die je in films te zien krijgt.



BELANGRIJK

De belangrijkste reden dat sommige mensen leren en intelligentie door elkaar halen, is dat ze ervan uitgaan dat een apparaat dat beter in iets wordt (en dus leert) ook zelfbewust wordt (en intelligent). Maar deze zwijgzaamheid over machine learning is nergens op gebaseerd. Je ziet hetzelfde fenomeen als mensen denken dat een computer bewust problemen voor ze veroorzaakt. Maar een computer kent geen emoties en reageert dus alleen op de invoer die hij ontvangt en de instructies die een programma geeft om deze invoer te verwerken. Echte AI zal ooit ontstaan als computers uiteindelijk in staat zijn de slimme combinatie uit de natuur na te bootsen:

» **Genetica.** Langzaam leren van generatie op generatie.

- » **Onderwijs.** Snel leren van georganiseerde bronnen.
- » **Verkenning.** Spontaan leren door middel van media en interactie met anderen.

De doelen van machine learning

Momenteel is AI op machine learning gebaseerd en verschilt machine learning wezenlijk van statistiek. Machine learning heeft natuurlijk een statistische basis, maar maakt andere aannamen dan statistiek doordat hun doelen verschillen. In tabel 1.1 vind je een vergelijking tussen machine learning en statistiek.

TABEL 1.1 **Vergelijking tussen machine learning en statistiek**

Techniek	Machine learning	Statistiek
Gegevensverwerking	Werkt met big data in de vorm van netwerken en grafen; ruwe data van sensoren of het web worden opgedeeld in trainings- en testgegevens.	Er worden modellen gebruikt om voorspellingen te doen bij kleine steekproeven.
Gegevensinvoer	Er wordt een aselechte steekproef genomen en getransformeerd om de nauwkeurigheid te maximaliseren van de voorspelling op compleet nieuwe voorbeelden.	Parameters interpreteren bestaande fenomenen en leggen de nadruk op magnitude.
Resultaat	Er wordt rekening gehouden met waarschijnlijkheid bij het vergelijken wat de beste uitkomst of beslissing kan zijn.	Bij de uitvoer wordt rekening gehouden met de verscheidenheid en onzekerheid van parameters.
Aannamen	De wetenschapper leert iets van de gegevens.	De wetenschapper maakt een aanname over de uitvoer en probeert deze te bewijzen.
Distributie	De distributie is onbekend of wordt genegeerd voordat iets van de gegevens wordt geleerd.	De wetenschapper gaat uit van een goed gedefinieerde distributie.
Uitkomst	De wetenschapper maakt het best passende, maar generaliseerbare, model.	Het resultaat wordt in de huidige gegevensdistributie ingepast.

Machine learning op de gebruikte hardware afstemmen

Enorme datasets vergen enorme hoeveelheden geheugen. Helaas blijft het hier niet bij. Bij grote hoeveelheden gegevens en geheugen heb je ook processors met meerdere kernen en hoge snelheden nodig. Een van de pro-

blemen die wetenschappers proberen op te lossen, is bestaande hardware efficiënter te gebruiken. In sommige gevallen is het niet mogelijk om dagen te wachten op een resultaat van machine learning. Wetenschappers hebben soms snel antwoorden nodig, zelfs als het resultaat niet helemaal correct is. Maar investeringen in betere hardware vergen ook investeringen in betere wetenschap. In dit boek maak je kennis met de volgende aandachtspunten om je ervaringen met machine learning te verbeteren:

