Neuronale Netze
Version 9.0

(N.1) Was sind neuronale Netze ? *[01]*

(N.1.1) Neurobiologie als Vorbild [01]
(N.1.2) Allgemeine Begriffsbestimmungen [03]
(N.1.3) Erstes Beispiel: ein Musternetz [04]
(N.1.4) Zweites Beispiel: ein logisches Netz [05]
(N.1.5) Verschiedene Netzwerk-Architekturen [06]
(N.1.6) Neuronale Netze und Computer [08]

(N.2) Wie lernen neuronale Netze ? *[09]*

(N.2.1) Vektorrechnung im Gehirn [09]
(N.2.2) Lernende Synapsen [10]
(N.2.3) Abstraktion und Typenbildung [12]
(N.2.4) Folgerungen für die Lernpraxis [13]
(N.2.5) Selbstorganisierende Netzwerke [14]

(N.3) Netzwerke und Sprache *[18]*

(N.3.1) Hierarchische Sprachnetze [18]
(N.3.2) Die richtige Lautgebung [19]
(N.3.3) Das Lernen von Grammatik-Regeln [20]
(N.3.4) Das Verstehen von Sprache [21]
(N.3.5) Semantische Netzwerke [22]
(N.3.6) Emotionen im Netz [23]

(N.4) Schlussbetrachtungen *[24]*

Quellennachweis: Das vorliegende Skriptum verwendet u.a. einige Teile aus dem großartigen Buch „Geist im Netz“ von Manfred Spitzer, erschienen 1996 im Verlag Spektrum.

(N.1) Was sind neuronale Netze ?

(N.1.1) Neurobiologie als Vorbild

Neuronale Netze können als technische Umsetzung der Gehirnfunktionen verstanden werden. Daher sollen zunächst die Nervenzellen und ihr Zusammenwirken im Zentralnervensystem untersucht werden. Eine Nervenzelle (auch Ganglienzelle oder Neuron genannt) besteht wie jede lebende Zelle aus einem Zellkörper, der von seiner Umgebung durch eine Zellmembran abgegrenzt ist. Ihr Aufbau weicht jedoch in auffälliger Weise von dem aller anderen Zellen des Körpers ab. Der Zellkörper trägt eine große Anzahl kurzer, stark verzweigter Auswüchse, die als Dendriten bezeichnet werden. Ein weiterer Auswuchs, das Axon, kann sehr lang werden (bis ca. 1 m) und teilt sich an seinem Ende in viele Zweige auf, die jeweils durch eine Synapse abgeschlossen sind. Jede Synapse ist, getrennt durch den dünnen synaptischen Spalt, mit dem Zellkörper oder einem Dendriten einer anderen Nervenzelle verbunden. Das hat zur Folge, dass jede Nervenzelle mit Tausenden von Synapsen bedeckt ist, wodurch eine hochgradige Vernetzung entsteht.

Die synaptischen Verbindungen ändern sich im Lauf der Zeit. Synapsen können wachsen, verkümmern oder sogar ganz verschwinden. Umgekehrt kann ein Axon neue Zweige mit den zugehörigen Synapsen ausbilden und dadurch mit weiteren Nervenzellen in Kontakt treten. Das führt zu Änderungen im Verhalten des Nervensystems, also zu Lernvorgängen.

Die Funktionsweise einer einzelnen Nervenzelle ist weitgehend bekannt. Sie beruht auf der elektrischen Spannung der Zellmembran gegenüber der Umgebung, dem so genannten Membranpotential. Im Ruhezustand hat die Nervenzelle ein einheitliches Membranpotential von -70 mV. Solange die Zelle nicht von außen angeregt wird, verharrt sie in diesem Zustand. Wird ihr Membranpotential an mindestens einer Stelle über einen bestimmten Schwellwert hinaus erhöht (auf -60 mV), so kommt durch innere Vorgänge (Ionenverschiebungen) eine Kettenreaktion in Gang, die zu einer schlagartigen Erhöhung des Membranpotentials auf etwa +30 mV führt. Durch Ausgleichsvorgänge kehrt die Zelle innerhalb einiger Millisekunden wieder in ihren Ruhezustand zurück.

Diese kurze Spannungsspitze, die den Erregungszustand der Nervenzelle kennzeichnet, nennt man Aktionspotential: die Zelle feuert. Das Aktionspotential pflanzt sich längs des Axons fort und erreicht schließlich die Synapsen. Diese werden zur Ausschüttung bestimmter chemischer Substanzen, der Neurotransmitter, veranlasst, die über die synaptische Spalte zu anderen Nervenzellen gelangen und deren Membranpotential erhöhen. Wenn an einer Nervenzelle genügend viele Synapsen gleichzeitig aktiviert sind, dann wird hier ebenfalls ein Aktionspotential ausgelöst.

Neben den Synapsen, die das Membranpotential erhöhen, also erregend wirken, gibt es auch hemmende Synapsen, die das Membranpotential erniedrigen und so der Aktivierung einer Nervenzelle entgegenarbeiten. Diese Hemmungsmöglichkeit ist für die Stabilität des Nervensystems wesentlich.

Damit ein Organismus lebensfähig ist, muss sein Nervensystem mit seiner Umgebung in Verbindung treten. Eindrücke von der Außenwelt werden durch Sinneszellen aufgenommen und über afferente Axone an das Nervensystem weitergeleitet. Umgekehrt regen efferente Axone Muskelzellen zur Kontraktion an, wodurch es zu Handlungen in der Außenwelt kommt.

Die Aktionspotentiale einer Nervenzelle sehen alle gleich aus (das bekannte "Alles-oder-Nichts-Gesetz"). Die Informationen, die im Nervensystem verarbeitet werden, sind also nicht durch die Größe der Spannungen, sondern durch die zeitlichen Abstände der Aktionspotentiale codiert. Das Nervensystem arbeitet demnach mit Frequenzmodulation.

Zum Begriff des neuronalen Netzes gelangt man, wenn man die beschriebenen Vorgänge zu einem stark vereinfachten abstrakten Modell zusammenfasst. Dieses Modell wird durch folgende Regeln beschrieben:

[1] *Aufbau einer Nervenzelle*: Eine Nervenzelle besitzt viele Eingänge, nämlich die synaptischen Verbindungen, und einen Ausgang, nämlich das Axon.

[2] *Zustände einer Nervenzelle*: Eine Nervenzelle kann zwei Zustände annehmen: den Ruhezustand und den Erregungszustand.

[3] *Verbindungen der Nervenzellen untereinander*: der Ausgang einer Nervenzelle führt zu den Eingängen anderer Nervenzellen.

[4] *Eingänge des Nervensystems*: Einige Nervenzellen (z.B. Sinneszellen) können durch Umweltreize erregt werden.

[5] *Ausgänge des Nervensystems*: Einige Ausgänge von Nervenzellen wirken (über Muskelzellen) auf die Umwelt ein.

[6] *Erregungsbedingung*: Eine Nervenzelle geht in den Erregungszustand über, wenn genügend viele ihrer Eingänge mit erregten Nervenzellen verbunden sind (Aktivierungsschwelle).

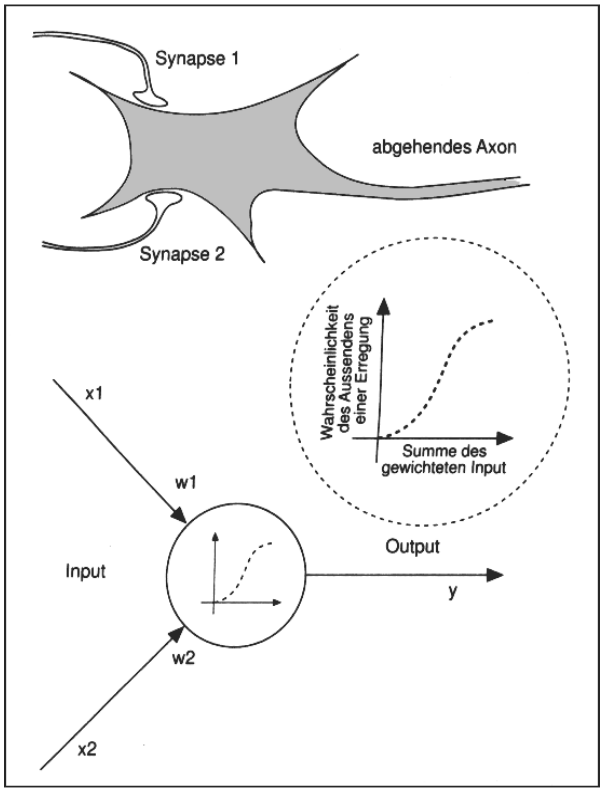
[7] *Unabhängigkeit der Nervenzellen*: Der Zustand einer Nervenzelle ist allein durch die Verhältnisse an ihren Eingängen bestimmt. Die einzelnen Zellen arbeiten also unabhängig voneinander.

Ersetzt man in diesem Modell das Wort "Nervenzelle" durch "Verarbeitungselement", so erinnert nichts mehr an ein biologisches Nervensystem. Nun kann man dem Modell eine technisch anmutende Interpretation geben und erhält so den Begriff des neuronalen Netzes. Ein solches Netz besteht aus einfachen Verarbeitungselementen, die über das ganze System verteilt sind und unabhängig voneinander ("parallel") arbeiten. Diese Arbeitsweise lässt sich unter dem Begriff der parallel verteilten Informationsverarbeitung zusammenfassen.

(N.1.2) Allgemeine Begriffsbestimmungen

Neuronale Netzwerke sind Modelle dafür, wie sich Neuronen im Gehirn verhalten. Mit entsprechenden Programmen lassen sich verschiedene neuronale Prozesse am Computer simulieren. Zu diesen Prozessen, die grundlegend für die Arbeitsweise unseres Gehirns sind, zählen Mustererkennung, Kategorienbildung (Abstraktion), Lernen und Gedächtnis, Sprache, emotionale Bewertungen und Entscheidungen. Die Anwendungsgebiete neuronaler Netzwerke sind vielfältig: in der naturwissenschaftlichen Forschung, in Technik, in Medizin und Psychologie.

Die Schaltelemente eines neuronalen Netzes sind Neuronen. Im Modell wird von verschiedenen biologischen Gegebenheiten abstrahiert, wie beispielsweise Form, mikroskopische Struktur, Zellphysiologie und Neurochemie. Neuronen werden als einfache Schaltelemente aufgefasst, welche viele Eingangssignale X_i (Inputs) in ein Ausgangssignal Y (Output) umwandeln. Ein vorhandenes Signal wird dabei durch die Zahl 1, kein Signal hingegen durch die Zahl 0 beschrieben. Die Stärke der synaptischen Übertragung eines einlaufenden Signals wird mathematisch durch eine Zahl zwischen -1 und +1 symbolisiert, je nachdem ob an der Synapse eine Hemmung oder Erregung des Neurons stattfindet. Diese Verbindungsstärken werden als Synapsengewichte (W_i) bezeichnet. Dabei bezeichnete der Index i die i-te Synapse von insgesamt n Inputleitungen. Das Neuron multipliziert nun jedes Inputsignal X_i mit dem entsprechenden Synapsengewicht X_i*W_i und addiert alle diese so gewichteten Inputs ($S = \sum X_i*W_i$). Als nächstes wird die Summe der gewichteten Eingangssignale mit einem internen Schwellwert G verglichen. Ist die gewichtete Inputsumme S größer als dieser Schwellwert G, dann feuert das Neuron, d.h. es entsteht an seiner Membran ein Aktionspotential. Eine so erfolgte Aktivierung des Neurons wird an der Outputleitung (Axon) durch die Zahl 1 symbolisiert, andernfalls ist der Output 0, das Neuron bleibt stumm.



Zusammenfassend lässt sich die Arbeitsweise eines einzelnen Neurons durch die Inputstärken X_i, die Synapsengewichte W_i und die Aktivierungsschwelle G beschreiben.

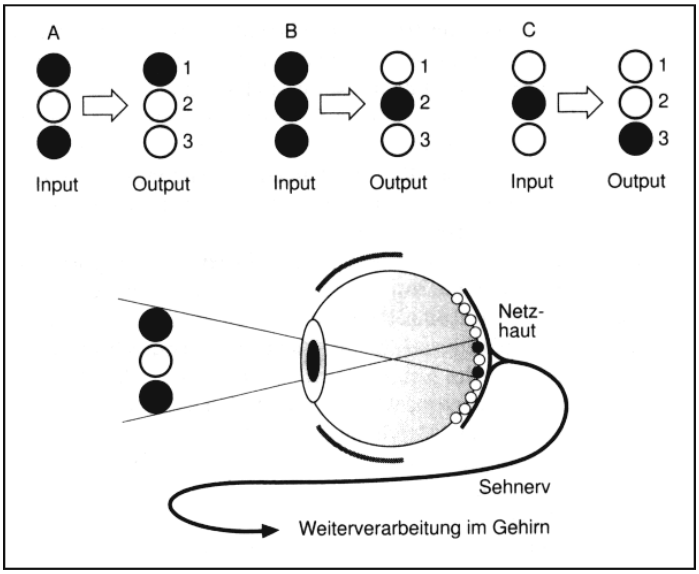
Die nebenstehende Abbildung veranschaulicht die grundlegenden Begriffsbestimmungen.

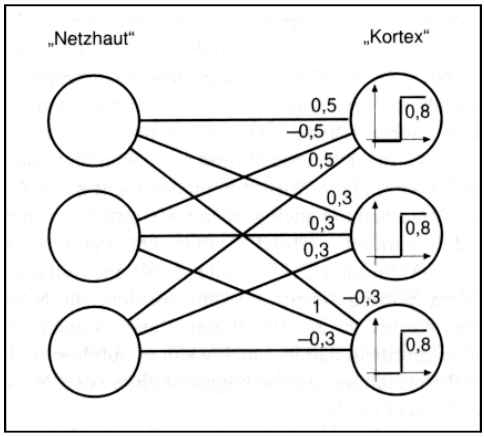
Eine fundamentale Erkenntnis aus der Neurobiologie besagt, dass immer dann, wenn zwei miteinander verbundene Neuronen gleichzeitig aktiv sind, die Verbindung zwischen ihnen stärker wird. Dadurch, dass die beiden Neuronen gleichzeitig feuern, erhöht sich somit die Leitfähigkeit der zwischen ihnen liegenden Synapse, also das Synapsengewicht. Verbindungen inaktiver Neuronen bleiben hingegen immer schwach. Dieser Sachverhalt bildet die Grundlage für jede Art des assoziativen Lernens und wird nach seinem Entdecker als **Hebbsche Lernregel** bezeichnet.

