

Konnektionismus für Linguisten

Unter Informatikern, Psychologen, Philosophen und Kognitionsforschern steht der Begriff *Konnektionismus* für eine zunehmend populäre neue Forschungsrichtung, die versucht ausgehend von gewissen Vorstellungen über die Funktionsweise *neuronaler Netzwerke* einen Modellansatz für wesentliche Aspekte intelligenten Handelns zu entwickeln. Ob sich der Konnektionismus auch zur Modellierung sprachlichen Verhaltens (im weitesten Sinne) eignet, ist eine zentrale Frage des Blockseminars. Eine Vorbedingung für eine positive Antwort ist die Überwindung der Kluft zwischen Symbolismus und Konnektionismus.

Inhaltliche Schwerpunkte

- Einführung in die konnektionistische Theorienbildung
- Konnektionismus und Symbolismus als zwei unterschiedliche Perspektiven einunddesselben kognitiven Systems.
- Neuronale Netze und nichtmonotone Logik

Inhaltsübersicht

- 1 Vorbetrachtungen
 - 1.1 Kurze Geschichte neuronaler Netze (bis 1988)
 - 1.2 Konnektionismus als Geisteshaltung
 - 1.3 Einige Standardeinwände gegen *den* Konnektionismus
 - 1.4 Zielsetzung des Seminars

- 2 Elementare Grundlagen neuronaler Netze
 - 2.1 Neuronen im ZNS
 - 2.2 Künstliche Neurone
 - 2.3 Lernregeln für künstliche Neurone
 - 2.4 Hopfield-Netze
 - 2.5 *Feedforward*-Netze
 - 2.6 Übersicht über weitere Netzwerktypen

- 3 Der angebliche Widerspruch zwischen konnektionistischen Systemen und Symbolsystemen
 - 3.1 Die Essenz der Konnektionismus-Symbolismus-Debatte
 - 3.2 Nichtmonotone Logik und neuronale Netze
 - 3.2.1 Einführung
 - 3.2.2 Formale Darstellung neuronaler Netze
 - 3.2.3 Informationszustände in Hopfield-Netze
 - 3.2.4 Asymptotische *Updates* und nichtmonotone Inferen
 - 3.2.5 Gewichtete Poole-Systeme
 - 3.2.6 Die Integration von Poole-systemen und Hopfield-Netzen
 - 3.2.7 Schlußfolgerungen

Literatur

I GRUNDLEGENDE SAMMELBÄNDE

- Anderson, J.A. & Rosenfeld, E. (Eds.) 1988. *Neurocomputing: foundations of research*. Cambridge, MA: MIT Press. (enthält die klassischen Arbeiten von McCulloch & Pitts, Hebb, Grossberg, Hopfield, Kohonen, Sejnowski).
- Anderson, J.A., Pellionisz, A., & Rosenfeld, E. (Eds.) 1990. *Neurocomputing 2: Directions for research*. Cambridge, MA: MIT Press. (Nachfolgeband vergleichbarer Qualität)
- A Dinsmore, J. (Ed.) 1992. *The symbolic and connectionist paradigms - closing the gap*. Hillsdale: Lawrence Erlbaum Associates.
- McDonald, C. & McDonald, G. (Eds.) 1995. *Connectionism - debates on psychological explanation*. Oxford: Blackwell (enthält den Aufsatz von Fodor & Pylyshyn und die "klassischen" Erwiderungen)
- Reilly, R.G. & Sharkey, N.E. (Eds.) 1992. *Connectionist approaches to natural language processing*. Hillsdale: Lawrence Erlbaum Associates. (Übersicht über Anwendungen von PDP in Syntax, Semantik und Psycholinguistik)
- A Rumelhart, D.E., McClelland, J.L., and the PDP Research Group. 1986. *Parallel distributed processing*, Volume 1+2. Cambridge, MA: MIT Press. (Darstellung der Grundlagen, nicht veraltet)

