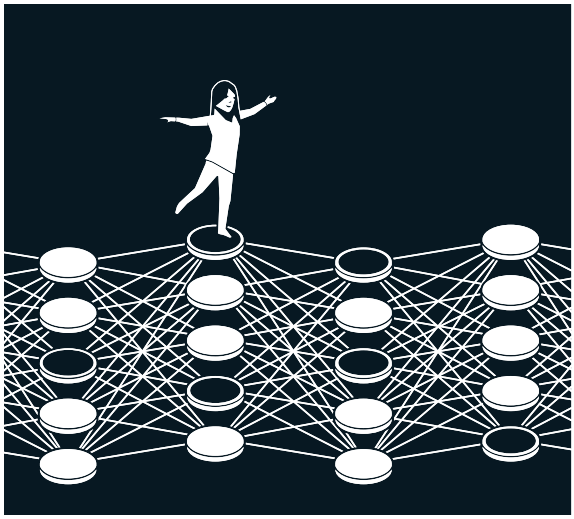
De använde fysiken för att hitta mönster i information

Årets pristagare har använt sig av fysikens verktyg för att konstruera metoder som ligger till grund för dagens kraftfulla maskininlärning. ***John Hopfield*** *skapade en struktur som kan lagra och återskapa information.* ***Geoffrey Hinton*** *uppfann en metod som självständigt kan hitta egenskaper i data och som blivit viktig för de stora artificiella neuronnät som används i dag.*



Numera har många sett hur datorer kan översätta mellan språk, tolka bilder och rentav föra en rimlig konversation. Det som kanske inte är lika känt är att samma typ av teknik länge har varit viktig inom forskningen, bland annat för att sortera och analysera stora datamängder. Utvecklingen av maskininlärning har exploderat de senaste femton till tjugo åren, och utnyttjar en typ av struktur som kallas artificiella neuronnät. När vi i dag pratar om *artificiell intelligens* är det ofta just den här typen av teknik vi syftar på.

Även om datorer inte kan tänka kan maskiner i dag efterlikna funktioner som minne och inlärning. Årets fysikpristagare har bidragit till att göra detta möjligt. Genom att använda grundläggande begrepp och metoder från fysiken har de utvecklat tekniker som använder strukturer i nätverk för att behandla information.

Maskininlärning skiljer sig från traditionella datorprogram som fungerar som en sorts recept. Programmen tar emot data som behandlas enligt en tydlig beskrivning och matar ut resultatet, ungefär som när en människa tar ingredienser och behandlar dem enligt receptets instruktioner för att få ett färdigt bakverk. Maskininlärning låter i stället datorn lära sig från exempel. På så sätt går det att tackla problem som är för diffusa och komplicerade för att hantera med steg för steg-instruktioner. Ett exempel är att tolka en bild och identifiera avbildade föremål i den.

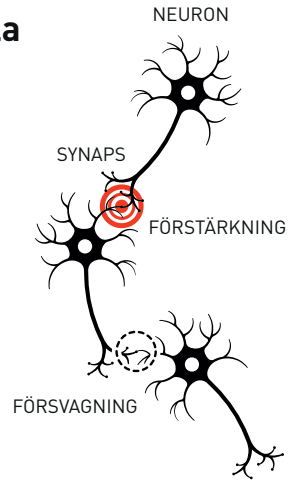
Efterliknar hjärnan

I ett artificiellt neuronnät behandlas informationen av strukturen i nätverket som helhet. Inspirationen kommer från början från försök att förstå hur hjärnan fungerar. Redan på 1940-talet hade forskare börjat resonera om matematiken bakom hjärnans nätverk av neuroner och synapser. En annan pusselbit kom från psykologen och hjärnforskaren Donald Hebb, som lade fram en hypotes om att inlärning sker genom att kopplingarna mellan neuroner förstärks när de arbetar tillsammans.