(N.1.3) Erstes Beispiel: ein Musternetz

In neuronalen Netzwerken sind Neuronen durch Signalleitungen miteinander verbunden. Anzahl, Anordnung und Stärke der Verbindungen bestimmen das Leistungsvermögen des Netzes. Um die grundsätzliche Funktionsweise solcher Netzwerke zu erläutern, soll im Folgenden ein einfaches Netz zur Mustererkennung entwickelt werden (es wird fortan kurz als Musternetz bezeichnet). Als Anwendungsbeispiel möge ein Frosch dienen, der an einem Teich sitzt. Sein Gehirn muss in jedem Augenblick die von den Augen kommenden Signale verarbeiten und auf sie angepasst reagieren. Die Neuronen in der Augennetzhaut (Retina) liefern ein Inputmuster von lauter 0 und 1 an die Neuronen des primären visuellen Areals in der Großhirnrinde (Cortex). Die retinalen Neuronen bilden die Inputschicht, die cortikalen die Outputschicht - von einer Zwischenschicht (Interneuronen) wird in dem einfachen Modell abgesehen. Jede der beiden Schichten soll aus jeweils drei Neuronen bestehen und jedes Inputneuron ist mit jedem Outputneuron verbunden. Von einer Vernetzung zwischen den Neuronen in derselben Schicht wird ebenfalls abgesehen.

Dieses Netzwerk soll beispielsweise drei verschiedene Inputmuster erkennen und darauf mit unterschiedlichem Output reagieren. Das Muster **A** ist ein sich nähernder Storch und die Reaktion besteht in einer Kontraktion der Oberschenkelmuskulatur zum rettenden Sprung. Wenn dieser nicht erfolgt, dient der Frosch dem Storch als Nahrung. Das Muster **B** ist eine vorbeischwirrende Fliege und die Reaktion ist die Aktivierung der Zungenmuskulatur zum Herausstrecken der Zunge und Fangen der Fliege. Geschieht dies nicht rasch und präzise, verhungert der Frosch. Das Muster **C** sei der leere blaue Himmel, der es dem Frosch ermöglicht, in aller Ruhe seine Verdauung zu aktivieren.





In unserem Musternetz werden feuernde Neuronen als volle Kreise und stumme Neuronen als leere Kreise dargestellt. Dem Muster **A** entsprechen die feuernden Inputneuronen X_1 und X_3 und das stumme Neuron X_2. Dieses Muster kann durch den Inputvektor (1/0/1) beschrieben werden. Die gewünschte Reaktion auf der Outputseite besteht darin, dass Neuron Y_1 feuert und die Neuronen Y_2 und Y_3 stumm sind, was dem Outputvektor (1/0/0) entspricht. Zu jedem Outputneuron gibt es einen Gewichtsvektor, der die Verbindungsstärken der einzelnen Synapsen darstellt, z.B. (+0.5/-0.5/+0.5) für Y_1. Im vorliegenden Musternetz soll jedes Outputneuron die gleiche Aktivierungsschwelle von G = 0.8 aufweisen. Wird Muster **A** wahrgenommen, so erhalten alle Neuronen der Outputschicht den Inputvektor (1/0/1). Dieser darf jedoch nur das erste Outputneuron Y_1 aktivieren. Das wird dadurch erreicht, dass der erste Input mit +0.5, der zweite mit -0.5 und der dritte mit +0.5 gewichtet wird. Die gesamte gewichtete Inputsumme von Muster **A** auf das erste Outputneuron beträgt dann S = 1*(+0.5) + 0*(-0.5)+1*(+0.5) = 1. Dieser Wert liegt über dem Betrag der Aktivierungsschwelle von 0.8; das Neuron feuert. Bei den anderen Outputneuronen bleibt hingegen die gewichtete Inputsumme von Muster **A** unterhalb der Aktivierungsschwelle.

Bei 3 Inputneuronen gibt es insgesamt 8 verschiedene Inputmuster (0/0/0), (0/0/1), ... (1/1/1). Auf Grund der vorgegebenen Synapsengewichte W_i kann für jedes Outputneuron Y die gewichtete Inputsumme $S = X_1*W_1+X_2*W_2+X_3*W_3 = \sum X_i*W_i$ berechnet werden. Nur wenn dieser Wert größer als 0.8 ist, feuert ein Outputneuron, ansonsten bleibt es stumm. Die nachfolgende Tabelle ordnet diese Zahlen übersichtlich an. In der Tabelle sind die oben besprochenen drei Inputmuster A, B, C und die erzielten Outputaktivierungen dunkel hinterlegt.

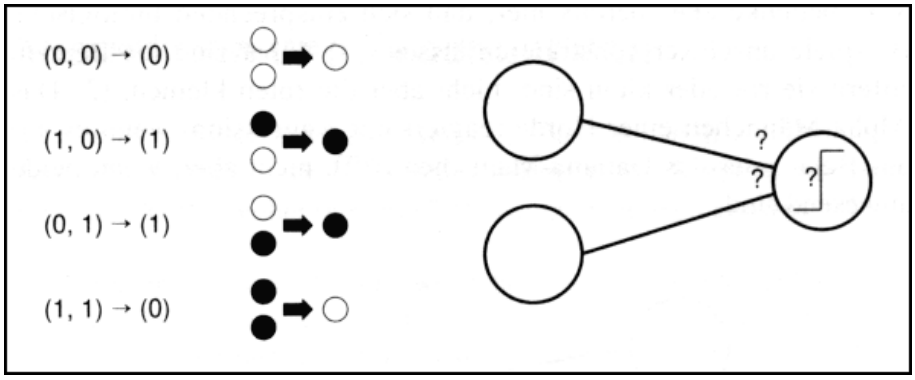
	X1	X2	X3	Y1	Y2	Y3	
	0	0	0	0.0	0.0	0.0	
	0	0	1	0.5	0.3	-03	
Input **C** →	**0**	**1**	**0**	-0.5	0.3	**1.0**	→ Output **C**
	1	0	0	0.5	0.3	-0.3	
	0	1	1	0.0	0.6	0.7	
Input **A** →	**1**	**0**	**1**	**1.0**	0.6	-0.6	→ Output **A**
	1	1	0	0.0	0.6	0.7	
Input **B** →	**1**	**1**	**1**	0.5	**0.9**	0.4	→ Output **B**

Das dargestellte Musternetz reduziert die tatsächlichen biologischen Verhältnisse auf ein Minimum. Aber gerade durch diese Reduktion und Generalisierung im Modell werden die Prinzipien der Mustererkennung in neuronalen Netzwerken deutlich gemacht. Dass eine differenzierte Mustererkennung zum Überleben unerlässlich ist, braucht wohl nicht besonders erwähnt werden.

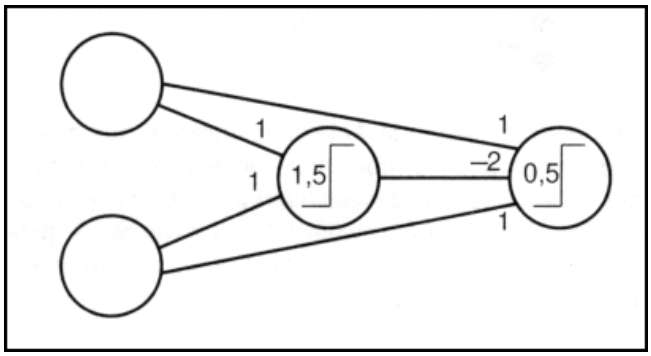
(N.1.4) Zweites Beispiel: ein logisches Netz

Neuronale Netze dienen nicht nur der Mustererkennung, welche durch lineare Vektorkombinationen dargestellt werden, sondern auch der nicht linearen Signalverarbeitung. Eine solche liegt beispielsweise dann vor, wenn zwei Eingangssignale X_1, X_2 entsprechend dem ausschließenden ODER (**XOR**) zu einem Ausgangssignal Y umgeformt werden sollen (Y = entweder X_1 oder X_2). Eine solche Datenverarbeitung findet beispielsweise dann statt, wenn man sich in einem Restaurant entscheiden soll, entweder Marillenknödel oder Sauerkraut zu essen, aber nicht beides zusammen. Ein biologischer Fall für die XOR-Verbindung liegt dann vor, wenn man weiß, dass Waldbeeren essbar sind, sofern sie rot oder klein sind, nicht aber die roten Kleinen. Ein soziologisches Beispiel liegt vor, wenn das Alpha-Männchen einer Horde mit Aggression reagiert, falls es auf das Beta- oder das Gamma-Männchen trifft, nicht aber wenn beide zugleich anwesend sind.

In Analogie zu dem Musternetz sähe ein zweischichtiges Netzwerk der XOR-Verknüpfung wie folgt aus. Im Diagramm stehen dabei jeweils links die vier möglichen Inputmuster (X_1,X_2) und rechts daneben der entsprechende Output (Y).



Bereits 1969 wurde von dem amerikanischen Informatiker Marvin Minsky mathematisch exakt nachgewiesen, dass sich das XOR-Problem durch ein zweischichtiges Netz nicht lösen lässt. Eine Lösung hingegen wird dadurch realisiert, dass zwischen der Input- und Outputschicht noch eine Zwischenschicht (Interneuronen) im Netz eingerichtet wird. In unserem Fall genügt bereits ein einziges Zwischenneuron (Z). Das dreischichtige XOR-Netz hat dabei nachfolgende Gestalt.



Aus dem Schaltplan ist ersichtlich, dass das Zwischenneuron Z nur dann aktiv wird, wenn beide Inputneuronen feuern (S_Z = 1*1+1*1 = 2 und 2 > 1.5). Es entspricht daher der **Konjunktion** „X_1 und X_2“. In diesem Fall hemmt es das Outputneuron (S_Y = 1*1+1*1+1*(-2) = 0 und 0 < 0.5). Das Zwischenneuron in der mittleren Schicht repräsentiert weder direkt den Input noch den Output, sondern eine bestimmte Verallgemeinerung des Inputs, nämlich die Konjunktion. Hier liegt somit der einfachste Fall einer **Abstraktionsleistung**, also einer Kategorienbildung vor. Dadurch wird eine effektivere, weil sparsamere Repräsentation der Außenwelt im neuronalen Netz ermöglicht.

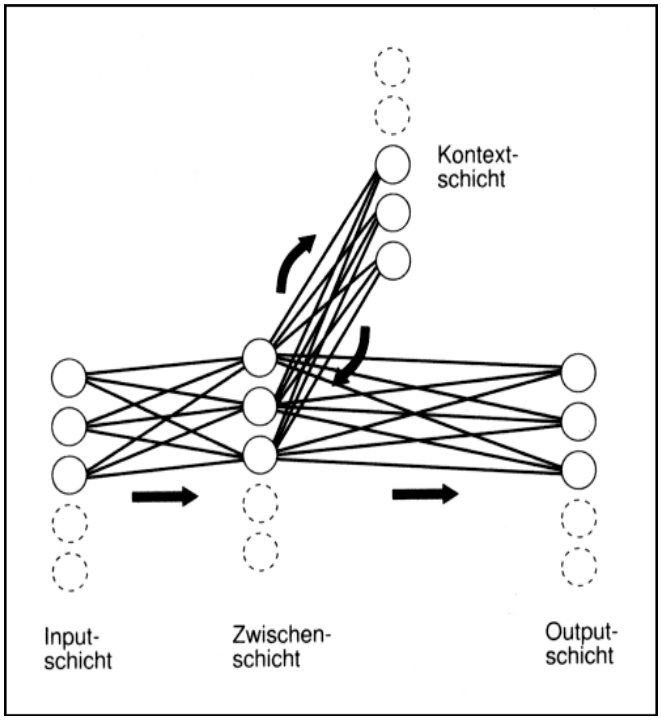
In den letzten Jahren wurden verschiedene Netzwerkarchitekturen entwickelt, welche die Funktion von Zwischenschichten bei der Erzeugung von Abstraktionsleistungen sehr deutlich aufzeigen.

(N.1.5) Verschiedene Neztwerk-Architekturen

In den letzten zwei Jahrzehnten des 20. Jahrhunderts wurden verschiedene Netzwerktypen für unterschiedliche Aufgaben entwickelt. Einige davon sollen im Folgenden kurz vorgestellt werden.

• *Elman-Netzwerke (J. Elman, 1991)*

Ein Elman-Netz ist ein konventionelles dreischichtiges Netzwerk, bei dem die Zwischenschicht mit einer so genannten zusätzlichen Kontextschicht verbunden ist. Die Anzahl der Neuronen in dieser Schicht ist identisch mit der Anzahl der Neuronen in der Zwischenschicht; und es bestehen direkte Leitungen mit der Verbindungsstärke 1 von der Zwischenschicht zur Kontextschicht; die Verbindungsstärken der Kontextschicht zur Zwischenschicht zurück sind hingegen modifizierbar.

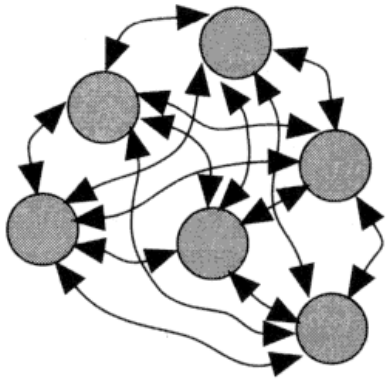


Die Anordnung bewirkt, dass die Kontextschicht als Input den Aktivierungszustand der Zwischenschicht ohne jede Veränderung erhält. Die Kontextschicht funktioniert wie ein Arbeitsspeicher, welcher dem zeitlich nachfolgenden Muster in der Zwischenschicht den vorher aufgenommenen Zustand der Zwischenschicht als zusätzlichen Input anbietet. Die Zwischenschicht erhält somit immer zweierlei Input: den jeweils neuen und den zeitlich vorangehenden. Dadurch wird es möglich, dass manche Muster gleichsam über die Zeit hinweg wirken können.

Mit Elman-Nezten wird die Repräsentation des (zeitbedingten) Kontextes von Information möglich. So wird die Bedeutung des Wortes „*Bank*“ von dem Kontext, in dem sie eingebettet ist, abhängig sein (beispielsweise „*Geld, Zinsen,* .....“ oder „*Park, Tisch,* .....“). Solche Kontexteffekte können nur dann wirksam werden, wenn der relevante Kontext auch tatsächlich im Netzwerk repräsentiert ist. Dies ist bei Elman-Netzen durch die zusätzliche Zwischenschicht der Fall.