II EINFÜHRUNG IN DIE KONNEKTIONISTISCHE THEORIENBILDUNG

- A Anderson, J.A. 1995. *An introduction to neural networks*. Cambridge: The MIT Press. (Lehrbuch aus Meisterhand)
- A Dorffner, G. 1991. *Konnektionismus*. Stuttgart: Teubner (Elementare Einführung, sehr anschaulich geschrieben)
- J. Elman etc. 1996. *Rethinking innateness : A connectionist perspective on development*. MIT Press. (Enthält leicht verständlichen Einführungsteil; Anwendungen bez. auf Spracherw.)
- K. Plunkett & J. Elman 1997. *Exercises in rethinking innateness*. MIT Press (enthält Simulationssoftware, Übungen zu Elman etc.)
<http://crl.ucsd.edu/innate/tlearn.html>
<ftp://ftp.crl.ucsd.edu/pub/neuralnets/tlearn/>
- Grossberg, M. 1989. *Nonlinear neural networks: Principles, mechanisms, and architectures*. *Neural Networks*, 1, 17-66.
- Hopfield, J.J. 1982. Neural networks and physical systems with emergent collective properties. *Proceedings of the National Academy Sciences*, 79, 2554-2558. (in: Anderson & Rosenfeld)

III DER ANGEBLICHE WIDERSPRUCH ZWISCHEN KONNEKTIONISTISCHEN SYSTEMEN UND SYMBOLSYSTEMEN.

- A Chalmers, J.D. 1993. *Connectionism and compositionality: Why Fodor and Pylyshyn were wrong*. *Philosophical Psychology*, 6, 305-319.
- A Fodor, J.A., Pylyshyn, Z.W. 1988. *Connectionism and cognitive architecture: A critical analysis*. *Cognition*, 28, 3-71.
- A Pollack, J.B. 1990. *Recursive distributed representation*. *Artificial Intelligence*, 46, 677-105.
- A Shastri, L. & Ajjangadde, V. 1993. *From simple associations to systematic reasoning: A connectionist representation of rules, variables and dynamic bindings using temporal*

synchrony. Behavioral and Brain Sciences, 16, 417-494.

A Smolensky, P. 1988. *On the proper treatment of connectionism*. Behavioral and Brain Sciences, 11, 1-74.

A Smolensky, P. 1987. *The constituent structure of connectionist mental states: A reply to Fodor and Pylyshyn*. Spindel Conference 1987: Connectionism and the philosophy of mind, The Southern Journal of Philosophy.

Vorschläge für Referate/Hausarbeiten

1. Assoziationsnetz zum Erlernen der Vergangenheitsformen englischer Verben (Past Tense-Modell von Rumelhart & McClelland).

Rumelhart, D.E. & McClelland, J.L. 1986. *On learning the past tenses of English verbs*. In: Rumelhart, D.E., McClelland, J.L., and the PDP Research Group. *Parallel distributed processing*, Vol. 1. Cambridge, MA: MIT Press.

Plunkett & Elman 1997 (s.ob.), Abschnitt 11 incl. Simulation.

2. Die Kritik von Pinker & Prince am Past Tense-Modell von Rumelhart & McClelland.

Pinker, S. & Prince, A. 1988. *On language and connectionism: Analysis of a parallel distributed processing model of language acquisition*. Cognition, 28, 73-193.

3. Die Einwände von Fodor und Pylyshyn gegen den Konnektionismus.

Fodor, J.A., Pylyshyn, Z.W. 1988. *Connectionism and cognitive architecture: A critical analysis*. Cognition, 28, 3-71

4. Über den richtigen Umgang mit Konnektionismus.

Smolensky, P. 1988. *On the proper treatment of connectionism*. Behavioral and Brain Sciences, 11, 1-74.

5. Konnektionismus und Kompositionalität. Die Entgegnung von Chalmers auf die Einwände von Fodor und Pylyshyn gegen den Konnektionismus.

Chalmers, J.D. 1993. *Connectionism and compositionality: Why Fodor and Pylyshyn were wrong*. Philosophical Psychology, 6, 305-319.