- » **Een nuttig resultaat behalen.** Je ontdekt dat je eerst een bruikbaar resultaat nodig hebt voordat je het kunt verfijnen. Bovendien kan het verbeteren van een algoritme te ver worden doorgevoerd, waardoor het resultaat erg kwetsbaar wordt (en misschien wel onbruikbaar buiten een specifieke dataset).
- » **De juiste vraag stellen.** Veel mensen ontvangen van machine learning niet de gewenste antwoorden omdat ze bij het verfijnen van hun algoritme vergeten om andere vragen te stellen. Bij een efficiënt gebruik van hardware is het soms nuttig om een stapje terug te doen en de vraag die je stelt onder de loep te nemen. Misschien stel je de verkeerde vraag; in zo'n geval kan zelfs de beste hardware het juiste antwoord niet geven.
- » **Te veel waarde aan intuïtie hechten.** Alle vragen voor machine learning beginnen als hypothese. Een wetenschapper gebruikt zijn intuïtie als uitgangspunt bij het ontdekken van het antwoord op een vraag. Mislukkingen komen meer voor dan succes tijdens het werken met machine learning. Je intuïtie voegt creativiteit aan machine learning toe, maar soms klopt je intuïtie niet en moet je je aannamen herzien.



Besef je eenmaal het belang van de omgeving op machine learning, dan ben je ook in staat om het belang in te zien van de juiste hardware en van de juiste balans om het gewenste resultaat te bereiken. Moderne systemen maken bij machine learning gebruik van GPU's (Graphical Processing Units). GPU's zorgen ervoor dat het proces een stuk sneller gaat. Het past niet in dit boek om gedetailleerd op GPU's in te gaan; je leest er meer over op <https://devblogs.nvidia.com/parallelforall/bidmach-machine-learning-limit-gpus/>.

Fantasiën over AI herkennen

Net als bij veel andere technologieën zijn er zowel bij AI als machine learning speelse en fantasierijke toepassingen. Sommige mensen gebruiken machine learning bijvoorbeeld om foto's in Picasso-achtige kunst om te toveren. Je leest hier meer over op <https://www.washingtonpost.com/news/innovations/wp/2015/08/31/this-algorithm-can-create-a-new->

van-gogh-or-picasso-in-just-an-hour/. Hier valt natuurlijk wel het een en ander op aan te merken. Ten eerste is het twijfelachtig of mensen dit soort kunstuitingen serieus nemen, want kunst bestaat niet alleen uit het maken van een interessante interpretatie van een bestaande voorstelling, maar meer uit de interpretatie van de kunstenaar zelf. Aan het eind van het artikel wordt erop gewezen dat de computer op dit punt alleen nog maar een bestaande stijl kan interpreteren en dus niet in staat is om zelf een nieuwe stijl te ontwikkelen. In de volgende paragrafen lees je meer over verschillende fantasieën over AI en machine learning.

Populaire toepassingen van AI en machine learning

AI leidt tot een tijdperk van innovatie dat vroeger alleen in sciencefiction voorkwam. Maar het is soms moeilijk om te bepalen of een specifieke AI-toepassing echt is of alleen maar een droom van een vastberaden wetenschapper. *De man van zes miljoen* (https://nl.wikipedia.org/wiki/Man_van_Zes_Miljoen) is een televisieprogramma dat ooit futuristisch leek. Toen het werd uitgezonden, dacht niemand dat er ooit gebruik van bionica zou worden gemaakt. Maar Hugh Herr heeft andere plannen: bionische benen zijn tegenwoordig echt mogelijk (<https://www.smithsonianmag.com/innovation/future-robotic-legs-180953040/>). Ze zijn natuurlijk nog niet voor iedereen beschikbaar; de gebruikte technologie wordt nu pas nuttig. Een andere tv-serie maakt het nog ingewikkelder: *The Six Billion Dollar Man* (<https://www.cinemablend.com/new/Mark-Wahlberg-Six-Billion-Dollar-Man-Just-Made-Big-Change-91947.html>). Het is duidelijk dat AI en machine learning de weg bereiden voor ongelofelijke technologieën waar nu al aan wordt gewerkt, al moet je wat je erover leest wel met een korreltje zout nemen.