• *Hopfield-Netzwerke (J. Hopfield 1986)*

Hopfield-Netzwerke sind Netze, welche aus einer einzigen Schicht bestehen, in der jedes Neuron mit jedem anderen Neuron verbunden ist - nur nicht mit sich selbst. Solche Netze werden auch als assoziative Netze bezeichnet. Außerdem besitzen einige Neuronen afferente (zuleitende) Verbindungen und einige andere Neuronen efferente (wegleitende) Verbindungen. Diese Eingangs- und Ausgangsneuronen stellen den Kontakt zur Außenwelt her. Im Hopfield-Netz soll die Hebbsche Lernregel volle Gültigkeit besitzen, d.h. wenn zwei verbundene Neuronen gleichzeitig aktiv sind, verstärkt sich ihre synaptische Verbindung.



Struktur eines Hopfield-Netzwerks. Jedes Neuron ist mit jedem anderen Neuron verbunden, nicht jedoch mit sich selbst. Das Netzwerk hat zudem (in der Abbildung nicht eingezeichnete) afferente und efferente Verbindungen nach „außen“, über die es Inputmuster erhält und seinen eigenen Aktivitätszustand weitergibt.

Auf Grund ihrer Architektur weisen Hopfield-Netze folgende wesentliche Eigenschaften auf:

[1] Inputmuster werden im Netz als Muster aktiver und inaktiver Neuronen gespeichert. Einmal erregte Netze entwickeln sich selbständig auf einen **stabilen Zustand** hin (**Attraktor**), den man als endgültige Gedächtnisspur des Inputmusters interpretieren kann.

[2] Ein Hopfield-Netz kann auf diese Weise nicht nur ein einziges Eingangsmuster speichern, sondern auch **mehrere verschiedene Muster**. Deren maximale Anzahl beträgt ungefähr 13 % der Anzahl der Neuronen des Netzwerkes.

[3] Der Abruf eines eingespeicherten Inhaltes kann auch bei **teilweiser Darbietung** dieses Inhaltes erfolgen. Erhält das Netz nur einen Teil des Aktivierungsmusters, dann sorgen die verstärkten Verbindungen in ihm dafür, dass das Muster vervollständigt wird (**Vervollständigungsfunktion**).

[4] Auch wenn der Input einem der gespeicherten Muster **nur ähnlich** ist, konvergiert der Aktivitätszustand des Netzwerkes zu diesem hin, d.h. auch ähnliche Inputs können eine eingespeicherte Gedächtnisspur aktivieren (**Verallgemeinerungsfunktion**).

• *Kohonen-Netzwerke (T. Kohonen, 1988)*

Kohonen-Netze sind wie Hopfield-Netze so genannte **selbstorganisierende Netze**. Eine genaue Beschreibung erfolgt in dem nachfolgenden Abschnitt [N.2.5].

• *Netze von Netzwerken*

Ein wichtiges Organisationsprinzip des Gehirns ist die Modularität. Eine Aufgabe wird dann effektiver erledigt, wenn sie in kleinere Teilaufgaben (Module) zerlegt wird.

Eine modulare Netzwerkarchitektur bedeutet, dass bestimmte Untersysteme sich auf bestimmte Funktionen spezialisieren. Dadurch kann die Anzahl der Verbindungsleitungen reduziert und die Geschwindigkeit der Lernvorgänge im Netzwerk beschleunigt werden. Als Beispiel für modulare Verarbeitung seien die verschiedenen visuellen Karten in der Großhirnrinde genannt. Von der Netzhaut werden die Inputsignale an das Feld V1 im primären optischen Cortex übermittelt. Von hier läuft die Information zu weiteren visuellen Arealen. So werden dann im Feld V4 die Farben, im Feld V5 die Bewegungen und im Feld V6 geometrische Formen erkannt.

Im Gehirn wird die Information modular verarbeitet. Von Interesse dabei ist sowohl der Grad der Modularität (bzw. Spezialisierung) als auch die Art und Weise des Zusammenspiels der einzelnen Module (Integration).

(N.1.6) Neuronale Netze und Computer

Neuronale Netzwerke arbeiten anders als herkömmliche Computer. Dieser Unterschied in der Informationsverarbeitung soll im Folgenden kurz erläutert werden.

- Im Computer werden die einzelnen Informationen (d.h. Bitfolgen, also Folgen von 0 und 1) hintereinander verarbeitet. Im Gegensatz zu dieser seriellen Bearbeitung erfolgt im neuronalen Netz die Verarbeitung der Signalmuster parallel.

- Die Geschwindigkeit der Signalverarbeitung an den Schaltelementen (z.B. Transistoren) eines Computers liegt im Bereich von Millionsteln von Sekunden, hingegen bei den Neuronen des Gehirns im Bereich von nur einigen Tausendsteln von Sekunden. Diese langsame Geschwindigkeit der Signalverarbeitung durch ein biologisches Neuron wird andererseits wieder ausgeglichen durch die parallele Informationsverarbeitung von mehreren Neuronen.

- Computer bestehen aus einigen Millionen von Schaltelementen, das Großhirn enthält über 20 Milliarden Neuronen. Ein Schaltelement des Computers hat nur wenige Verbindungsanschlüsse, ein Neuron kann bis zu 10000 Synapsen aufweisen. Die Vernetzungsdichte neuronaler Netze ist ungleich höher als bei einem Computer.

- Die Hardware moderner Computer muss absolut zuverlässig und fehlerlos sein. Ein kleiner Schaden und schon steht der Computer. Computer sind daher fehlerintolerant. Ganz anders hingegen neuronale Netze. Der Ausfall einzelner Neuronen führt nicht zum Totalausfall, sondern bewirkt nur einen langsamen Leistungsabfall. Neuronale Netze sind fehlertolerant.

- Die Hardware eines Computers kann durch die Eingabe von Inputmustern nicht verändert werden, sehr wohl aber die Vernetzungsstärke in neuronalen Netzen. Das Gehirn kann auf Inputveränderungen mit Anpassungen seiner Hardware reagieren (Neuroplastizität).

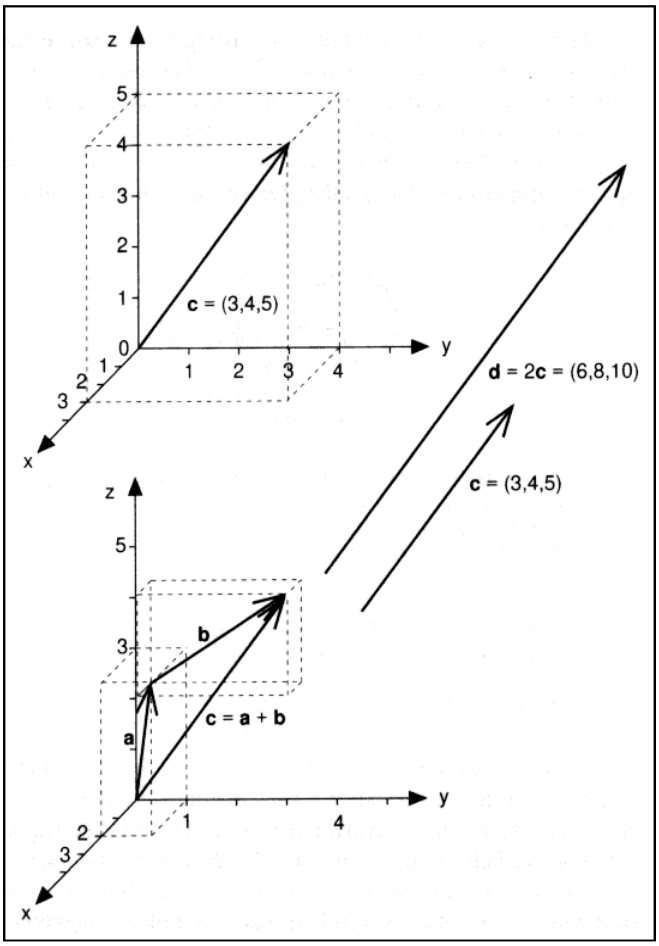
- Im Computer erfolgt eine genaue Unterscheidung von Daten und Adressen. Daten werden an Adressen abgespeichert und von dort wieder aufgerufen. Daten werden von im Zentralprozessor ablaufenden Programmbefehlen bearbeitet. In neuronalen Netzen gibt es diese Trennung von Adressen, Daten und Befehlen nicht. Die gespeicherten Inhalte **sind** die Adressen. Die Daten werden durch veränderte Synapsengewichte an bestimmten Orten im Netz repräsentiert. Datenverarbeitung in neuronalen Netzen erfolgt durch keine einprogrammierten Befehle, sondern durch den Feed-back-Mechanismus in angeleiteten Netzen oder den Mechanismus der lateralen Assoziation in selbstorganisierenden Netzen. Eine genaue Erklärung dazu erfolgt im nächsten Abschnitt.

(N.2) Wie lernen neuronale Netze ?

(N.2.1) Vektoren im Gehirn

In Gehirnen wird die Umwelt repräsentiert. Erst dieses innere Abbild ermöglicht es, auf Umweltreize angemessen zu reagieren. Es wäre jedoch viel zu aufwendig (Energie, Zeit, Material), wenn jeder Punkt der Realität einem Bildpunkt im Gehirn entspräche. Die Reizvielfalt wird daher komprimiert verschlüsselt (Codierung), indem nur wesentliche Merkmale extrahiert und abgespeichert werden. Eine besonders effiziente Form der Repräsentation ist die Codierung der Information in Vektorform. Es gibt heute sehr viele Hinweise dafür, dass Gehirne einfache Vektorrechnung betreiben.

Ein solcher Vektor ist eine gerichtete Größe, die als Pfeil in einem Merkmalsraum dargestellt werden kann. Als Beispiel sollen die Farbreize unseres Auges dienen. Jede Farbe wird als additive Mischfarbe der drei Grundfarben Rot (R), Grün (G) und Blau (B) aufgefasst; d.h. der Merkmalsraum wird durch ein dreidimensionales Koordinatensystem aufgespannt, dessen Achsen den drei Grundfarben entsprechen (R/G/B). Jedem Farbreiz entspricht ein bestimmter Punkt in diesem System bzw. ein Vektor, der vom Systemursprung zu diesem Punkt zeigt (Ortsvektor). Seine Koordinaten geben die Anteile der Grundfarben am Farbreiz wieder. Wenn für die Grundfarben jeweils 256 Intensitätsstufen zur Verfügung stehen (0 - 255), dann sind 256*256*256=16777216 Farbmischungen möglich, also um einiges mehr als unser Auge noch zu unterscheiden vermag. So entspricht der Farbe Purpur der Vektor (250/5/83), d.h. Rot=250, Grün=5 und Blau=83. Schwarz wird durch (0/0/0) und Weiß durch (255/255/255) codiert. Die nachfolgende Abbildung zeigt Beispiele von Vektoren in einem dreidimensionalen Raum.



Ein Reizmuster für unser Gehirn kann somit durch einen bestimmten Merkmalsvektor beschrieben werden.

Der **Länge** $|\mathbf{c}|$ des Vektors **c** entspricht die Reizintensität. Eine **Normierung** des Vektors bedeutet, dass unter Beibehaltung seiner Richtung seine Länge auf den Wert **1** abgeändert wird. Die Reizqualität bleibt gleich, die Reizintensität hingegen wird verändert.

Folgende einfache Operationen mit Vektoren können durchgeführt werden: die **Multiplikation** mit einer Zahl bedeutet eine Streckung oder Stauchung der Vektorlänge unter Beibehaltung der Richtung, also eine Variation der Reizintensität. Die **Addition** zweier Vektoren bedeutet eine Mischung der Reizmerkmale. Werden beide Operationen zusammen ausgeführt, dann spricht man von einer **Linearkombination** der beteiligten Vektoren, z.B. **c = 2a + 3b**.

In neuronalen Netzen spielt die **Ähnlichkeit** von Inputmustern eine wesentliche Rolle. Wie kann nun diese Ähnlichkeit von Mustervektoren im Merkmalsraum quantitativ gemessen werden ?

Normiert man zwei verschiedene Vektoren derart, dass ihre Richtungen unverändert bleiben und ihre Längen auf 1 geändert werden, dann sind sie offenkundig umso ähnlicher, je kleiner der Winkel zwischen ihren Richtungen ist. In der Vektorrechnung kann dieser Richtungsunterschied durch das so genannte Skalarprodukt ausgedrückt werden. Berechnet wird das Skalarprodukt (a*b), indem man die entsprechenden Koordinaten der beiden Vektoren a und b miteinander multipliziert und diese Produkte dann addiert. Wenn a = ($a_1/a_2/$ ..... $/a_N$) und b = ($b_1/b_2/$ ..... $/b_N$), dann gilt (a*b) = $a_1b_1+a_2b_2+.....a_Nb_N = \sum a_i*b_i$. Je größer das Skalarprodukt (a*b) der Vektoren ist, umso größer ist der Cosinus des von den Vektoren eingeschlossenen Winkels α und umso kleiner ist der Winkel selbst: $\cos \alpha = (a*b) / (|a| * |b|)$. Dieses Skalarprodukt erweist sich somit als ein gutes Maß für die Ähnlichkeit zweier Vektoren.

Die Gewichtung eines Inputmusters mithilfe der Synapsenstärken im neuronalen Netz entspricht mathematisch dem Skalarprodukt von Inputvektor und Synapsengewichtsvektor:

$$(X_1/X_2/ ..... /X_N) * (W_1/W_2/ ..... /W_N) = X_1W_1+X_2W_2+.....X_NW_N = \sum(X_i*W_i) = S$$

Auf Grund obiger Überlegungen gilt daher, dass je ähnlicher der Synapsengewichtsvektor dem Inputvektor ist, umso größer ist die gewichtete Inputsumme S und umso größer wird die Aktivierung des Neurons sein. Die Menge jener Inputmuster im Merkmalsraum, die auf Grund ihrer Passung zum Synapsengewichtsvektor zu einer Aktivierung des Neurons führen, bezeichnet man als das **rezeptive Feld** des Neurons.

Lernen im neuronalen Netz bedeutet immer eine Änderung von Synapsenstärken. Prinzipiell kann man beim Lernen in neuronalen Netzwerken zwei Grundtypen unterscheiden:

[1] Überwachtes oder angeleitetes Lernen *(d.h. mit einem äußeren Trainer)*

[2] Unüberwachtes oder selbstorganisierendes Lernen *(d.h. ohne einen äußeren Trainer)*

Beim **angeleiteten Lernen** werden die Synapsenstärken schrittweise so geändert, dass der tatsächliche Output einem gewünschten Output angeglichen wird. Wie das erreicht wird, beschreibt der nächste Abschnitt [N.2.2]. In **selbstorganisierenden Systemen** hingegen werden die Synapsenstärken den einlangenden Inputmustern angepasst, was im n-dimensionalen Merkmalsraum einer Rotation des Synapsengewichtsvektors in Richtung des Inputvektors entspricht. Wie das erreicht wird, beschreibt der Abschnitt [N.2.5]. Auf jeden Fall betreiben neuronale Netze andauernd und sehr effektiv praktische Vektorrechnung.