6. Konnektionismus und Konstituentenstruktur. Die Entgegnung von Smolensky auf die Einwände von Fodor und Pylyshyn gegen den Konnektionismus.

Smolensky, P. 1987. *The constituent structure of connectionist mental states: A reply to Fodor and Pylyshyn*. Spindel Conference 1987: Connectionism and the philosophy of mind, The Southern Journal of Philosophy.

7. Zeitliche Synchronisierung als Antwort auf die Herausforderung von Fodor und Pylyshyn.

Shastri, L. & Ajjangadde, V. 1993. *From simple associations to systematic reasoning: A connectionist representation of rules, variables and dynamic bindings using temporal synchrony*. Behavioral and Brain Sc., 16, 417-450.

8. Konnektionismus und die Pragmatik der Adjektive

Lahav, R. 1993. *The combinatorial-connectionist debate and the pragmatics of adjectives*. Pragmatics and Cognition, 1, 71-88.

9. Ein Beispiel zur Optimalitätstheorie

Prince, A. & Smolensky, P. 1991. *Optimality theory*. Technical Report #2. Rutgers University Center for Cognitive Science.

10. Experimente mit konnektionistischen Netzen

Bharat, R. & Drosen, J. 1994. *Neural network computing*. Wincrest/ McGraw-Hill. (Hopfield-Netze)
K. Plunkett & J. Elman 1997. *Exercises in rethinking innateness*. MIT Press

1 Vorbetrachtungen

1.1 Kurze Geschichte neuronaler Netze (bis 1988)

Anfänge:

McCulloch & Pitts 1943: *A logical calculus of the ideas immanent in nervous activity*

Hebb 1949: lernfähige neuronale Netze

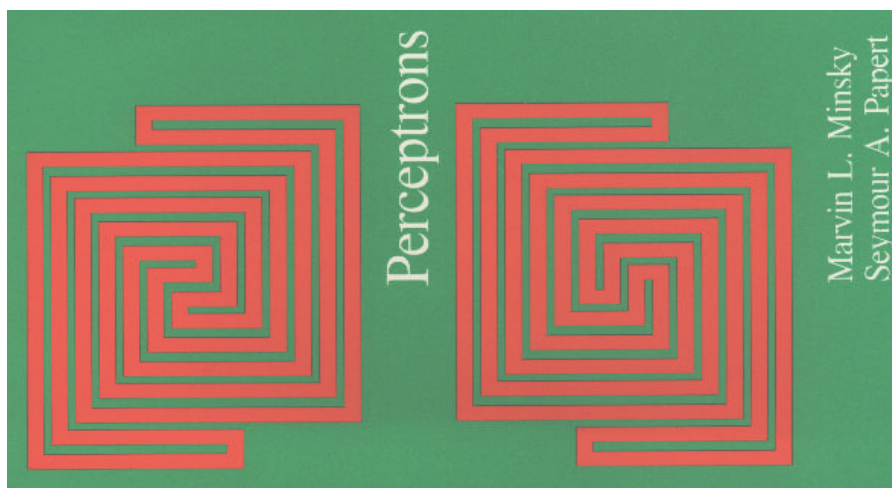
Rosenblatt 1959: Perceptron ("Erstmals besitzen wir eine Maschine, die zu eigenständigem Denken statistischen als nach logischen Grundsätzen verfahren, scheint damit klar vorgezeichnet zu sein.")

Widrow & Hoff 1960: Adaline (Mustererkennung mit Netzwerk aus adaptiven neuron-ähnlichen Einheiten)

Der Niedergang

Minsky & Papert 1969 : *Perceptrons* ("Perceptrons haben als Mustererkennungs- oder Lernmaschinen allenthalben Beachtung gefunden und sind in zahlreichen Büchern, Zeitschriftenartikeln und dickleibigen 'Reports' erörtert worden. Doch dieses Schrifttum ist zum

Nachweis, das Perceptrons grundsätzlich unfähig sind, selbst simple Aufgaben zu lösen, z.B. die Bestimmung ob die Zahl der Punkte auf einem Rasterfeld gerade oder ungerade ist, ob eine Figur zusammenhängend ist oder nicht, ...)