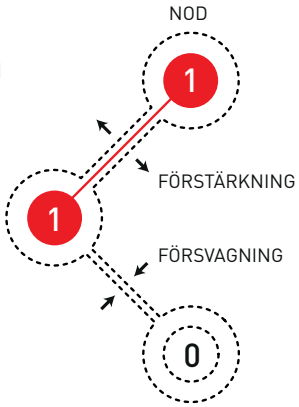
De här idéerna följdes senare av försök att återskapa funktionen i hjärnans nätverk genom att bygga upp artificiella neuronnät som en simulering i en dator. Där efterliknas hjärnans neuroner av noder som får anta olika värden, och synapserna motsvaras av kopplingar mellan noderna som kan göras starkare eller svagare. Donald Hebbs hypotes används än i dag som en av de grundläggande reglerna för att uppdatera artificiella neuronnätverk i en process som kallas *träning*.

Naturliga neuroner och artificiella

Neuronerna i hjärnan är levande celler som har ett avancerat inre maskineri. De kan skicka signaler till varandra genom synapser. Vid inlärning stärks kopplingarna mellan vissa neuroner, och andra försvagas.



Artificiella neuronnät är uppbyggda av noder som kodas med ett värde. Noderna har kopplingar till varandra. När nätverket tränas görs kopplingarna starkare mellan noder som är aktiva samtidigt, annars görs de svagare.



I slutet av 1960-talet kom några nedslående teoretiska resultat som fick många forskare att misstänka att sådana neuronnät aldrig skulle bli verkligt användbara. Intresset för artificiella neuronnät svalnade men väcktes igen på 1980-talet när flera viktiga idéer fick genomslag, bland dem arbeten av årets fysikpristagare.

Associativt minne

Tänk dig att du försöker komma ihåg ett lite ovanligt ord som du använder sällan, som vad det där lutande golvet heter som ofta finns i biosalonger och föreläsningssalar. Du söker i minnet. Det påminner om *gradient* … kanske *grad…ering*? Inte riktigt rätt. *Gradäng*, där har vi det!

Den här processen att söka bland liknande ord för att hitta det rätta påminner om det associativa minne fysikern John Hopfield uppfann 1982. *Hopfieldnätverket* kan lagra mönster och har en metod för att återskapa dem. När nätverket matas med ett ofullständigt eller lite trasigt mönster kan metoden hitta det av de lagrade mönstren som är mest likt.

John Hopfield hade tidigare använt sin bakgrund i fysik för att utforska teoretiska problem inom molekylärbiologi. När han blev inbjuden till ett möte om neurovetenskap kom han i kontakt med forskning om hjärnans struktur. Där blev han fascinerad av vad han fick höra, och började fundera över dynamiken i enkla nätverk av neuroner. När neuronerna agerar tillsammans kan de ge upphov till nya kraftfulla egenskaper som inte är uppenbara för den som bara tittar på nätverkets enskilda delar.

1980 lämnade John Hopfield sin tjänst vid Princeton University, där hans forskningsintressen hade fört honom utanför de områden som hans fysikkollegor ägnade sig åt, och flyttade tvärs över kontinenten. Han hade nämligen antagit ett erbjudande om att bli professor i kemi och biologi vid Caltech (California Institute of Technology) i Pasadena i södra Kalifornien. Där fick han tillgång till datorresurser som han kunde använda för att fritt experimentera och utveckla sina idéer om neuronnät.

Han lämnade däremot inte sin förankring i fysiken, där han hämtade inspiration för att förstå system med många smådelar som tillsammans kan ge upphov till nya intressanta fenomen. Särskilt hade han nytta av att ha lärt sig saker om magnetiska material som får speciella egenskaper genom atomernas *spinn* – en egenskap som gör varje atom till en liten magnet. Spinnet i angränsande atomer återkopplar till varandra, så att det till exempel kan bildas domäner med spinn i samma riktning. Han kunde använda fysiken som beskriver hur material utvecklas när spinnen påverkar varandra för att göra en modell av ett nätverk med noder och deras kopplingar.