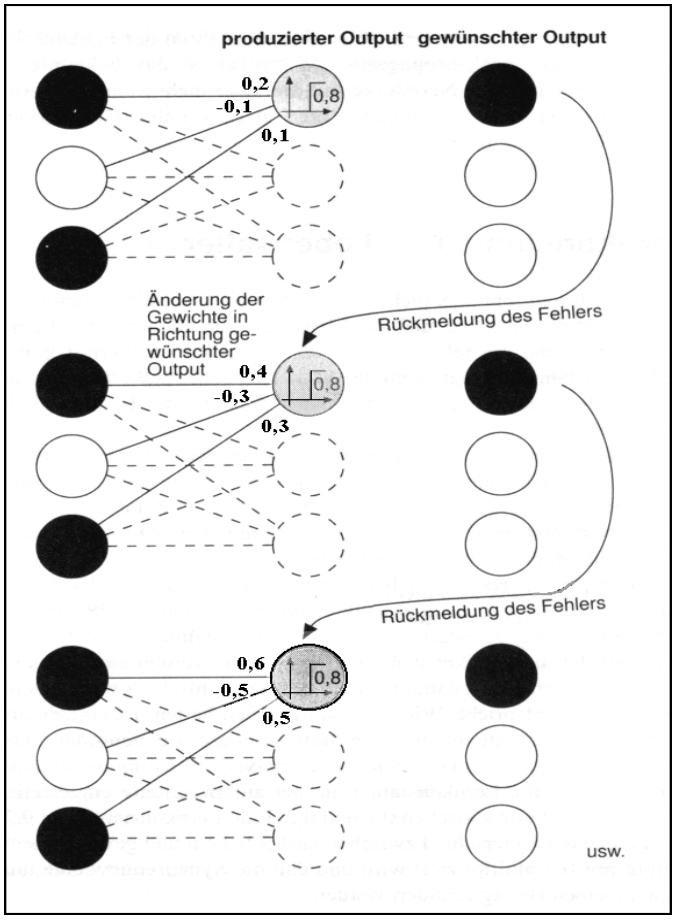
(N.2.2) Lernende Synapsen

Neuronale Netze lernen, indem die Verbindungsstärken der Synapsen verändert werden. Die dahinter stehenden neurobiologischen Mechanismen sind die Langzeitpotenzierung (LTP) und die Langzeitdepression (LTD). Sie bilden die Grundlage für die so genannte **Hebbsche Lernregel**. Durch LTP bzw. LTD wird die Signalübermittlung an der Synapse verstärkt bzw. abgeschwächt. Mathematisch bedeutet das entweder eine Erhöhung oder eine Erniedrigung des Synapsengewichtes.

Wenden wir uns vorerst dem **Lernen in überwachten Systemen** zu. Vor dem Beginn einer Lernphase sind die Synapsengewichte des neuronalen Netzes zufällig in einem kleinen Wertebereich, beispielsweise zwischen -0.2 und +0.2 verteilt. Betrachten wir das Musternetz des vorangehenden Abschnittes. Die Erkennung von Muster **A** besteht darin, dass dem Inputvektor (1/0/1) der Outputvektor (1/0/0) zugeordnet wird. Man erreicht dieses Ziel durch entsprechende Synapsengewichtung, beispielsweise am ersten Outputneuron durch den geeigneten Gewichtsvektor (+0.6/-0.5/+0.5). Die gewichtete Inputsumme 1*0.6+0*(-0.5)+1*0.5 ergibt 1.1 und ist somit höher als die Aktivierungsschwelle von 0.8. Das Outputneuron feuert.

Nehmen wir nun an, dass vor der Lernphase die Synapsengewichte des ersten Outputneurons die Werte (+0.2/-0.1/+0.1) haben. Die damit gewichtete Inputsumme ist 1*0.2+0*(-0.1)+1*0.1 = 0.3, was noch unter der Aktivierungsschwelle des Neurons liegt. Der Wert der Inputsumme 0.3 weicht von der Aktivierungsschwelle von 0.8 um einen Fehler von 0.5 ab. Nach einer **Rückmeldung** (Feed-back) dieses Fehlers an das neuronale Netz erfolgt eine Neueinstellung der Synapsengewichte in einer solchen Weise, dass der Fehler (DELTA) kleiner wird. Je nach Richtung (Vorzeichen) des Fehlers wird zu den alten Gewichten ein kleiner Korrekturwert (z.B. d = ±0.2) addiert. Der erste Lernschritt ergibt dann einen neuen Gewichtsvektor von (+0.4/-0.3/+0.3), was zu einer gewichteten Inputsumme von 1*0.4+0*(-0.3)+1*0.3 = 0.7 führt, was aber noch immer unter der Aktivierungsschwelle von 0.8 liegt. Der Fehler beträgt jetzt nur mehr 0.1. Durch neuerliche Fehlerrückmeldung wird nach demselben Verfahren im nächsten Lernschritt ein geänderter Gewichtsvektor von (+0.6/-0.5/+0.5) eingestellt, was eine gewichtete Inputsumme von 1*0.6+0*(-0.5)+1*0.5 = 1.1 ergibt. Diese liegt bereits über 0.8. Das Outputneuron feuert. **Anmerkung**: In der Praxis wird der Korrekturwert nicht konstant gehalten, sondern er ist direkt proportional zum jeweiligen Fehler (d = k*DELTA); so werden drastische, überschießende Änderungen der Synapsengewichte und damit resultierende Instabilitäten des Systems vermieden.

Dieses schrittweise Lernen durch Fehlerrückmeldung und Fehlerminimierung in zweischichtigen Netzen **(DELTA-Regel)** wird in der nachfolgenden Abbildung schematisch dargestellt.



Das Lerntraining von mehrschichtigen Netzen kann nicht nach der oben beschriebenen einfachen **DELTA-Regel** erfolgen, weil auch die Auswirkungen der Zwischenschichten auf die Outputschicht berücksichtigt werden müssen. Änderungen der Synapsengewichte in der Outputschicht allein genügen hier nicht mehr. Zur Lösung dieses Problems wurde ein mathematisches Verfahren entwickelt, welches den Fehler zunächst der Outputschicht zurückmeldet und dort die Synapsengewichte verändert. Dann wird der Fehler weiter zur nächsten Zwischenschicht gemeldet und die Synapsengewichte in dieser Schicht werden verändert. Die Fehlermeldung durchläuft das gesamte Netzwerk so lange rückwärts (backpropagation), bis jene Neuronenschicht erreicht wird, die ihre Signale direkt von der Inputschicht erhält. Dort finden dann die letzten Gewichtsänderungen statt. Dieser so genannte **Backpropagations-Algorithmus** besteht in einer fortschreitenden Rückwärtsmeldung des Fehlers und erfordert komplizierte und umfangreiche mathematische Berechnungen.

Die oben beschriebene einfache DELTA-Regel und auch der allgemeinere Backpropagations-Algorithmus gelten nur bei **angeleitetem Lernen**. Bei **selbstorganisierenden Netzen** hingegen erfolgt ein spontanes Lernen. Dieses wird erstens dadurch ermöglicht, dass die Neuronen in der Outputschicht untereinander verbunden sind und zweitens müssen unbedingt sehr ähnliche Inputmuster sehr häufig eingegeben werden. Eine ausführliche Beschreibung selbstorganisierender Netzwerke erfolgt im Abschnitt [N.2.5].

Bei angeleitetem Lerntraining (überwachtes Lernen, supervised learning) erfolgt im neuronalen Netz die Anpassung der Synapsengewichte in kleinen Schritten. Dazu muss aber ein und dasselbe Inputmuster in der Menge aller eingegebenen Inputmuster wiederholt vorkommen. Nur dadurch ist gewährleistet, dass die Änderungen der Synapsengewichte in der gewünschten Richtung erfolgen. Andernfalls würden die Gewichtsmodifikationen zu radikal ausfallen und bei den nächsten andersartigen Inputmustern hin und her springen. Das Netz würde keine bleibenden Modifikationen (Speicherwirkungen) aufweisen. Es würde andauernd vergessen. Eine **häufige Musterwiederholung**, eine **langsame Lerngeschwindigkeit** und eine dauernde **Fehlerrückmeldung** sind die unerlässlichen Voraussetzungen für ein erfolgreich angeleitetes Lernen in neuronalen Netzen.

Die **Lerngeschwindigkeit** eines neuronalen Netzes ist durch die so genannte **Lernkonstante k** bestimmt, welche immer auf kleine Zahlenwerte zwischen 0 und 1 eingestellt wird. Eine Lernkonstante von $k = 0.5$ bedeutet, dass der Unterschied zwischen tatsächlichem und gewünschtem Output (also die Differenz zwischen der gewichteten Inputsumme S und der Aktivierungsschwelle G, was ja dem Fehler DELTA entspricht) mit 0.5 multipliziert wird und dass die Synapsengewichte um diesen kleinen Betrag geändert werden: $W_i(neu) = W_i(alt) + k*DELTA$. Diese kleinen Änderungen werden solange weitergeführt bis der gewünschte Output endgültig erreicht ist. Der Zahlenwert der Lernkonstanten bestimmt offensichtlich die Lerngeschwindigkeit. Die Alltagserfahrung zeigt, dass diese vom jeweiligen Lebensalter des Individuums abhängig ist. In jüngeren Jahren ist die Lerngeschwindigkeit relativ hoch, während sie mit zunehmendem Lebensalter abnimmt, was biologisch durchaus sinnvoll erscheint.

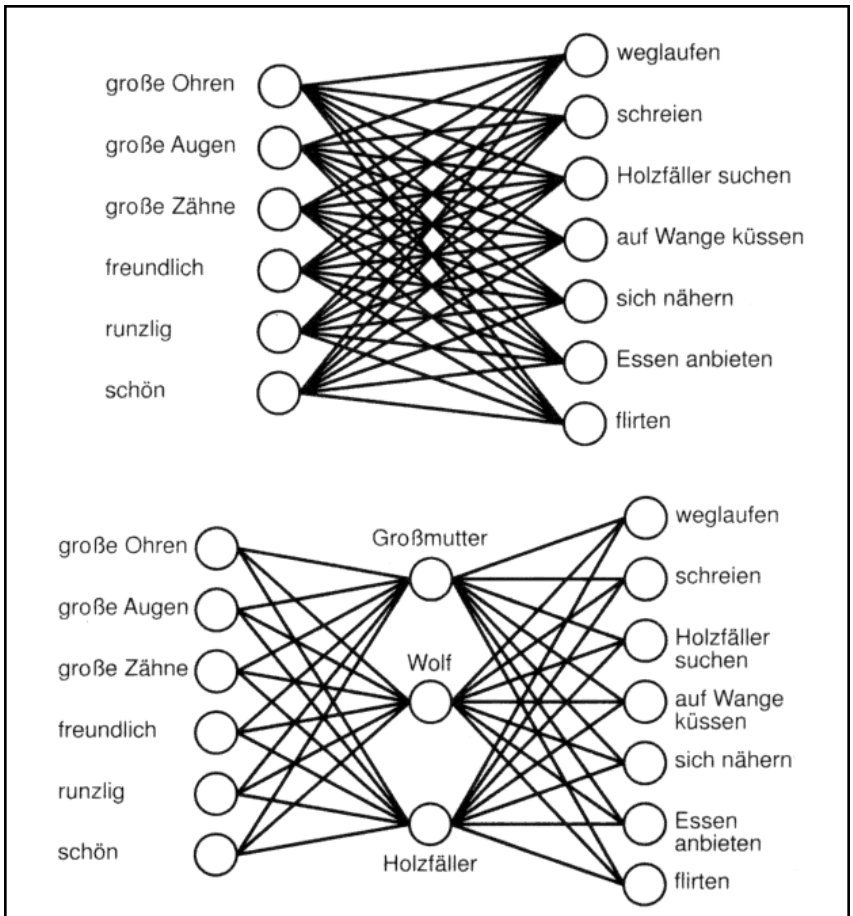
(N.2.3) Abstraktion und Typenbildung

Jones & Hoskins entwickelten 1987 ein neuronales Netz, welches das Verhalten von Rotkäppchen simuliert. Dem bekannten Märchen zufolge läuft Rotkäppchen durch den Wald, wobei ihm drei Lebewesen mit typischen Merkmalen begegnen können. Ein Wolf (große Augen, große Ohren, große Zähne), die Großmutter (große Augen, freundlich, faltig) und ein Holzfäller (große Ohren, freundlich, hübsch). Diesen insgesamt sechs vorgegebenen Reizmerkmalen stehen sieben Verhaltensweisen gegenüber, welche von Rotkäppchen ausgeführt werden können. Weglaufen, Schreien und den Holzfäller suchen (bei einer Begegnung mit dem Wolf); Annähern, Essen anbieten und auf die Wange küssen (bei einer Begegnung mit der Großmutter); Annähern, Essen anbieten und flirten (bei einer Begegnung mit dem Holzfäller).

Ein zweischichtiges Netz kann durch Anwendung der einfachen **DELTA-Regel** trainiert werden, seine Synapsengewichte so zu verändern, dass auf das entsprechende Inputmuster mit dem gewünschten Verhaltensmuster reagiert wird. Das neuronale Netz hat dann das Verhalten von Rotkäppchen gelernt.

Man kann das Verhalten von Rotkäppchen aber auch durch ein dreischichtiges Netz simulieren, wobei man zwischen Input und Output eine Zwischenschicht einschaltet, welche beispielsweise nur aus drei Interneuronen besteht. Durch Anwendung der Backpropagations-Regel wird das Netz trainiert. Dabei stellen die Neuronen der Mittelschicht ihre synaptischen Verbindungen so ein, dass sie die häufig miteinander auftretenden Inputreize (Merkmalsvektoren) zu **Prototypen** bzw. **Kategorien** zusammenfassen. Das bedeutet, dass jeweils ein bestimmtes Zwischenneuron nur dann aktiviert wird, wenn ein Inputmuster aus der Menge der häufig miteinander auftretenden Inputs erscheint. **Die Zwischenneuronen abstrahieren** somit aus der Inputschicht die drei Begriffe Wolf, Großmutter und Holzfäller und aktivieren ihrerseits durch passende Veränderung der Outputsynapsen die gewünschten Outputmuster, d.h. die richtigen Verhaltensweisen.