BELANGRIJK

Om het toekomstige gebruik van AI en machine learning te laten overeenkomen met de concepten die sciencefiction in de loop van de jaren heeft gepresenteerd, moeten programmeurs, datawetenschappers en andere belanghebbenden hulpmiddelen ontwikkelen. In hoofdstuk 8 lees je meer over enkele nieuwe tools voor het werken met AI en machine learning, al zijn deze hulpmiddelen nog erg beperkt. Er komt geen magie bij kijken, al kan iets op tovenarij lijken als je niet weet wat er achter de schermen gebeurt. Om ervoor te zorgen dat speelse toepassingen van AI en machine learning echt nuttige toepassingen worden, moeten ontwikkelaars, wetenschappers en anderen doorgaan met het bouwen van nuttige hulpmiddelen die op dit moment misschien moeilijk voorstelbaar zijn.

Praktische toepassingen van AI en machine learning

AI en machine learning maken tegenwoordig deel uit van een groot aantal toepassingen. Het enige probleem is dat de technologie zo goed werkt dat je niet eens doorhebt dat ze worden gebruikt. Misschien komt het als een verrassing dat veel apparaten in je huis al gebruikmaken van beide technologieën. Ze worden ook in je auto en op kantoor gebruikt. Er zijn al miljoenen toepassingen voor zowel AI als machine learning, al blijven die meestal veilig uit het zicht, zelfs als ze behoorlijk spectaculair van aard zijn. Hier volgen slechts enkele manieren waarop je AI kunt gebruiken:

- » **Fraudedetectie.** Je ontvangt een telefoontje van de bank met de vraag of je een specifieke aankoop hebt gedaan. De bank is niet nieuwsgierig, maar wijst je alleen op de mogelijkheid dat iemand anders misschien je creditcard heeft gebruikt. De AI die deel uitmaakt van de code van de bank detecteerde een onbekend uitgavenpatroon en gaf hierover een waarschuwing.
- » **Resourceplanning.** Veel organisaties moeten het inzetten van de beschikbare bronnen efficiënt plannen. Een ziekenhuis kan bijvoorbeeld aan de hand van de behoeften van een patiënt, de beschikbaarheid van specialisten en de hoeveelheid tijd die de patiënt verwacht wordt te blijven, vaststellen waar die patiënt terecht moet komen.
- » **Complexe analyse.** Mensen hebben bij complexe analyses vaak hulp nodig omdat er letterlijk te veel factoren zijn om rekening mee te houden. Dezelfde verzameling symptomen kan bijvoorbeeld bij meerdere problemen horen. Een arts of andere expert heeft bij het tijdig stellen van een diagnose soms hulp nodig als hij het leven van een patiënt wil redden.
- » **Automatisering.** Elke vorm van automatisering heeft baat bij AI om onverwachte veranderingen of gebeurtenissen te verwerken. Een probleem met sommige moderne vormen van automatisering is dat een onverwachte gebeurtenis, zoals een object dat zich op de verkeerde plaats bevindt, het automatiseringsproces kan laten stoppen. Door er AI aan toe te voegen, kan de automatisering onverwachte gebeurtenissen afhandelen en doorgaan alsof er niets is gebeurd.
- » **Klantenservice.** Als je tegenwoordig de klantenservice belt, krijg je misschien niet eens een mens aan de lijn. De automatisering ervan is zo goed dat de gebruikte scripts en bronnen de overgrote meerderheid van je vragen kan beantwoorden. Met de juiste intonatie (die AI ook levert) heb je misschien niet eens door dat je met een computer praat.
- » **Veiligheidssystemen.** Veel veiligheidssystemen in moderne apparaten vertrouwen op AI om in noodsituaties de bediening over te nemen.

Automatische remsystemen gebruiken tegenwoordig bijvoorbeeld vaak AI om de auto te stoppen aan de hand van gegevens die het voertuig biedt, zoals de richting van een slip.