Papert später: "Als das Buch *Perceptrons* 1969 erschien, operierte die KI-Forschung nicht in einem Elfenbeinturm. Es ging um viel Geld. Unsere Motivation verdanken wir, wie wir in unserem Buch ganz offen zugegeben haben, zum Teil dem Umstand, daß Geldmittel und Forschungsenergien für etwas vergeudet wurden, das mir noch immer als ein untauglicher Versuch erscheint, konnektionistische Methoden in die Praxis umzusetzen."

Dunkle Jahre

Grossberg 70er Jahre: Adaptive Resonanztheorie
James A. Anderson, Teuvo Kohonen.

Neokonnektionismus

Hinton & Anderson 1979, 1981: Konferenz in San Diego, Sammelband *Parallel models of associative memory*

Hopfield 1982: Neuronale Netze und physikalische Systeme mit emergenten kollektiven Eigenschaften

Ackley, Hinton, & Sejnowsky 1985: Boltzmann-Maschine

Rumelhart & McClelland 1986: *Backpropagation* in mehrschichtigen Netzen

Sejnowski & Rosenberg 1986: NETalk (Netzwerk für lautes Lesen)

Rumelhart & McClelland 1986: Lernen der Vergangenheitsformen englischer Verben.

Schelte

Fodor & Pylyshyn 1987: Konnektionismus und kognitive Architektur (basiert auf Vortrag am UCSD; Konnektionismus liefert keine in seiner Architektur begründete Garantie für Systematizität. Es gibt nur einen Weg zur garantierten Systematizität: die klassische Architektur von Symbolsystemen.

Pinker & Prince 1988: Kritik am *Past Tense*-Modell von Rumelhart & McClelland.

1.2 Konnektionismus als Geisteshaltung

eliminativer Konnektionismus

Gespräch mit Sejnowski: "Wir wußten von Anfang an, daß das laute Lesen ein schwieriges Problem ist. Als wir loslegten, glaubte sogar Geoff Hinton, der große Stücke von diesen Netzwerkmodellen hält, das laute Lesen übersteige das Leistungsvermögen eines solchen einfachen Modells. Er meinte, wir würden dazu eine Million Neurone und nicht nur ein paar hundert benötigen. Aber wir haben trotzdem mit NETalk weitergemacht, in der Erwartung, wir würden zumindest einige Einsichten gewinnen, die zur Leistungssteigerung von neuronalen Netzen beitragen könnten. Doch wie sich gezeigt hat, kann selbst ein so bescheidenes Netzwerk mit 300 Neuronen zu 95 Prozent Gesetzmäßigkeiten - und nicht nur die Gesetzmäßigkeiten, sondern auch die Ausnahmen - der englischen Phonetik erfassen. Nun, ein kognitiver Wissenschaftler der alten Schule würde das Problem so anpacken, daß er sich hinsetzt und Tausende von Regeln und Miniregeln und Kommentare zur englischen Sprache niederschreibt. Das würde ein enormes Einfühlungsvermögen erfordern und wäre die Arbeit eines ganzen Lebens. Aber hier haben wir ein simples 300-Neuronen Netzwerk, das imstande ist, viele dieser Regeln und Regelmäßigkeiten sowie deren Ausnahmen zu extrahieren, und es schafft das über Nacht (stolzes Lächeln)."

implementativer Konnektionismus

Die eigentliche Rolle des Konnektionismus besteht darin, ein bestehendes Symbolsystem zu implementieren. (Z.B. Schnelle)

hybrides System (konnektionistisch/symbolisch)