Die nachstehende Abbildung zeigt eine schematische Darstellung des Rotkäppchen-Netzes, einmal in der zweischichtigen und dann in der dreischichtigen Ausführung.



Oben ist ein zweischichtiges Netzwerk abgebildet, das die „Funktionen" von Rotkäppchen simuliert. Unten ist ein dreischichtiges Netzwerk mit drei Zwischenneuronen abgebildet, das die Funktionen ebenfalls erfüllt.

Das Rotkäppchen-Netz zeigt Folgendes: **Erstens** können in neuronalen Netzen mit Zwischenschichten verallgemeinerte Repräsentationen von Inputmerkmalen gebildet werden (Abstraktion). **Zweitens** entstehen diese Abstraktionen nicht durch vorgegebene Einprogrammierung von Regeln, sondern sie werden schrittweise erzeugt, wenn entsprechend regelhafte Input-Output-Strukturen dem Training zu Grunde liegen. Dieses Ergebnis ist für das Verstehen von Gestalt- und Begriffsbildung von kaum zu unterschätzender Bedeutung.

(N.2.4) Folgerungen für die Lernpraxis

Aus den bisher gewonnenen Erkenntnissen lassen sich wichtige Konsequenzen für die Lernpraxis unserer Kinder ziehen.

• *Kinder lernen spielend*

Wie oben schon mehrfach ausgeführt, basiert das Lernen darauf, dass Input-Output-Beziehungen immer wieder durchgespielt werden und die Synapsenverbindungen im Netzwerk sich langsam so verändern, dass die Wahrscheinlichkeit des richtigen Outputs immer größer wird. Wie kann Lernen aber stattfinden, wenn die Konsequenz eines falschen Outputs zum Tod des Organismus führt ? Das Erlernen solcher lebenswichtiger Verhaltensweisen ist offenbar nur dann möglich, wenn sie ohne Schaden für den Organismus, gleichsam vor dem Ernstfall, immer wieder geübt werden können. Genau das aber ist Sinn und Zweck des Spielens.

• *Kinder lernen an guten Beispielen*

Beispiele (Inputmuster) sind aus der Sicht des Lernerfolges im Netzwerk dann gut, wenn sie sich mit konstanter innerer Strukturierung häufig wiederholen. Kinder brauchen Struktur und Konstanz. Nichts ist schädlicher als chaotischer Input, denn wenn der Input keine **Regelhaftigkeit** aufweist, können keine Muster erkannt und daher kein geordneter Output produziert werden.

• *Kinder lernen durch das Wahrnehmen der Handlungsfolgen*

Nur wenn dem Netzwerk der Fehler eines Outputs rückgemeldet wird, können die Netzwerkverbindungen entsprechend angepasst werden. Unterbleibt ein solches **Feed-back**, kann das Netz nicht modifiziert werden. Das unterstreicht die Wichtigkeit eines systematischen Feed-backs.

• *Praktizieren statt Predigen*

Die eigentliche Arbeitsweise von neuronalen Netzen besteht darin, dass dem Netz keine explizite Regel einprogrammiert wird, sondern das Netz durch regelhafte Inputmuster und durch dauernde Fehlerrückmeldung die Regelhaftigkeit des Inputs selbst erkennt. Darin besteht die Abstraktionsleistung, dass invariante Inputmerkmale durch verstärkte Neuronenverbindungen im Netz repräsentiert werden. **Gehirne sind Regel-Erkennungs-Maschinen**. Wird die zu erkennende Regel vom Trainer bereits vorgesagt, dann unterbleibt die eigentliche Arbeit des Netzwerkes. Predigen und Schimpfen erweisen sich somit als ineffiziente Trainingsmethoden. Viel wirkungsvoller ist dagegen das Praktizieren von guten Beispielen. Das Einzige, was beim Predigen und Schimpfen vom Netz gelernt wird, ist nicht der vorgesagte Inhalt, sondern die Erkenntnis, dass in der Welt irgendwer immer predigt und schimpft. Nicht die wohlgemeinte Botschaft der elterlichen Rüge wird gemerkt, sondern nur dass die Eltern rügen.

• *Kinder brauchen Zeit*

Wie bereits ausführlich erläutert, sichern nur kleine Veränderungen der Synapsengewichte die gewünschte Anpassung des Outputs. Diese aber brauchen Zeit. Nur eine langsame Lerngeschwindigkeit führt zu stabilen Modifikationen. Der Trainer ist daher aufgefordert, **Geduld und Nachsicht** zu üben.

(N.2.5) Selbstorganisierende Netzwerke

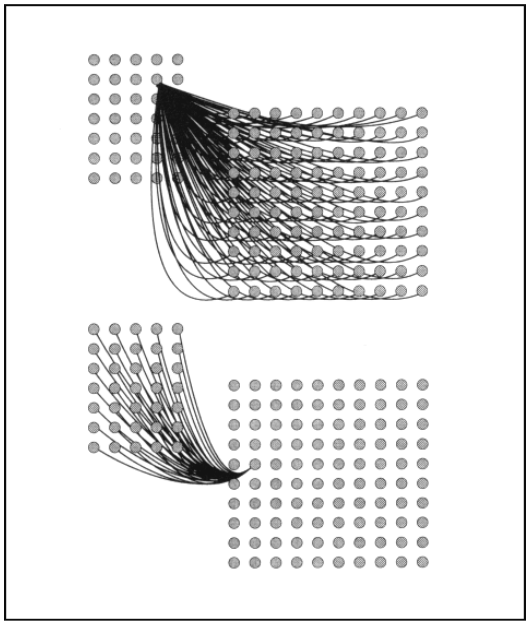
Selbstorganisierende Netzwerke werden nach ihrem Erfinder, dem finnischen Ingenieur Teuvo Kohonen, auch Kohonen-Netze genannt (1988). Sie bestehen im einfachsten Fall aus einer Inputschicht und einer Outputschicht. Diese Schichten sind als zweidimensionale Matrizen konzipiert, und es gelten folgende zwei wesentliche Voraussetzungen:

[1] Jedes Neuron der Inputschicht ist mit jedem Neuron der Outputschicht verbunden. Für die Einstellung der Verbindungsstärken gilt die **Hebbsche Lernregel**. Diese besagt ja, dass häufige gleichzeitige Aktivierung zweier Neuronen deren gegenseitige Verbindung stärkt (what fires together, wires together).

[2] Alle Neuronen der Outputschicht sind miteinander verbunden. Dabei gilt die so genannte **Nachbarschaftsfunktion**: Je näher Neuronen am Erregungszentrum liegen umso stärker werden sie erregt. Je weiter die Neuronen entfernt sind, umso schwächer werden sie erregt bzw. sogar gehemmt.

In der Biologie werden diese Wirkungen durch das Zusammenspiel der Neurotransmitter mit den entsprechenden Membranrezeptoren erreicht.

Die nachfolgende Abbildung zeigt ein einfaches Kohonen-Netzwerk, welches aus einer 5*7-Inputschicht und aus einer 10*10-Outputschicht besteht. Nur die Verbindungen jeweils eines Neurons sind dargestellt. Die Verbindungen der Outputneuronen untereinander sind nicht eingezeichnet. Das Netz enthält also 35*100 = 3500 Input-Output-Leitungen und (100*99)/2 = 4950 Querverbindungen nur in der Outputschicht.



Die oben beschriebene Architektur des Kohonen-Netzes leistet Erstaunliches. Durch wiederholte Eingabe einer begrenzten Anzahl von unterschiedlichen Inputmustern (beispielsweise die 26 Buchstaben des Alphabets) kann jedem dieser Inputmuster ein bestimmter Ort in der Outputschicht zugeordnet werden. Im Fall des Alphabets wird jedem Buchstaben ein bestimmtes Outputneuron zugewiesen. Nach dem Lerntraining entsteht in der Outputschicht eine so genannte **topografische Eigenschaftskarte** der Inputmuster. Dies geschieht aber alles ohne äußere Anleitung (Supervision). Lernen erfolgt hier durch Selbstorganisation des Netzes, welche nur auf wiederholten Reizeingaben und der internen Anwendung von der Hebbschen Lernregel und der lateraler Assoziation beruht. Wie das möglich ist, soll im Folgenden genauer erklärt werden.

Wird beispielsweise an der Inputschicht ein Buchstabe im 5*7-Raster dargestellt, so erhält jedes Neuron der Outputschicht eine Kopie dieses Musters übermittelt, jedoch modifiziert durch die jeweiligen Synapsengewichte. Vor dem Lerntraining sind die Synapsengewichte an jedem Outputneuron unterschiedlich und zufällig verteilt. So können durch den Input verschiedene Outputneuronen aktiviert werden, soferne nur ihre Inputsumme größer als ihre Aktivierungsschwelle ist. Irgendein Outputneuron wird dabei am stärksten aktiviert. Dieses heißt das **gewinnende Neuron**.

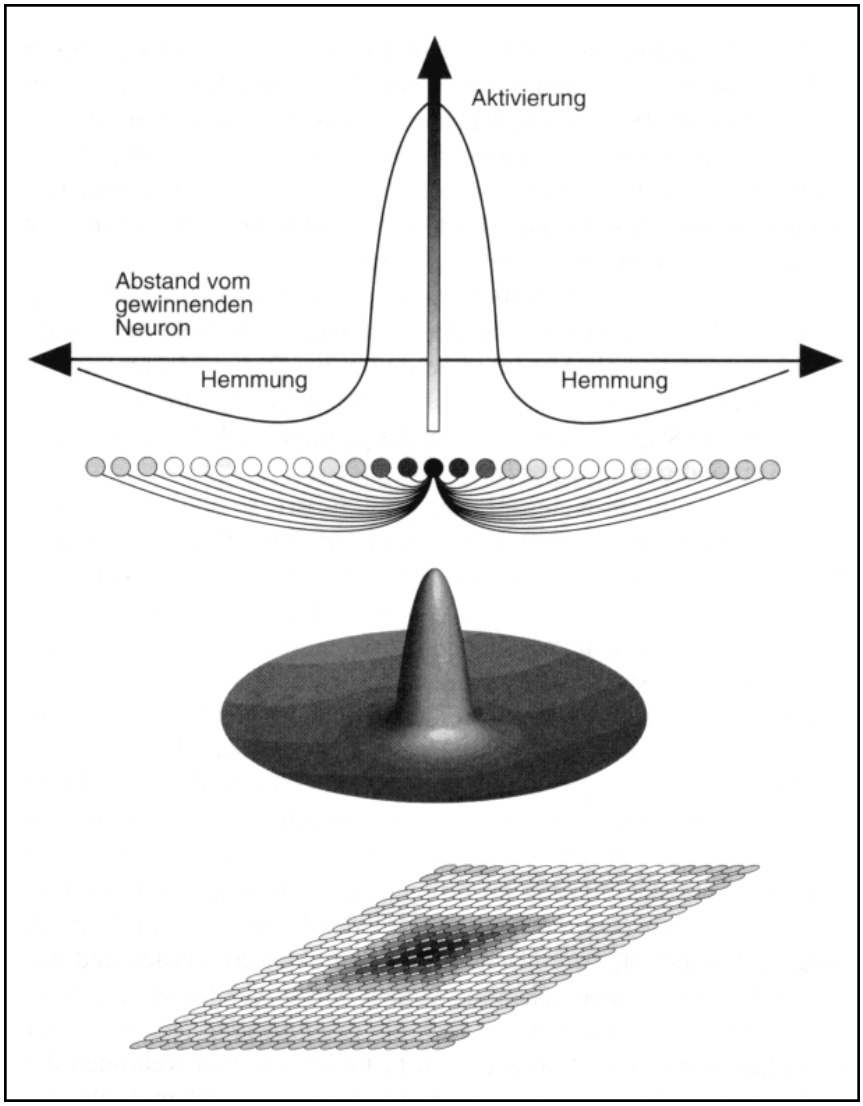
Wovon hängt es ab, welches Neuron gewinnt ? Offensichtlich gewinnt ein Neuron dann, wenn der Inputvektor zu seinem Synapsengewichtsvektor am besten passt. Das ist genau dann der Fall, wenn die mit 1 besetzten Inputleitungen auf Synapsen mit großen bzw. positiven Gewichten treffen, während die mit 0 besetzten Inputleitungen auf Synapsen mit kleinen bzw. negativen Gewichten treffen. In diesem Fall wird die gewichtete Inputsumme möglichst groß, und dementsprechend stark fällt die Aktivierung des Outputneurons aus. In der Sprache der Mathematik lässt sich dieser Sachverhalt sehr einfach ausdrücken: es wird dasjenige Neuron gewinnen, dessen Synapsengewichtsvektor mit dem Inputvektor am besten übereinstimmt, weil dann das skalare Produkt dieser beiden Vektoren und damit die gewichtete Inputsumme am größten ist.

Bis jetzt hat das Kohonen-Netz noch nichts gelernt. Das Lernen vollzieht sich interessanter Weise nur am gewinnenden Neuron und an dessen unmittelbaren Nachbarn. Warum ? Bei einer neuerlichen Darbietung des gegebenen Buchstabenmusters sind die entsprechenden Inputneuronen und das gewinnende Neuron zugleich aktiv. Auf Grund der **Hebbschen Lernregel** kommt es zu einer Verstärkung der Verbindungen zwischen den aktiven Eingangsneuronen und dem gewinnenden Neuron, wohingegen die Verbindungen zwischen den inaktiven Neuronen der Inputschicht und dem gewinnenden Neuron in der Outputschicht unverändert schwach bleiben.

Zusätzlich wird in der Outputschicht die **Nachbarschaftsfunktion** wirksam, so dass Neuronen nahe am Erregungszentrum stärker erregt werden als weiter entfernte. Dadurch kommt es zur Gewinnmaximierung, sodass nur ein einziges Outputneuron - eben das gewinnende - bei einem bestimmten Inputmuster besonders stark aktiviert wird. Das dem gegebenen Inputmuster in solcher Art und Weise zugeordnete gewinnende Neuron hat seine unmittelbare Umgebung mitaktiviert, sodass auch ähnliche Inputmuster mit großer Wahrscheinlichkeit ihr gewinnendes Neuron in dieser Umgebung finden werden. **Je häufiger** ein bestimmter Input oder dazu ähnliche Muster dargeboten werden, **um so ausgedehnter** wird jenes Neuronenareal in der Outputschicht, welches diesen Inputs zugeordnet ist. Dies wird als **Neuroplastizität** bezeichnet.