- » **Machine-efficiëntie.** AI kan bij de besturing van een machine voor optimale efficiëntie zorgen. Dankzij de kunstmatige intelligentie worden bronnen zo ingezet dat het systeem de snelheid of andere doelen niet overschrijdt. Elk beetje vermogen wordt precies zo gebruikt als nodig is om de gewenste diensten te leveren.

Deze lijst is niet meer dan het begin van de mogelijkheden. Kunstmatige intelligentie wordt ook op veel andere manieren gebruikt. Maar het kan ook geen kwaad om toepassingen van machine learning te bekijken die buiten het normale bereik vallen van wat velen als het domein van AI beschouwen. De volgende toepassingen van machine learning associeer je misschien niet met AI:

- » **Toegangscontrole.** In veel gevallen is toegangscontrole een kwestie van ja of nee. De smartcard van een werknemer geeft op een vergelijkbare manier toegang tot een bron als de manier waarop mensen al eeuwen sleutels gebruiken. Sommige sloten bieden de mogelijkheid om tijden en datums in te stellen waarop de toegang mogelijk is, maar eenvoudig toegangsbeheer is niet voor alle doeleinden geschikt. Met machine learning kun je aan de hand van de functie en behoefte van een werknemer vaststellen of hij toegang tot een bron moet krijgen. Een werknemer krijgt dan bijvoorbeeld toegang tot een trainingsruimte als de training bij zijn functie past.
- » **Dierenbescherming.** Oceanen lijken groot genoeg om dieren en schepen probleemloos te laten samen zijn. Toch worden jaarlijks nog veel dieren door schepen aangevaren. Een algoritme voor machine learning kan schepen leren dieren te vermijden aan de hand van de kenmerken en geluiden van zowel dieren als schepen.
- » **Wachttijden voorspellen.** De meeste mensen hebben er een hekel aan te wachten als ze geen flauw idee hebben hoelang dit gaat duren. Met machine learning kan een toepassing de wachttijd bepalen aan de hand van het aantal beschikbare werknemers, hun hoeveelheid werk, de complexiteit van de problemen die ze proberen op te lossen, de beschikbaarheid van bronnen enzovoort.

Nuttig en alledaags zijn

In een speelfilm krijg je al snel het idee dat AI een enorme impact gaat maken en er zijn in het echt inderdaad enkele ongelofelijke toepassingen van AI, maar over het algemeen zijn de meeste toepassingen ervan nogal alle-

daags, om niet te zeggen ronduit saai. In een recent artikel wordt bijvoorbeeld beschreven hoe Verizon de taal R (zie hoofdstukken 4 en 5 voor meer informatie over R) gebruikt om gegevens over beveiligingslekken te analyseren (<https://www.computerworld.com/article/3001832/data-analytics/how-verizon-analyzes-security-breach-data-with-r.html>). In deel 5 lees je meer over praktische toepassingen van dit soort analyses. Het uitvoeren van deze analyses is nogal saai als je het vergelijkt met andere mogelijkheden die AI te bieden heeft, maar het voordeel ervan is dat Verizon geld bespaart met de analyse in R en dat de resultaten ervan bovendien beter zijn.

Python-ontwikkelaars (zie hoofdstukken 6 en 7 voor meer informatie over de taal Python) hebben een enorme hoeveelheid bibliotheken tot hun beschikking die machine learning vereenvoudigen. De website Kaggle (<https://www.kaggle.com/competitions>) organiseert wedstrijden voor Python-ontwikkelaars en gebruikers van R om hun vaardigheden op het gebied van machine learning in te zetten bij het maken van praktische toepassingen. Het resultaat van deze wedstrijden kom je achteraf vaak tegen in producten die mensen daadwerkelijk gebruiken. R kan nog niet zonder ondersteuning van wetenschappers met statistische kennis, maar softwareontwikkelaars zijn al druk bezig voor Python nieuwe bibliotheken te maken die het werken met complexe gegevens en machine learning vereenvoudigen (zie <http://www.kdnuggets.com/2015/06/top-20-python-machine-learning-open-source-projects.html> voor de top 20 met Python-bibliotheken).