Nätverket sparar bilder i ett landskap

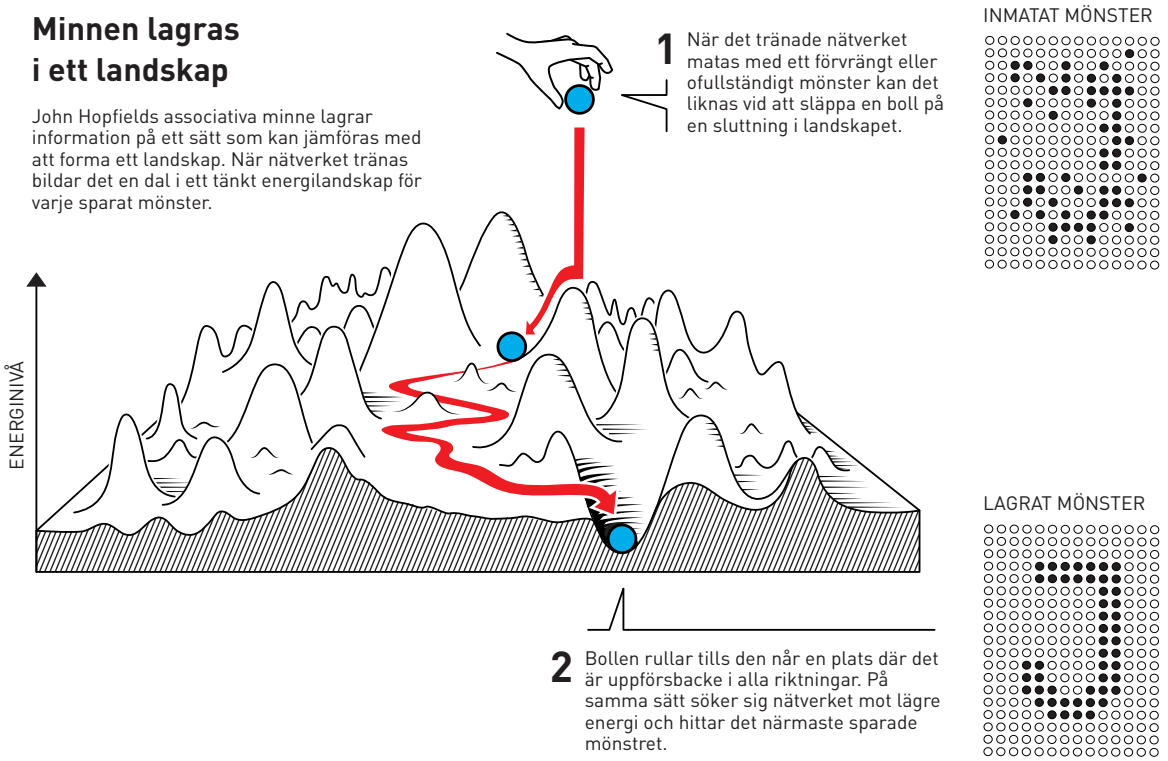
Nätverket som John Hopfield ställde upp innehåller noder som alla är hoplänkade med alla de övriga genom kopplingar som kan ha olika styrka. De olika noderna kan lagra varsitt värde – i John Hopfields första arbete kunde värdena vara antingen 0 eller 1, som bildpunkter i en svartvit bild.

John Hopfield beskrev tillståndet i nätverket som helhet med en egenskap som fungerar precis som energin i fysikens spinnsystem. Energin beräknas med en formel som använder alla värdena i noderna och alla styrkor på kopplingarna mellan dem. Hopfieldnätverket programmeras genom att en bild matas in i noderna som ges värdet svart (0) eller vitt (1). Sedan justeras kopplingarna i nätverket med hjälp av energiformeln så att bilden som ska sparas får låg energi. När nätverket sedan matas med ett annat mönster finns en regel för att gå igenom noderna en efter en och kontrollera om nätverket som helhet får lägre energi om värdet i den aktuella noden ändras. Om det visar sig att energin blir lägre om just den här svarta punkten blir vit i stället så får den alltså växla färg. Proceduren fortsätter på det viset ända tills det inte längre går att hitta några ändringar som sänker energin. När det läget nås har nätverket ofta återställt den ursprungliga bilden som det tränades på.