Die Hemmung aller anderen Neuronen in der Outputschicht führt dazu, dass eine Aktivierung dieser Outputneuronen durch den betreffenden Input oder durch ihm ähnliche Inputmuster praktisch ausgeschlossen ist. Nach einigen Hundert Darbietungen der Inputmuster repräsentieren bestimmte Outputneuronen ganz bestimmte Inputmerkmale. In der Outputschicht ist eine **topografische Eigenschaftskarte** der Inputmuster entstanden.

Die nachfolgende Abbildung zeigt die Outputschicht in der Umgebung eines gewinnenden Neurons (Bildmitte). Die Aktivierungsstärke der Neuronen in der Umgebung verteilt sich dabei wie ein Mexikanerhut um das gewinnende Neuron.



Der Lernvorgang in einem **Kohonen-Netzwerk** lässt sich zusammenfassend folgendermaßen beschreiben.

[1] Ein Eingangssignal wird von den Inputneuronen aufgenommen und als Inputvektor der Outputschicht zugeleitet.

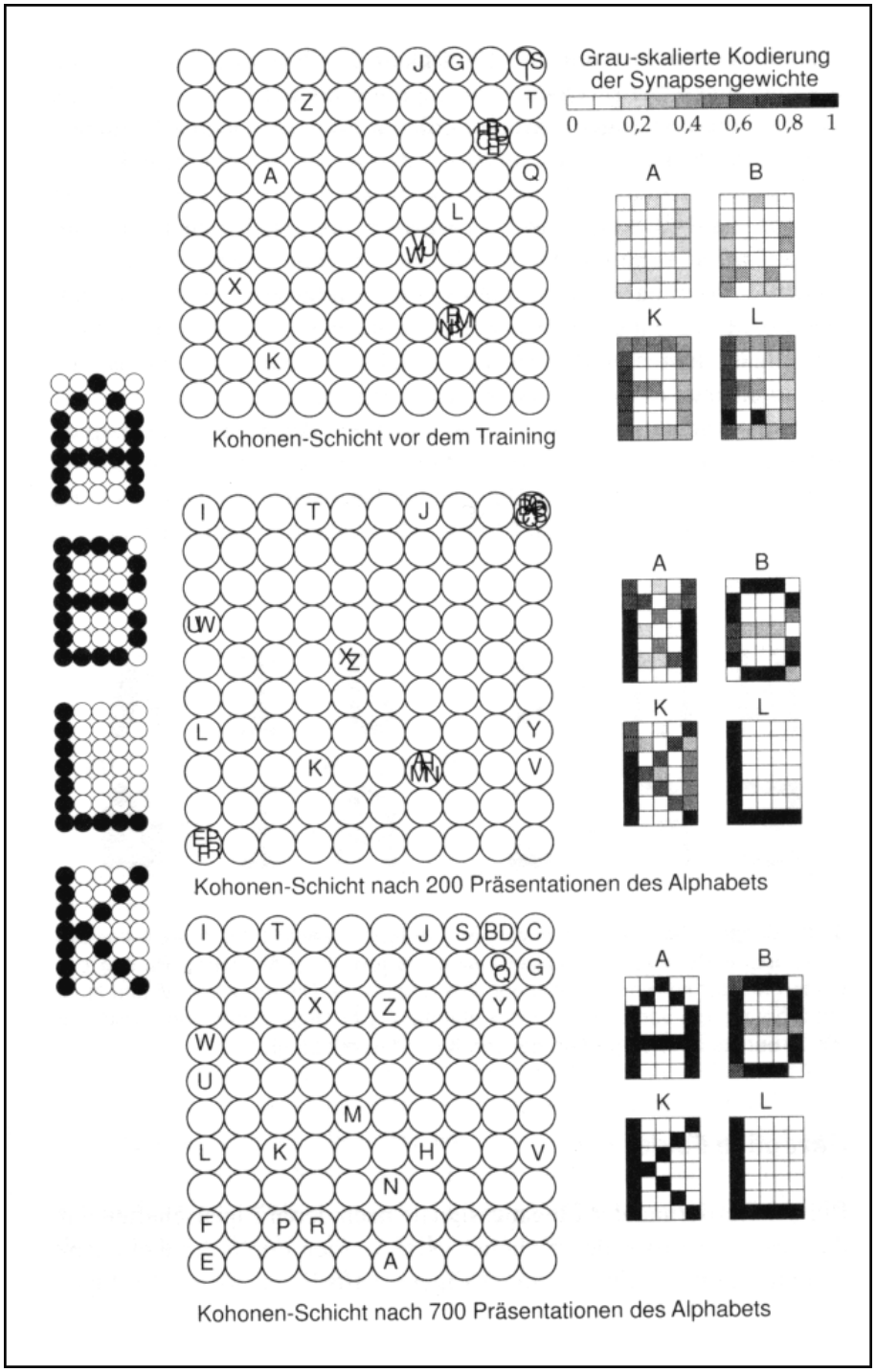
[2] Auf Grund ihrer internen Verschaltung wird in der Outputschicht das Neuron mit dem Erregungsmaximum ermittelt (gewinnendes Neuron).

[3] Beim eigentlichen Lernen wird der Synapsengewichtsvektor des gewinnenden Neurons schrittweise immer mehr dem Inputvektor angenähert. In geringerem Ausmaß geschieht dies auch an den Neuronen in einer unmittelbaren Umgebung des gewinnenden Neurons. Alle anderen Outputneuronen bleiben inaktiv.

[4] Der beschriebene Vorgang wird bei Schritt **[1]** mit einem neuen Inputmuster wiederholt.

Nach vielen Hunderten von solchen Lernschritten wird jedem speziellen Inputmuster ein Neuron bzw. eine Gruppe von Neuronen zugeordnet, welche immer dann aktiv werden, wenn das Inputmuster auftritt. Das funktioniert aber nur, wenn häufig das gleiche oder ihm ähnliche Inputmuster dargeboten werden. Mit der Landkarten-Repräsentation von Inputmustern ist aber wesentlich mehr gemeint als die Abbildung von Inputeigenschaften auf klar definierten Orten der Outputschicht: **Erstens** wird ähnlicher Input auf der Karte nahe beieinander repräsentiert, unähnlicher Input findet sich auf weiter entfernten Orten der Karte. **Zweitens** gilt, dass häufiger Input auf einer größeren Fläche der Karte repräsentiert wird als seltener Input (**Neuroplastizität**).

Die folgende Grafik zeigt die Mustererkennung von Buchstaben in einem Kohonen-Netz. Links sind in einem 5*7-Inputraster vier Buchstaben-Muster dargestellt. In der Mitte sind drei Lernstadien der 10*10-Outputschicht zu sehen. Rechts sind die den Inputvektoren zugeordneten Synapsengewichte für das jeweilig gewinnende Neuron dargestellt. Den Verbindungsstärken entsprechen dabei bestimmte Grauwerte. Die Simulation zeigt insgesamt, dass sich die Synapsengewichte zunehmend besser auf den Input einstellen. Auch ist deutlich zu erkennen, wie sich die Karte mit fortschreitendem Lerntraining nach Ähnlichkeit und Häufigkeit des Inputs formiert.



Selbstorganisierende Netzwerke sind auf Grund der oben ausführlich dargelegten Mechanismen durch zwei wesentliche Leistungen gekennzeichnet: **Erstens** kommt es in der Outputschicht des Netzes zu einer Anordnung, in der bestimmte Neuronen (bzw. Neuronengruppen) generelle Merkmale der vorgesetzten Inputmuster repräsentieren. Das ist deswegen so, weil nach erfolgtem Lerntraining auch ähnliche Inputs die Aktivierung eines Outputneurons auslösen können. Dadurch kommt es auf der neuronalen Eigenschaftskarte zu einer Reduzierung der Dimensionalität des Merkmalsraumes. Aus der Reizvielfalt werden einige wenige Haupteigenschaften extrahiert - das sind diejenigen, in denen die Inputmuster am deutlichsten variieren. Irgendwie erinnert diese **Abstraktionsleistung** neuronaler Netze an das statistische Verfahren der Faktorenanalyse !

Zweitens wird durch die Häufigkeit der Darbietung des gleichen Inputs das ihn repräsentierende Areal auf der Outputschicht vergrößert. Diese **Neuroplastizität** lässt sich bei professionellen Gitarrespielern oder auch bei Blinden, die häufig mit ihren Fingern in der Blindenschrift lesen, nachweisen. Die grafische Darstellung der aktiven sensorischen Areale für die entsprechenden Finger in der hinteren Zentralwindung der Gehirnrinde (Cortex) mithilfe moderner bildgebender Verfahren hat folgende Ergebnisse gezeigt (T. Elbert, 1995): **[1]** Der Cortex baut sich zeitlebens gemäß dem von ihm verarbeiteten Input um. **[2]** Der cortikale Umbau erfolgt besonders ausgeprägt in jüngeren Lebensjahren, d.h. die Lerngeschwindigkeit sinkt mit zunehmendem Alter.

(N.3) Netzwerke und Sprache

Neuronale Netze erkennen Muster, sie sind zu Abstraktionsleistungen fähig. Selbstorganisierende Netzwerke extrahieren Regelhaftigkeiten aus Inputmustern und bilden diese nach Häufigkeit, Ähnlichkeit und Wichtigkeit auf Eigenschaftskarten ab. Zusätzliche Zwischenschichten repräsentieren als Arbeitsgedächtnis den zeitbedingten Kontext eingegebener Information.

Diese Erkenntnisse und Funktionsprinzipien lassen sich in verschiedenen Gebieten der Psychologie anwenden. Die Beispiele reichen von der Regelung der selektiven Aufmerksamkeit und Gedächtnisprozessen über Spracherwerb und Sprachproduktion bis zu Denkstörungen und psychotischen Erkrankungen. In all diesen unterschiedlichen Gebieten wurden in den letzten Jahren erfolgreich neuronale Netzwerke verwendet, um einfache und doch mächtige Erklärmodelle aufzustellen. Aus der Fülle der Forschungsergebnisse sollen im Folgenden exemplarisch einige Untersuchungen aus der Sprachpsychologie herausgegriffen werden.

(N.3.1) Hierarchische Sprachnetze

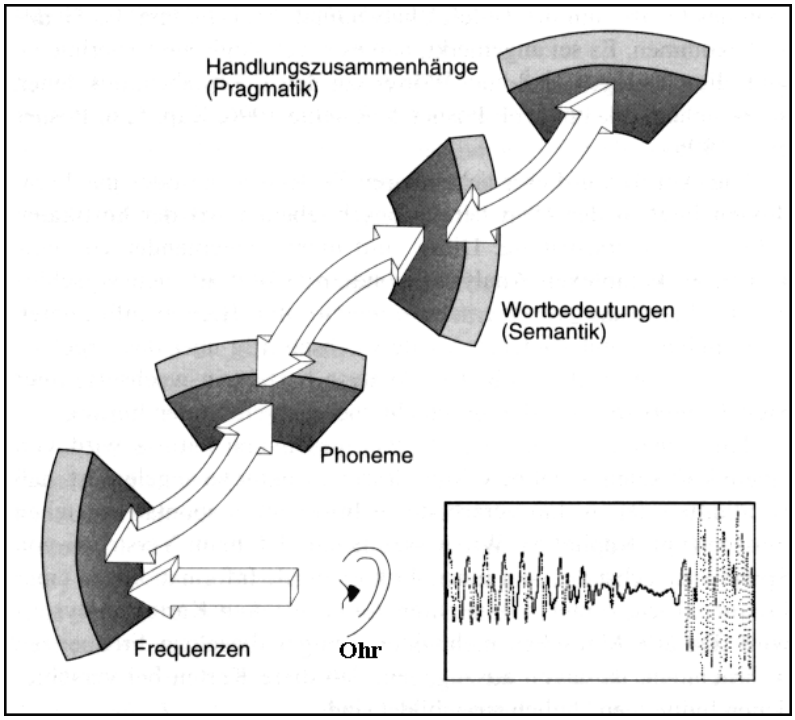
Im Lehrbuch sind die Grundlagen der Sprachpsychologie behandelt worden. Hier nun soll die Verwendung von neuronalen Netzen zur Erklärung menschlichen Sprachverhaltens herangezogen werden. Zunächst zwei alltägliche Phänomene, die jedem bekannt sind. Wenn uns beispielsweise ein Wort auf der Zunge liegt, so wissen wir zwar, was wir sagen wollen, aber es fällt uns das entsprechende Wort dazu nicht ein. Auch in Zuständen von Ermüdung kann es zu solchen Wortfindungsstörungen kommen. Dieser Sachverhalt beweist eindeutig, dass im Gehirn die Repräsentation der Bedeutungen von der Repräsentation der zugeordneten Wortfolgen getrennt ist.

Wörter werden zu Sätzen verknüpft. Die Wörter eines Satzes bilden den Umgebungszusammenhang (Kontext) eines bestimmten Wortes. Die Bedeutung eines einzelnen Wortes kann aus seinem Kontext verstanden werden. Im Satz „*der ?isch zappelt im Netz*“ wird das unvollständige Wort „*?isch*“ unschwer zu „*Fisch*“ und nicht zu „*Tisch*“ ergänzt. Im Satz „*der Mann ist zur Bank gegangen*“ ist die Bedeutung des Wortes „*Bank*“ noch immer nicht eindeutig. Um zu entscheiden, ob es sich dabei um eine Sitzgelegenheit oder ein Geldinstitut handelt, muss ein noch größerer Kontext (d.h. die umgebenden Sätze im Satzgefüge) betrachtet werden.

Diese und ähnliche Beispiele legen nahe, das Sprachverhalten nicht durch ein einziges neuronales Netzwerk sondern durch ein Modell von mehreren parallel arbeitenden Netzen zu beschreiben.