De relatie tussen AI en machine learning

Machine learning is slechts een van de vereisten om een AI-systeem te worden. Het onderdeel machine learning stelt een apparaat met AI in staat om de volgende taken uit te voeren:

- » zich aanpassen aan nieuwe omstandigheden die de oorspronkelijke ontwerper niet heeft voorzien;
- » patronen in allerlei gegevensbronnen detecteren;
- » nieuw gedrag creëren aan de hand van herkende patronen;
- » beslissingen nemen aan de hand van het succes of falen van dit gedrag.

Algoritmen die gegevens manipuleren vormen de kern van machine learning. Bij machine learning moet een geschikt algoritme worden

gebruikt om het gewenste resultaat te bereiken. Bovendien moeten de gegevens geschikt zijn voor analyse met het gewenste algoritme, dat anders zorgvuldig door een wetenschapper moet worden voorbereid.

AI omvat veel andere disciplines om het denkproces met succes te simuleren. Naast machine learning maakt AI over het algemeen ook gebruik van:

- » **Verwerking natuurlijke taal.** Het toestaan van taalinput en het omzetten in een formaat dat bruikbaar is voor een computer.
- » **Begrip natuurlijke taal.** Het ontcijferen van de taal zodat er op de geboden betekenis kan worden gereageerd.
- » **Kennisrepresentatie.** Het vermogen om informatie zo op te slaan dat snelle toegang mogelijk is.
- » **Planning (in de vorm van doel zoeken).** Het vermogen om aan de hand van opgeslagen informatie bijna onmiddellijk conclusies te trekken (dus met een korte onderbreking, die mensen niet opmerken, maar een computer wel).
- » **Robotica.** De mogelijkheid om in een fysieke vorm op verzoeken van een gebruiker te reageren.

Het zal je waarschijnlijk verbazen hoeveel disciplines er nodig zijn om een AI-systeem tot stand te brengen. Om die reden lees je in dit boek slechts over een deel van wat een AI-systeem bevat. Maar ook het onderdeel machine learning kan heel ingewikkeld worden, want het is een complexe taak om de wereld via de gegevensinput van een computer te leren kennen. Denk maar aan alle beslissingen die je voortdurend neemt zonder erover na te denken. Alleen al het concept van iets zien en weten of je er succesvol mee kunt communiceren, wordt een complexe taak.

De specificaties van AI en machine learning

Terwijl wetenschappers blijven werken aan een technologie en hypothesen in theorieën omzetten, wordt de technologie meer *techniek* (waarbij theorieën worden geïmplementeerd) dan *wetenschap* (waarbij theorieën worden gemaakt). Naarmate de regels voor een technologie duidelijker worden, werken groepen deskundigen samen om deze regels schriftelijk te definiëren. Het resultaat hiervan zijn *specificaties* (een groep regels waar iedereen het over eens is).

Vervolgens worden implementaties van deze specificaties tot norm of *standaard* die beheerd worden door een bestuursorgaan, zoals de IEEE (Institute of Electrical and Electronics Engineers) of een combinatie van de ISO/IEC (International Organization for Standardization/International Electrotechnical Commission). AI en machine learning bestaan allebei lang genoeg voor het opstellen van specificaties, al zijn er momenteel nog geen standaarden voor beide technologieën ontwikkeld.

De basis voor machine learning is wiskunde. Algoritmen bepalen hoe big data op specifieke manieren moeten worden geïnterpreteerd. In deel 3 van dit boek lees je meer over de wiskundige basis voor machine learning. Je ontdekt dat algoritmen invoergegevens op specifieke manieren verwerken en voorspelbare uitvoer leveren aan de hand van gegevenspatronen. Alleen de gegevens zelf zijn niet voorspelbaar. Je hebt AI en machine learning nodig om de gegevens op zo'n manier te ontcijferen dat je de patronen erin kunt zien en ze kunt begrijpen.