Beide Arten von Systemen haben ihre Berechtigung: Konnektionismus für periphere (perzeptiv/motorisch) Teilsysteme, Symboltheorie für zentrales (gesteuertes) Teilsystem. (Z.B. Hendler 1989)

integrativer Konnektionismus

Symboltheorie und konnektionistische Theorie sind zwei Sichtweisen auf ein gegebenes Berechnungssystem. (Dies kontrastiert mit der Auffassung hybrider Systeme: zwei Module innerhalb eines Gesamtsystems). Ziel ist es, gewisse Gesetzmäßigkeiten, die für symbolische Systeme bestehen, als emergente Eigenschaften neuronaler Netze abzuleiten. Die partielle Reduktion von Gesetzmäßigkeiten eliminiert Symboltheorien nicht, sondern gibt ihren Inhalten eine tiefere Rechtfertigung. (Z.B. Smolenskys *Harmonietheorie* 1986)

1.3 Einige Standardeinwände gegen *den* Konnektionismus

1. Die angenommenen *künstlichen Neurone* haben nichts mit der neurophysiologischen Wirklichkeit gemeinsam, daher können neuronale Netze auch nichts zur Emergenz mentalen Verhaltens beitragen (Aber: kinematische Gastheorie und Thermodynamik; Universalien/Invarianten in der Theorie dynamischer Systeme).
2. Beschränkte Leistungsfähigkeit neuronaler Netze (Aber: versteckte Einheiten -> Turingmaschine).
3. Neuronale Netze bilden Verhalten ab wie Landkarten im Maßstab 1:1 (Aber: (i) man kann die Netze mathematisch untersuchen und damit zu gewissen Generalisierungen gelangen, s. Hopfield; (ii) man kann die Netze zerschneiden und dann ihr Verhalten untersuchen - Aphasie-Daten).
4. Sprache als teilweise angeborenes/instinktives/geprägtes/herangereiftes Verhalten unterscheidet sich von anderen erworbenen Verhaltensformen wie Klavierspielen, Backen, Kochen, Schreibmaschineschreiben etc. Dieser Unterschied wird in den bestehenden konnektionistischen Modellen an keiner Stelle ausgedrückt. (Aber: Edelmans Neuraldarwinismus postuliert zwei Grundtypen des "Lernens": (i) Selektion von Neuronen und Nervenbahnen, (ii) biochemische Veränderungen an Synapsen; nichtüberwachtes, kompetitives Lernen).
5. Konnektionismus kann Verhalten nicht wirklich erklären, bestenfalls liefert er eine Implementation eines vorgegebenen Symbolsystems. (s. Fodor & Pylyshyn)

1.4 Zielsetzung:

1. Einführung in die konnektionistische Theorienbildung
2. Der angebliche Widerspruch zwischen konnektionistischen Systemen und Symbolsystemen. Die Einwände von Fodor & Pylyshyn und ihre Entkräftung.
3. Konnektionismus und Symbolismus als zwei unterschiedliche Perspektiven einunddesselben kognitiven Systems.
4. Linguistische Anwendungen des integrativen Konnektionismus (Optimalitätstheorie; dynamische Semantik und die Logik von Propositionen; Pragmatik und konversationelle Implikaturen).