Om man bara sparar ett enda mönster verkar det här kanske inte så märkvärdigt. Det ligger nära till hands att undra varför man inte bara sparar själva bilden och jämför med en annan bild som ska testas. Det speciella med John Hopfields metod är att det går att spara flera bilder samtidigt och att nätverket oftast kan skilja mellan dem.

Att söka igenom nätverket efter ett sparat tillstånd liknade John Hopfield vid att rulla en kula i ett landskap med toppar och dalar och en viss friktion som bromsar rörelsen. Släpps kulan på ett visst ställe i landskapet kommer den att rulla ner i den närmaste dalen och stanna där. Om nätverket ges ett mönster som ligger nära ett av de sparade mönstren kommer det på samma sätt att stega sig fram tills det har hamnat längst ner i en dal i energilandskapet och därmed hittat det närmaste mönstret i sitt minne.

Hopfieldnätverket kan användas för att återskapa data som innehåller brus eller som delvis har suddats ut.



John Hopfield och andra har fortsatt att utveckla detaljerna i Hopfieldnätverkets funktion. Bland annat kan nätverket använda noder som kan lagra vilket värde som helst och inte bara noll eller ett. Om du tänker på noderna som pixlar i en bild kan de ha olika färger i stället för bara svart eller vitt. Förbättrade metoder gör att det går att spara fler bilder och att skilja mellan dem även om de är ganska lika. Det går lika bra att identifiera eller rekonstruera vilken information som helst som är uppbyggd av många datapunkter.

Klassificering med hjälp av 1800-talsfysik

Att komma ihåg en bild är en sak, men att tolka vad den föreställer kräver lite mer.

Barn kan redan som ganska små peka på olika djur och självsäkert säga att det är en hund, eller katt, eller ekorre. Ibland blir det fel, men ganska snart är det rätt nästan hela tiden. Barnet lär sig det utan att ha fått se några diagram eller förklaringar av begrepp som *djurart* eller *pälsdjur*. Efter att barnet har stött på några exempel på varje sorts djur faller de olika kategorierna på plats i huvudet. Genom erfarenheter av omgivningen lär sig människor känna igen en katt, eller att uppfatta ett ord, eller att komma in i ett rum och märka att något inte är som det brukar.

När John Hopfield publicerade artikeln om sitt associativa minne arbetade Geoffrey Hinton vid Carnegie Mellon University i Pittsburgh, USA. Tidigare hade han studerat experimentell psykologi och artificiell intelligens i England och Skottland. Han funderade över hur maskiner kan lära sig att hantera mönster på liknande sätt som människor, och hitta egna kategorier för att sortera information och tolka den. Tillsammans med kollegan Terrence Sejnowski utgick Geoffrey Hinton från Hopfield-nätverket och utvidgade det till något nytt med hjälp av fler idéer från statistisk fysik.

Den statistiska fysiken beskriver system som består av många likadana delar, som till exempel molekyler i en gas. Det är svårt eller omöjligt att spåra alla individuella molekyler i gasen, men det går att betrakta dem som en gemensam helhet och beräkna övergripande egenskaper som tryck eller temperatur. Det finns många olika möjligheter för gasmolekylerna att sprida ut sig i volymen med olika individuella hastigheter, som ändå ger upphov till samma övergripande egenskaper hos gasen.

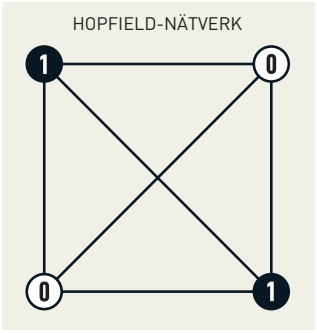
Med statistisk fysik går det att analysera vilka olika tillstånd som de individuella delarna gemensamt kan befinna sig i, och räkna ut hur troligt det är att de uppstår. Vissa tillstånd är mer sannolika än andra. Det beror på hur mycket energi som finns tillgänglig, vilket beskrivs med en ekvation från 1800-talsfysikern Ludwig Boltzmann. Geoffrey Hintons nätverk utnyttjade just den här ekvationen, och metoden publicerades 1985 med det slående namnet *Boltzmannmaskinen*.