Die einzelnen Stufen der Sprachverarbeitung werden von bestimmten Arealen auf der Großhirnrinde bewältigt, in welchen Landkarten (d.h. ortspezifische Repräsentationen) der verarbeitenden Information spontan entstehen. Auf tiefster Stufe wird die vom Innenohr kommende Information zunächst durch eine **Frequenzkarte** analysiert (tonotopische Karte). Da alle Menschen mehr oder weniger dieselben Frequenzen wahrnehmen, ist es naheliegend, dass diese Karten bei verschiedenen Menschen sehr ähnlich entwickelt sind. Auf der nächsten Stufe der Sprachverarbeitung werden aus den Frequenzspektren **lautliche Einheiten** (Phoneme) gebildet. Diese Laute sind aber für verschiedene Sprachen nicht identisch. Aus einem Grundinventar von ungefähr 90 wohl unterschiedenen Phonemen lernt das Kleinkind nur jene zu gebrauchen, die auch in seiner Sprachwelt verwendet werden (das sind im Englischen ca. 40 Phoneme). So haben die Engländer mit der deutschen „*Gemütlichkeit*“ ihre Schwierigkeiten und die Japaner mit dem deutschen „***R**adfahr**e**r*“. Ein „***ü***“ kommt im Englischen und ein „***r***“ im Japanischen nicht vor, sodass das „***ü***“ als „***u***“ und das „***r***“ als „***l***“ ausgesprochen werden. Was nicht gehört wird, kann nicht unterschieden werden und wird daher auch nicht kodiert und ist damit auch nicht verfügbar.

Über der Schicht der Phoneme folgt die **Schicht ihrer Bedeutungen**. Diese semantische Sprachebene ist ein **assoziatives Netz**, dessen Knoten (d.h. Neuronen bzw. Neuronengruppen) die Bedeutung von Wörtern repräsentieren. Die Assoziationen werden durch die verbindenden Nervenbahnen hergestellt, wobei die Stärke einer Assoziation durch die Verbindungsstärke der Synapse gegeben ist. Über diesem semantischen Netz liegt schließlich noch das Netzwerk der **Handlungszusammenhänge**, das einer letzten **Kontextschicht** entspricht. Der Kontext kann den Sinn eines Satzes determinieren, der Satz die Bedeutung eines Wortes und das Wort den Klang eines Phonems. In diesem hierarchisch-modularen Modell unseres Sprachverhaltens läuft der Informationsfluss nicht nur von unten nach oben (bottom up), sondern auch von oben nach unten (top down). Außerdem erfolgt er nicht seriell, sondern in den einzelnen Modulen parallel. Die nachfolgende Abbildung soll diese Verhältnisse deutlich machen.



(N.3.2) Die richtige Lautgebung

T. Sejnowsky und C. Rosenberg entwickelten ein dreischichtiges Netz (**NETtalk**, 1987), welches geschriebenen englischen Text in gesprochene englische Sprache umwandelt. Das stellt in der englischen Sprache ein erhebliches Problem dar, weil die Zusammenhänge von Schrift und Laut äußerst unregelmäßig sind. Man vergleiche nur den Klang von „***i***“ in den Wörtern „***i**magine*“ und „***i**con*“. So könnte auf Grund des Klanges der Phoneme das Wort „***fish***“ auch als „***ghoti***“ geschrieben werden („***f***“ von „*enou**gh***“, „***i***“ von „*w**o**men*“ und „***sh***“ von „*na**ti**on*“).

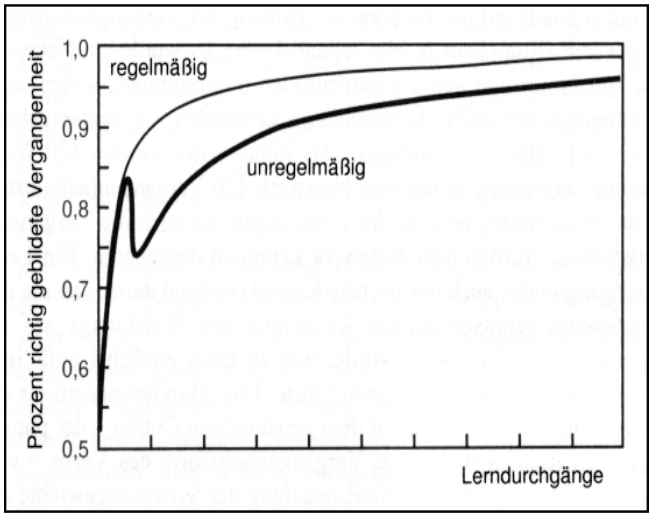
Dem neuronalen Netz **NETtalk** wurden 29 verschiedene Grundmuster eingegeben (26 für alphabetische Buchstaben und 3 für Satzzeichen und Wortgrenzen). Da die Aussprache im Englischen weitgehend von den voranliegenden und nachfolgenden Buchstaben abhängt (Kontext), wird der Inputschicht jeweils eine Kette von 7 Buchstaben gleichzeitig angeboten. Somit besteht die Inputschicht aus 29*7 = 203 Einheiten. Die Zwischenschicht enthält 80 Verarbeitungseinheiten. Die Outputschicht besteht aus 34 Einheiten für die akustischen Merkmale der Phoneme. Die Outputschicht ist zusätzlich mit einem digitalen Sprachsynthesizer verbunden, welcher die Outputsignale in hörbare Laute umwandelt. So wird es möglich, der Maschine beim Lesen eines englischen Textes zuzuhören.

Insgesamt umfasst das Netz mehrere tausend Synapsengewichte, die eingestellt werden müssen. Zu Beginn des Lerntrainings haben diese Gewichte kleine zufällige Werte. Beim Training wird eine relativ geringe Zahl von Wörtern (1024) zehn Mal wiederholt eingegeben und die jeweiligen Abweichungen von den richtigen phonetischen Outputmustern dem System zurückgemeldet (Backpropagations-Algorithmus). Dadurch kommt es zu schrittweisen Veränderungen der Synapsenstärken in Richtung des gewünschten Outputs. In der Zwischenschicht des Netzes bilden sich abstrakte Prototypen wie Vokale, Konsonanten und Silbengrenzen.

Es ist faszinierend der Maschine zuzuhören, wie sie lernt Englisch zu lesen und zu sprechen. Zuerst hört man, wegen der anfänglich zufälligen Verbindungen, nur eine wirre Abfolge von Geräuschen, die in eine Art von kindlichem Lallen übergeht. Nach zehn Stunden Training werden sowohl die gelernten als auch neue unbekannte Wörter mit einer Trefferrate von 95 % phonetisch richtig widergegeben.

(N.3.3) Das Lernen von Grammatik-Regeln

D. Rumelhart und J. McClelland entwickelten 1986 ein zweischichtiges neuronales Netz, welches von englischen Verben die Vergangenheitsform bildet. Das Netzwerk ist mit jeweils 460 Input- und 460 Output-Neuronen programmiert, wobei jedes Inputneuron mit jedem Outputneuron verbunden ist (das ergibt 460*460 = 211 600 Verbindungen). Im klassischen Versuch wurde das Netz mit 420 verschiedenen Verben trainiert, darunter sowohl regelmäßige (z.B. kaufen - kaufte) als auch unregelmäßige (z.B. sitzen - saß). Bei einem Lerndurchgang wurde den Inputneuronen der Wortklang der Gegenwartsform von einem Verb dargeboten, was zu einer vorerst zufälligen Aktivierung der Outputneuronen führte, welche über einen angeschlossenen Sprachsynthesizer ausgegeben wurde. Dieser klangliche Output wurde mit dem richtigen Klangbild der Vergangenheitsform von dem Verb verglichen und der Fehler (DELTA) zur Neueinstellung der Synapsengewichte in den Verbindungsleitungen zwischen Input- und Output-Schicht verwendet. Nach insgesamt 80 000 Durchgängen hatte das Netz die richtigen Zuordnungen gelernt. Es konnte für ein in Gegenwartsform eingegebenes Verb den richtigen phonetischen Code der Vergangenheitsform generieren. Auch dann, wenn neue Verben dargeboten wurden, arbeitete das Netz fast fehlerlos. Bei regelmäßigen Verben lag die Trefferquote um 92%, bei unregelmäßigen Verben um 84%.

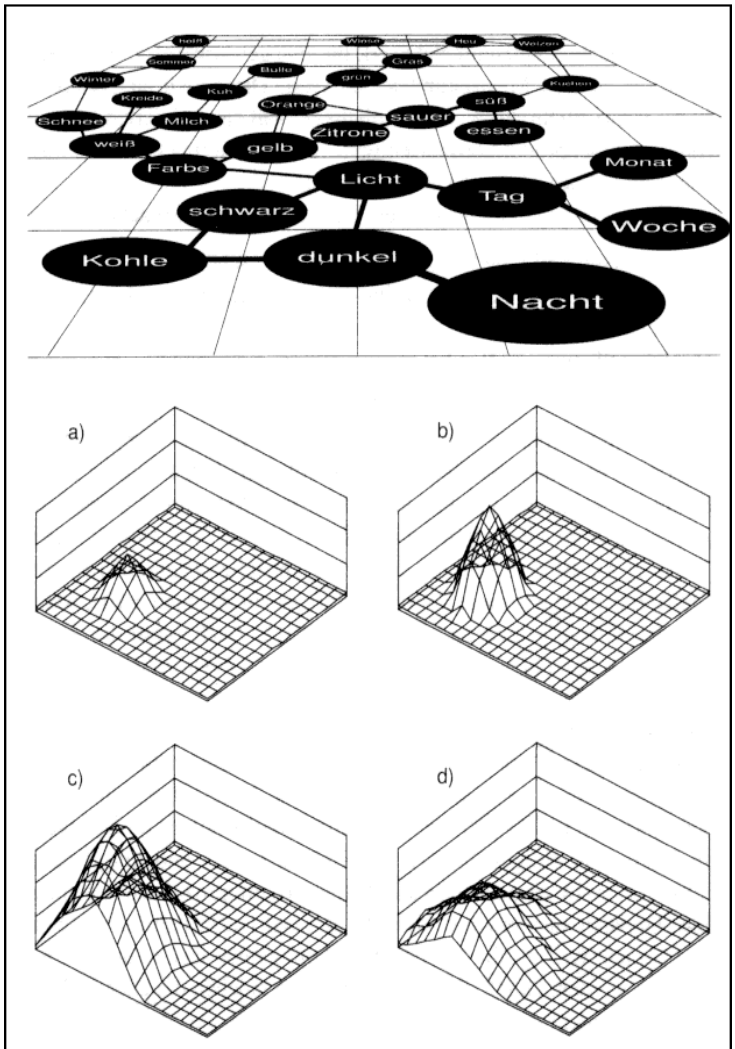


Besonders hervorzuheben ist die Tatsache, dass die Lernkurve des Modells mit der tatsächlichen Lernkurve von Kindern übereinstimmt. Die Vergangenheitsform der regelmäßigen Verben wird in stetiger Weise immer besser produziert, während die Erzeugung der Vergangenheitsform unregelmäßiger Verben zunächst ebenfalls immer besser wird, dann jedoch abfällt und erst später sich wieder verbessert.

Aus den verblüffenden Übereinstimmungen in der Art und Weise und auch im Phasenverlauf des Lernverhaltens von neuronalen Netzwerken und Kindern ergeben sich weitreichende Konsequenzen. Zu keiner Zeit wird dem Netz eine Grammatik-Regel explizit vorgegeben. Es werden nur wohl unterscheidbare Inputmuster wiederholt dargeboten. Häufigkeit und Ähnlichkeit dieser Inputs bewirken dann, dass sich die Verbindungsstärken zwischen einigen Hunderten von Neuronen optimal einstellen. Der einzige Eingriff besteht darin, dass die Abweichungen vom gewünschten Output dem Netzwerk von außen zurückgemeldet werden. Alles andere leistet das Netz selbständig. Diese und ähnliche Experimente zeigen, dass die Annahme eines angeborenen Spracherlernungsmechanismus, der in jedem Gehirn arbeitet, grundsätzlich falsch ist. Grammatik-Regeln sind anfangs nirgendwo im Kopf lokalisierbar, sie müssen anhand guter und klarer Beispiele schrittweise erlernt werden.

(N.3.4) Das Verstehen von Sprache

Das Verständnis eines Wortes wird sehr oft dadurch überprüft, dass seine Wortbedeutung mithilfe von anderen Wörtern erklärt werden muss. Also bestehen Verbindungen (**Assoziationen**) zwischen den einzelnen Wortbedeutungen. Dadurch wird ein semantisches Netz gebildet, dessen Knoten (d.h. Neuronen bzw. Neuronengruppen) die Bedeutung der Wörter repräsentieren. Die Assoziationen werden durch die verbindenden Nervenbahnen hergestellt, wobei die Stärke der Assoziation durch die Verbindungsstärken der entsprechenden Synapsen gegeben ist. Die nachfolgende Abbildung zeigt das Schema eines semantischen Netzwerks.



Semantisches Netzwerk, schematische Darstellung (oben). Zeitlicher Verlauf der Aktivierung eines Knotens im semantischen Netzwerk (unten). Das Ausmaß der Aktivierung ist auf der nach oben gerichteten Achse des Koordinatensystems aufgetragen. Die Aktivierung beginnt binnen 200 bis 300 Millisekunden nach Beginn der Wahrnehmung eines Wortes (a), wird stärker (b) und breitet sich aus (c), um dann bei wahrscheinlich weiterer Ausbreitung wieder abzuflachen (d).

Wenn wir Sprache verstehen oder selbst sprechen, so werden Knoten in semantischen Netzwerken aktiviert. Bei der Aussage *„die Nacht ist dunkel“* werden der *„Nacht“*-Knoten und der *„dunkel“*-Knoten aktiv.

Innerhalb weniger Millisekunden kann sich die Aktivierung von einem Knoten auf dessen nähere Umgebung ausbreiten (**spreading activation**). Durch die Mitaktivierung tauchen assoziierte Wörter in nachfolgenden Sätzen unserer Spontansprache häufiger auf und werden auch in so genannten Wort-Entscheidungs-Aufgaben schneller erkannt. Wer das Wort *„dunkel“* wahrnimmt, kann auf Grund der Mitaktivierung auch das Wort *„Licht“* schneller und besser als andere erkennen.

Je häufiger zwei Wörter zugleich wahrgenommen werden (Kontiguität) und je plausibler ihre logische Beziehung ist (z.B. Teil – Ganzes, oder Ursache - Wirkung), desto so näher liegen ihre Bedeutungsknoten im Netz. Je näher nun zwei Netzwerk-Knoten liegen, um so leichter kann zu einem vorgegebenen Wort das andere Wort reproduziert werden.