2 Elementare Grundlagen neuronaler Netze

2.1 Neurone im ZNS

Aufbau eines Neurons

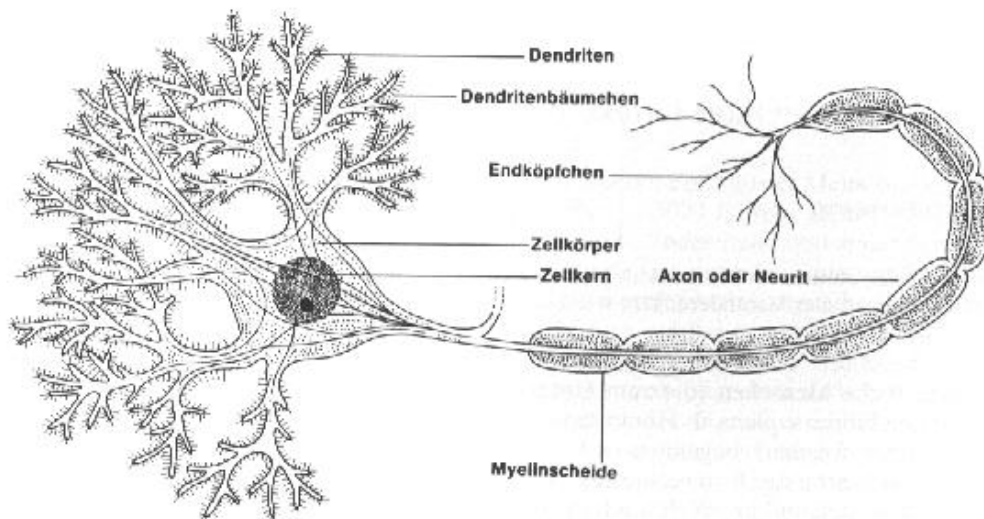


Abbildung 1: Im Körper gibt es die verschiedensten Arten von Neuronen oder Ganglienzellen, aber sie haben verschiedene Merkmale miteinander gemein, die in diesem Diagramm dargestellt sind. Die von anderen Neuronen eintreffenden Impulse stoßen auf die dendritische Zone, die hier aus der Membran der Dendriten (insbesondere den dendritischen Bäumchen) und dem Zellkörper besteht. Die resultierenden elektrischen Veränderungen in diesem Abschnitt können Erregungen auslösen, die sich entlang des Neurits ausbreiten und ihrerseits Reaktionen in den dendritischen Zonen der nachfolgenden Ganglienzellen hervorrufen oder diese beeinflussen, mit denen die dendritischen Bäumchen über die Endköpfchen in Verbindung stehen. Man beachte auch die Myelinscheide, eine segmentierte Fettschicht, die die Neuriten vieler Ganglienzellen umgibt, was die charakteristische weiße Farbe der Leitungsbahnen und Nerven erklärt.

Im menschlichen Gehirn existieren ca. 10^{10} Neurone; Zahl der Nervenzellen nach Geburt praktisch festgelegt. Nervenzellen selbst können sich nicht teilen (vermehren); Ausläufer (Dendriten und Axone) können jedoch wachsen/absterben. Diese Art "Plastizität" bleibt bis zur Pubertät erhalten (und in gewissem Umfang darüber hinaus).

Zellkörper: verantwortlich für Zellstoffwechsel (Stabilisierung der gesamten Nervenzelle)

Zellmembran: dient nicht allein der Aufrechterhaltung der physischen Form der Zelle; sie ist aktiver Bestandteil der chemischen Reaktionen, die zur Auslösung elektrischer Nervenimpulse führen. Im Innern einer Nervenzelle findet sich eine erhöhte Konzentration von Kalium, außen eine erhöhte Konzentration von Natrium. Im Ruhezustand ist eine gewisse Zahl von Kalium⁺-Ionen vom Innern der Zelle durch die Membran nach Außen diffundiert und liefert ein Ruhepotential von -70 mV (negative Ladung an der Innenseite).

Axon (Neurit): Überträgt Signale zu anderen Nervenzellen; bis zu 1 m lang; oft umgeben von Myelin-Hülle (fetthaltige Markscheide), unterbrochen von "Schnürringen" ermöglicht Leitungsgeschwindigkeiten bis zu 120 m/s. Feine Verzweigungen am Ende, abgeschlossen von "Endköpfchen" mit Synapsen.

Dendrit: verzweigte Ausläufer des Zellkörpers (bis zu 200000/Zelle), die zur Ankopplung von Axonen anderer Nervenzellen dienen (bis zu 10000 Ankopplungen/Neuron). Dendritische Zonen sind für äußere Stimulierungen empfindlicher als andere Teile der Zelle.