Att känna igen nya exempel av samma sort

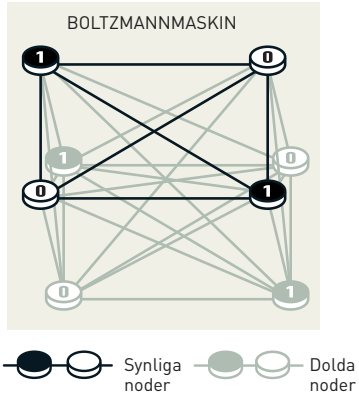
Boltzmannmaskinen används vanligtvis med två olika typer av noder. Information matas in till den ena gruppen, som kallas synliga noder. De övriga enheterna bildar ett dolt lager. De dolda nodernas värden och kopplingar bidrar också till energin för nätverket som helhet.

Maskinen körs genom att tillämpa en regel för att uppdatera värdena i de ingående noderna en i taget. Så småningom kommer maskinen att hamna i ett tillstånd där mönstret i noderna kan förändras men egenskaperna för nätverket som helhet förblir desamma. Då kommer varje möjligt mönster att ha en viss sannolikhet som bestäms av nätverkets energi genom Boltzmanns ekvation. När maskinen har stannat har den skapat ett nytt mönster. Boltzmannmaskinen är alltså ett tidigt exempel på en generativ modell.

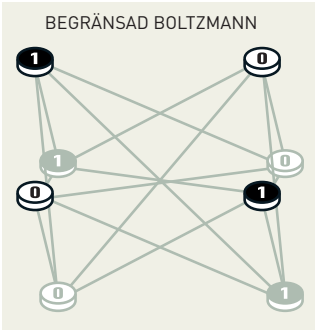
Olika sorters nätverk



John Hopfields associativa minne är uppbyggt så att alla noder är kopplade till alla övriga. Information matas in och läses ut från alla noderna.



Geoffrey Hintons Boltzmannmaskin sätts med fördel upp i två lager, där information matas in och läses ut på en grupp av noder som kallas *synliga*. De är kopplade till *dolda* noder, som påverkar hur nätverket fungerar som helhet.



I en begränsad Boltzmannmaskin finns inga kopplingar mellan noder i samma lager. Det är vanligt att använda sådana maskiner i en kedja efter varandra. Efter träning av den första begränsade Boltzmann-maskinen används de dolda nodernas innehåll för att träna nästa, och så vidare.

Boltzmannmaskinen kan utan instruktioner lära sig något genom att få se ett antal exempel. Träningen görs genom att uppdatera värdena i nätverkets kopplingar på ett sådant sätt att de exempelmönster som matas in i de synliga noderna under träningen får så hög sannolikhet som möjligt att uppstå när maskinen sedan körs. Om samma mönster skulle komma tillbaka flera gånger i träningen blir sannolikheten för just det mönstret ännu högre. Träningen påverkar också sannolikheten för att mata ut nya mönster som liknar de exempel som maskinen har tränats på.

Den tränade Boltzmannmaskinen kan känna igen bekanta drag i ny information som den inte har sett tidigare. Tänk dig att du träffar ett syskon till din kompis, och att du genast ser att de måste vara släkt. På liknande sätt kan Boltzmannmaskinen känna igen ett helt nytt exempel om det hör till någon kategori som finns i träningsmaterialet, och skilja det från material som är helt olikt.

I sin ursprungliga form är Boltzmannmaskinen ganska ineffektiv och tar lång tid på sig att hitta lösningar. Den blir intressantare när den utvecklas på olika sätt, vilket Geoffrey Hinton har fortsatt att utforska. Senare versioner är utglesade, genom att kopplingarna mellan vissa av enheterna har tagits bort. Det visar sig att det kan göra maskinen ännu mer effektiv.

Under 1990-talet tappade många forskare intresset för artificiella neuronnät, men Geoffrey Hinton var en av dem som fortsatte att arbeta på det spåret. Han bidrog också till att sätta fart på den nya explosionen av intressanta resultat. Bland annat utvecklade han 2006 tillsammans med sina medarbetare Simon Osindero, Yee Whye Teh och Ruslan Salakhutdinov en metod för att *förträna* ett nätverk med hjälp av en serie Boltzmannmaskiner i lager på lager ovanpå varandra. Förträningen gav kopplingarna i nätverket ett bättre utgångsläge, vilket gjorde träningen för att känna igen drag i bilder effektivare.