In deel 4 vind je meer informatie over de specificaties in de vorm van algoritmen die specifieke taken uitvoeren. Heb je eenmaal deel 5 bereikt, dan snap je de reden dat iedereen instemt met specifieke sets regels voor het gebruiken van algoritmen om bepaalde taken uit te voeren. Het gaat erom dat je een algoritme gebruikt dat het best past bij je gegevens om de specifieke doelen te bereiken die je hebt opgesteld. Professionals implementeren algoritmen met behulp van talen die het best bij een specifieke taak passen. Machine learning maakt gebruik van Python en R, en tot op zekere hoogte ook van Matlab, Java, Julia en C++. (Meer hierover lees je op <https://www.quora.com/What-is-the-best-language-to-use-while-learning-machine-learning-for-the-first-time>.)

Onderscheid maken tussen creativiteit en techniek

De reden dat AI en machine learning allebei tot de wetenschap in plaats van de techniek worden gerekend, is dat er een zekere mate van creativiteit is vereist om goede resultaten te bereiken. Het artistieke element van machine learning uit zich in verschillende vormen. Je moet er bijvoorbeeld rekening mee houden hoe de gegevens worden gebruikt. Sommige gegevens fungeren als referentie om een algoritme zo te trainen dat specifieke resultaten worden bereikt. De overige gegevens bieden de uitvoer die wordt gebruikt om de onderliggende patronen te begrijpen. Er zijn geen specifieke regels voor het wegen van gegevens; wetenschappers die met de gegevens werken, moeten zelf bepalen of een specifieke verhouding optimale uitvoer produceert.



Het opschonen van gegevens geeft ook een zekere artistieke kwaliteit aan het resultaat. De manier waarop een wetenschapper de gegevens voor gebruik voorbereidt, is ook belangrijk. Sommige taken komen regelmatig voor, zoals het verwijderen van dubbele records. Een wetenschapper kan er ook voor kiezen om gegevens op een bepaalde manier te filteren of alleen naar een deel van de gegevens te kijken. Hierdoor kan de opgeschoonde dataset voor machine learning van de ene wetenschapper afwijken van de opgeschoonde dataset van een andere wetenschapper.

Je kunt de algoritmen ook op verschillende manieren afstellen of de werking ervan verfijnen. Het gaat erom dat er uitvoer wordt geproduceerd die de gewenste patronen onthult, zodat je de gegevens kunt begrijpen. Wordt er bijvoorbeeld een afbeelding bekeken, dan moet een robot bepalen met welke elementen van de afbeelding hij wel en niet kan communiceren. Het antwoord op deze vraag is belangrijk als de robot bepaalde elementen moet vermijden om op schema te blijven of om specifieke doelen te bereiken.

Bij het werken in een omgeving voor machine learning moet je ook rekening houden met invoergegevens. De microfoon in een smartphone van merk A produceert bijvoorbeeld niet precies dezelfde invoergegevens als de microfoon in een smartphone van merk B. De kenmerken van de microfoons verschillen, maar het interpreteren van de stemopdrachten van de gebruiker moeten wel hetzelfde resultaat geven. Omgevingsgeluiden veranderen de invoerkwaliteit van de gesproken opdracht en de smartphone kan ook te maken hebben met bepaalde vormen van elektromagnetische interferentie. Je begrijpt dat de variabelen waarmee een ontwerper bij het creëren van een omgeving voor machine learning geconfronteerd wordt zowel groot als complex zijn.

De kunst achter de techniek is een onmisbaar onderdeel van machine learning. De ervaring die een wetenschapper bij het werken met gegevensproblemen opdoet, is essentieel omdat het hem de mogelijkheid biedt om waarden toe te voegen die het algoritme beter laten werken. Een nauwkeurig afgestemd algoritme kan het verschil betekenen tussen een robot die met succes een pad langs obstakels neemt en een robot die ze allemaal raakt.