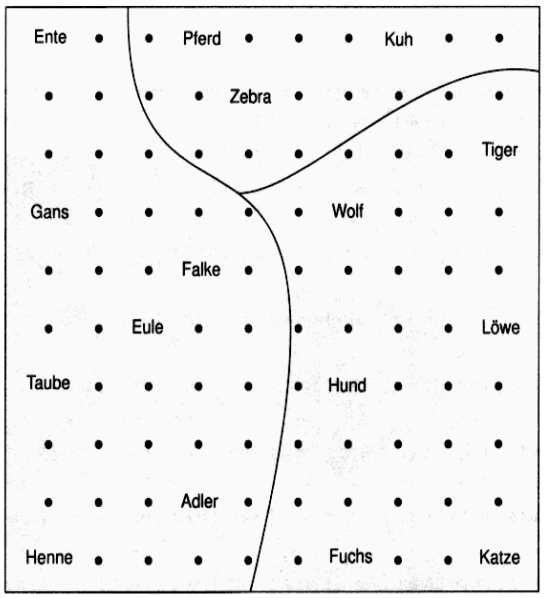
Wie aber kommt es zur Organisation unserer mentalen semantischen Netzwerke ? Diese Frage soll im nächsten Abschnitt behandelt werden.

(N.3.5) Semantische Netzwerke

H. Ritter und T. Kohonen konnten 1989 erstmals zeigen, dass semantische Netzwerke vom Typ der selbstorganisierenden Eigenschaftskarten sind (siehe auch Abschnitt [N.2.5]). Die Bedeutung der Wörter ist durch bestimmte Eigenschaften gekennzeichnet - genauso wie einfache grafische Inputmuster, wie beispielsweise Buchstaben. Ritter und Kohonen verwendeten 16 Tiernamen, die durch einen Merkmalsraum von 13 Eigenschaften charakterisiert sind.

		Taube	Henne	Ente	Gans	Eule	Falke	Adler	Fuchs	Hund	Wolf	Katze	Tiger	Löwe	Pferd	Zebra	Kuh
ist	klein	1	1	1	1	1	1	0	0	0	0	1	0	0	0	0	0
	mittel	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0
	groß	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1
hat	2 Beine	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
	4 Beine	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1
	Haare	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1
	Hufe	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
	Mähne	0	0	0	0	0	0	0	0	0	1	0	0	1	1	1	0
	Federn	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
kann	jagen	0	0	0	0	1	1	1	1	0	1	1	1	1	0	0	0
	rennen	0	0	0	0	0	0	0	0	1	1	0	1	1	1	1	0
	fliegen	1	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0
	schwimmen	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0

So kann beispielsweise eine Katze (bzw. Katzenhaftigkeit) durch den Eigenschaftsvektor (1/0/0/0/1/1/0/0/0/1/0/0/0) dargestellt werden. Die Koordinate 1 zeigt das Vorhandensein und die Koordinate 0 die Abwesenheit einer Eigenschaft an. Jedem Tiernamen entspricht somit ein bestimmter Eigenschaftsvektor. Diese Vektoren werden als Inputmuster dargeboten und an eine Outputschicht weitergeleitet, die aus 10*10 = 100 Outputneuronen besteht. Nach den Mechanismen der Hebbschen Lernregel und der Nachbarschaftsfunktion bilden sich in der Outputschicht gewinnende Neuronen, die jeweils eine der 16 Wortbedeutungen repräsentieren. Nach 2000 Lernschritten zeigt die Outputkarte nachfolgende Struktur:

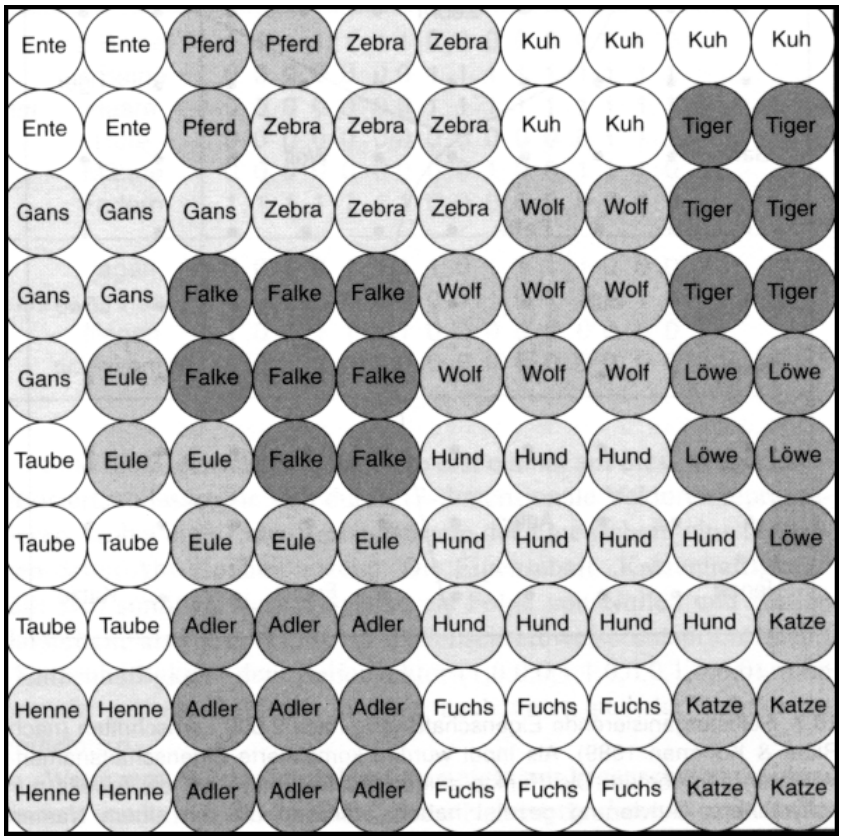


Selbstorganisierende Eigenschaftskarte nach 2 000 Lernschritten (nach Ritter & Kohonen 1989). Als Input wurden kombinierte Eigenschaftsnamen-Vektoren dargeboten. Die Punkte zeigen die Neuronen an, die jeweils nur schwächere Aktivierung gezeigt haben, während die mit einem Namen bezeichneten Neuronen die „Gewinner" repräsentieren. Es ist sehr deutlich zu sehen, wie die Namen sinnvoll auf dem Netz angeordnet sind: links die Vögel, oben die Paarhufer und rechts die Raubtiere. Auch innerhalb der Gruppen sind ähnliche Tiere auf der Karte nahe beieinander repräsentiert, unähnliche hingegen liegen weiter voneinander entfernt. Diese Ordnung entstand spontan im Netzwerk, ohne Zutun von außen.

Die auf der Karte eingezeichneten schwarzen Punkte sind nicht völlig unbeteiligt. In Wirklichkeit repräsentieren sie als direkte Umgebung eines gewinnenden Neurons ebenfalls die entsprechende Wortbedeutung, nur wesentlich schwächer. Für jede Wortbedeutung hat sich somit eine kleine Neuronengruppe auf der Karte gebildet. So funktioniert die Bedeutungskarte selbst dann noch, wenn ein Teil ihrer Neuronen ausfällt.

Die Forschungen von Ritter und Kohonen zeigen, dass ein selbstorganisierendes System die räumliche Verteilung von Gedächtnisspuren (also Knoten mit ihren Verbindungsstärken) in der Weise lernen kann, dass die Karte ein direktes Abbild von begrifflichen Beziehungen darstellt. Dabei werden einander ähnliche Inputs an benachbarten Orten der Outputschicht repräsentiert. Die Repräsentationsorte von unähnlichen Inputs liegen voneinander weiter entfernt. Diese Ergebnisse zeigen die beeindruckenden Leistungen neuronaler Netze beim Verständnis von Wortbedeutungen.

In der letzten Abbildung ist jede Zelle mit dem Wort markiert, welches dem Input entspricht, der sie am stärksten aktiviert. Wie zu sehen, sind die Neuronen in einer Umgebung des gewinnenden Neurons jeweils ebenfalls auf diesen Input spezialisiert - nur etwas schwächer. Für jedes Inputmuster hat sich somit eine kleine Neuronenpopulation gebildet, die diesen Input repräsentiert. Damit funktioniert die Karte auch dann noch, wenn ein Teil der Neuronen ausfällt.



(N.3.6) Emotionen im Netz

Emotionen bzw. Gefühle werden als bestimmte Knoten in semantischen Netzen aufgefasst (Z. V. Segal, 1996). In diesen Knoten ist ein Wissensinhalt (Mutter, Tod, usw.) mit einer affektiven Valenz (angenehm, unangenehm usw.) verknüpft. Im realen Gehirn spielt dabei die chemische Einwirkung bestimmter Transmitterstoffe auf die entsprechenden synaptischen Verbindungen eine wesentliche Rolle (Neuromodulation). Bei depressiven Störungen beispielsweise liegt ein Mangel an Serotonin vor. Serotoninmangel innerhalb des Nervensystems einerseits und negative äußere Reizsituationen (Verlust, Misserfolg usw.) andererseits führen dazu, dass in semantischen Netzwerken depressive Knoten überrepräsentiert sind. Dadurch kommt es zu einer grundsätzlichen Verschiebung der affektiven Gewichtung sowohl von bereits gelernten als auch von neuen Erfahrungen (**negative affective bias**).

Dieser Sachverhalt kann einfach durch das oben beschriebene Modell der sich ausbreitenden Aktivierung (**spreading activation**) erklärt werden. Werden semantische Netzwerke bevorzugt mit affektiv negativen Inhalten trainiert, dann stellt sich eine Übergewichtung dieser Inhalte ein, sodass auch bei weiterem Training mit an sich neutralen Inputmustern deren Verarbeitung im Sinne einer negativen Bewertung erfolgt. Dieses Phänomen wird als **Depressionsspirale** bezeichnet und kann mit neuronalen Netzwerken im Computer simuliert werden. Das Beispiel zeigt, dass mit Netzwerkmodellen zumindest einige Teilaspekte des emotionalen Geschehens erklärt werden können.

Abschließend sei erwähnt, dass in den letzten Jahren auch in der Psychiatrie erfolgreich neuronale Netze zur Modellbildung eingesetzt werden, beispielsweise zur Erklärung der Denkstörungen von Schizophrenen.

(N.4) Schlussbetrachtungen

Die moderne Neurowissenschaft hat mit dem neuronalen Netzwerk ein sparsames und zugleich äußerst mächtiges Modell für reale neurobiologische Prozesse im Gehirn entwickelt. Dieses Modell ist in hohem Maße effizient, weil es einfache und plausible Erklärungen liefert und sogar Vorhersagen von konkretem Verhalten ermöglicht. Neuronale Netze sind vor allem imstande, aus häufigen oder ähnlichen Inputmustern allgemeine Merkmalskombinationen (d.h. Begriffe) zu extrahieren und diese zu repräsentieren. Neuronale Netze sind somit zur Abstraktion fähig.

Bemerkenswert ist, dass die beiden Grundmechanismen, die Hebbsche Lernregel und die Nachbarschaftsfunktion, anscheinend auch für **soziale Beziehungen** ihre Gültigkeit haben. Eine Verbindung wird um so stärker sein, je mehr die Partner zugleich aktiv sind (what fires together, wires together). Bei Inaktivität verkümmert jede Beziehung. Genauso gilt, dass sich jedes Individuum einen engen Freundschaftskreis mit hoher Bindung schafft; wohingegen entfernte und fremde Individuen eher abgelehnt werden. Es hat den Anschein, als ob soziale Netze die Mechanismen neuronaler Netze widerspiegeln. Man denke beispielsweise an den Mechanismus des gewinnenden Neurons (the winner takes it all).

Zuletzt muss noch die philosophische Konsequenz aus den Erkenntnissen der modernen Neurowissenschaften angesprochen werden. Dazu sollen drei führende Forscher auf diesen Gebieten zu Wort kommen, der Neurobiologe und Nobelpreisträger Francis Crick, der Psychiater Manfred Spitzer und der Philosoph Paul Churchland.

Die erstaunliche Hypothese besagt Folgendes: die Menschen, ihre Freuden und Leiden, ihre Erinnerungen, ihre Ziele, ihr Sinn für die eigene Identität und Willensfreiheit - bei Alledem handelt es sich in Wirklichkeit nur um das Verhalten einer riesigen Ansammlung von Nervenzellen und dazugehörigen Molekülen (Francis Crick, 1994).

*Man kann davon ausgehen, dass jeder Mensch mit einer bestimmten neuromodulatorischen Grundausstattung geboren wird. Die chemischen Neuromodulatoren wie Serotonin, Noradrenalin und Dopamin bestimmen sein Temperament bzw. sie **sind** sein Temperament. Unterschiede in ihrem Vorhandensein zwischen einzelnen Menschen legen damit den Grundstein für Verhaltensdispositionen und Reaktionsweisen. Wir sind jedoch diesem genetischen Schicksal nicht völlig ausgeliefert, diese Neuromodulatoren und ihre Effekte sind vielmehr auch beeinflussbar durch unsere individuellen Lernerfahrungen (Manfred Spitzer, 1996).*

Das weiträumig reziprok gekoppelte neuronale Netzwerk zwischen Großhirnrinde und Thalamus im Zwischenhirn weist die sieben wichtigen Eigenschaften auf, welche man auch dem Bewusstsein zuschreibt (Kurzzeitgedächtnis, selektive Aufmerksamkeit, Integrationsfähigkeit von polymodalen Inputs, variable Interpretationsfähigkeit von Daten, von sensorischen Inputs unabhängige kognitive Aktivitäten, Träumen, Tiefschlaf). In den Begriffen der Neuroinformatik können wir beschreiben, wie jede einzelne dieser Eigenschaften generiert wird, und es ist denkbar, dass physische Strukturen unseres Gehirns sie ebenfalls derart erzeugen. Meine Hypothese sieht also folgendermaßen aus: Kognitive Aktivitäten tauchen dann und nur dann im Bewusstsein auf, wenn sie als Muster von elektrischen Signalen und synaptischen Verbindungsstärken (d.h. als Vektoren) innerhalb des neuronalen Netzwerkes repräsentiert werden (Paul Churchland, 1997).

Offensichtlich haben die Erkenntnisse der modernen Neurowissenschaften den alten, abendländischen **Dualismus** von Materie und Geist, bzw. Leib und Seele, überwunden. Damit löst sich auch das dualistische Dilemma der Wechselwirkung von Materie und Geist auf. Paul Churchland bezeichnet das Gehirn als Seelenmaschine, Manfred Spitzer als Regelerkennungsmaschine. Mit der Entwicklung des menschlichen Gehirns hat die Evolution einen faszinierenden Höhepunkt erreicht.

Bewusste Erlebnisse sind das Resultat informationsverarbeitender Prozesse in einem dichten und komplex strukturierten Netzwerk von Milliarden von Neuronen. In diesem Sinne sind sie Systemfunktionen des Gehirns. Die Frage nach einem immateriellen Sein hinter der neuronalen Aktivität kann von der naturwissenschaftlichen Psychologie nicht beantwortet werden; sie ist sehr wohl aber eine Frage des persönlichen (philosophischen) Standpunktes !