Boltzmannmaskinen används ofta som en del i större nätverk. Till exempel kan den användas för att rekommendera filmer och tv-serier baserat på tittares personliga smak.

Maskininlärning i dag och i framtiden

Med sina arbeten från 1980-talet och framåt har John Hopfield och Geoffrey Hinton bidragit till grunden för den revolution inom maskininlärning som började omkring 2010.

Den utveckling vi ser i dag har blivit möjlig på grund av tillgången till stora mängder data som kan användas för att träna nätverken, och genom den enorma ökningen av datorkraft. Dagens artificiella neuronnät är ofta mycket stora och uppbyggda i många lager. Sådana nätverk kallas för djupa neuronnät och träningen av dem kallas djupinlärning.

En blick på John Hopfields artikel om associativt minne från 1982 ger perspektiv på utvecklingen. Han använde där ett nätverk med 30 noder. Om alla noder är hopkopplade med alla de övriga blir det 435 kopplingar. Noderna har sina värden, kopplingarna olika styrka, och totalt blir det under 500 parametrar att hålla reda på. Han testade också ett nätverk med 100 noder, men det blev för otympligt på den dator han hade tillgång till. Det kan vi jämföra med de största språkmodellerna i dag, som är uppbyggda som nätverk som kan innehålla över en biljon parametrar (en miljon miljoner).

Många forskare sysslar i dag med att utveckla olika användningsområden för maskininlärning. Det återstår fortfarande att se vilka tillämpningar som blir mest användbara. Samtidigt pågår en omfattande diskussion om etiska frågor kring hur tekniken utvecklas och används.

Eftersom fysiken har bidragit med verktyg för att utveckla maskininlärning är det intressant att se att fysiken som forskningsfält också drar nytta av artificiella neuronnät. Maskininlärning har länge använts inom områden som kanske är bekanta från tidigare Nobelpris i fysik. Bland annat användes maskininlärning för att sålla igenom och behandla stora datamängder i sökandet efter Higgspartikeln. Andra tillämpningar går ut på att reducera bruset i mätningar av gravitationsvågor från kolliderande svarta hål, eller i sökandet efter exoplaneter, med mera.

På senare år har tekniken också kommit att användas för att beräkna och förutse egenskaper hos molekyler och material – till exempel att beräkna proteinmolekylers struktur, som avgör deras funktion, eller att räkna ut vilka nya versioner av ett material som kan ha bäst egenskaper för att användas till effektivare solceller.

LÄS MER

Mer information om årets priser, bland annat en vetenskaplig bakgrundsartikel på engelska, finns på Kungl. Vetenskapsakademiens webbplats, *www.kva.se*, och på *www.nobelprize.org*. Där kan man också titta på presskonferenser, Nobelföreläsningar och annat videomaterial. Mer information om utställningar och aktiviteter kring Nobelprisen och Ekonomipriset finns på *www.nobelprizemuseum.se*.

Kungl. Vetenskapsakademien har beslutat utdela Nobelpriset i fysik 2024 till

JOHN J. HOPFIELD

Född 1933 (91 år) i Chicago, IL, USA. Fil.dr 1958 vid Cornell University, Ithaca, NY, USA. Professor vid Princeton University, NJ, USA.

GEOFFREY HINTON

Född 1947 (76 år) i London, Storbritannien. Fil.dr 1978 vid The University of Edinburgh, Storbritannien. Professor vid University of Toronto, Kanada.

”för grundläggande upptäckter och uppfinningar som möjliggör maskininlärning med artificiella neuronnätverk”

Vetenskapsredaktörer: Ulf Danielsson, Olle Eriksson, Anders Irbäck och Ellen Moons, Nobelkommittén för fysik
Text: Anna Davour
Illustrationer: Johan Jarnestad
Redaktör: Sara Gustavsson