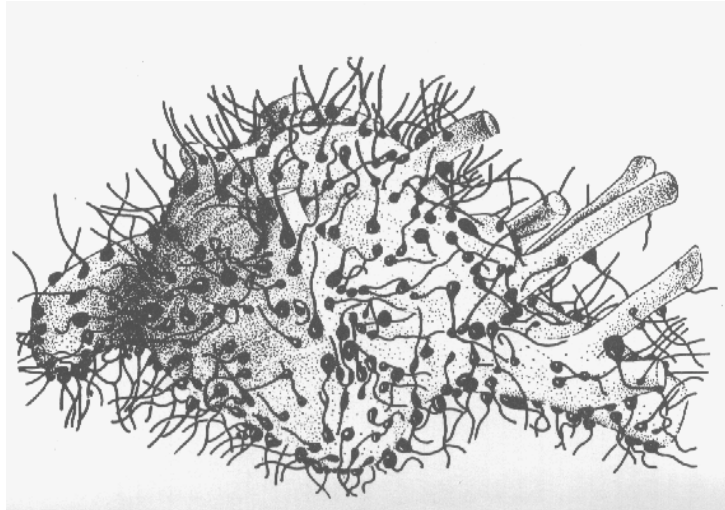


Abb. 2: Zellkörper einer motorischen Vorderhornzelle mit Hunderten von Nervenendigungen, die von anderen Nervenzellen stammen und synaptischen Kontakt mit der Zelle gewinnen.

Entstehung eines Aktionspotentials

Erregungsübertragung an der Synapse:

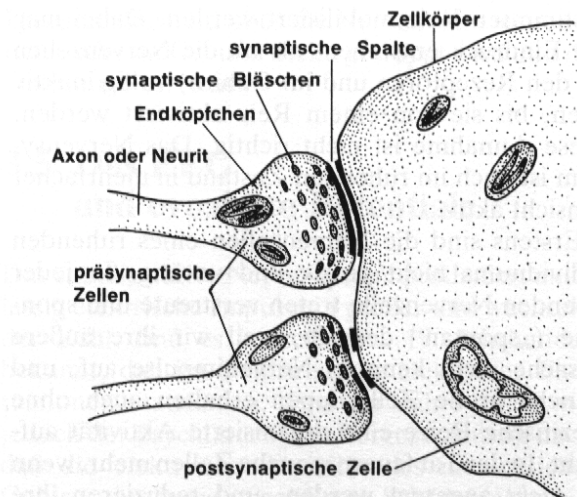


Abb. 3: An jeder Synapse nähert sich ein Axonende (Endköpfchen) eines Neurons der Zellmembran des nächsten Neurons. Die Abbildung zeigt zwei Endköpfchen, die beide viele synaptische Bläschen enthalten.

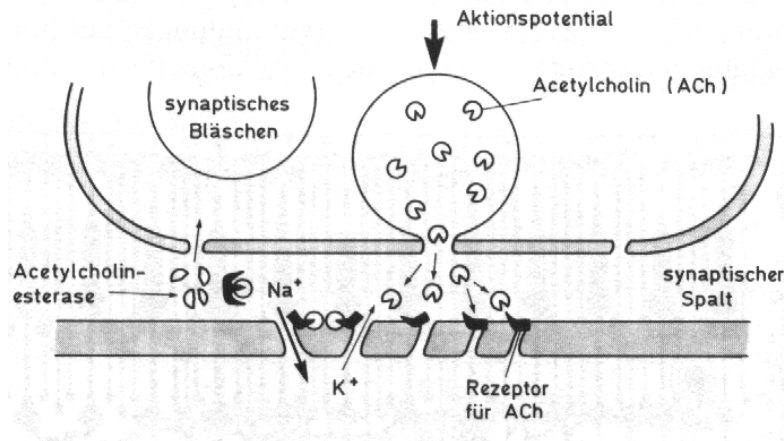


Abb. 4: Erregungsübertragung an der Synapse. Das in der Synapse ankommende Aktionspotential setzt den 'Transmitter', hier das Acetylcholin, aus den Bläschen frei. Seine Moleküle suchen gezielt entsprechende Rezeptorstellen an der gegenüberliegenden Membran auf. Dadurch werden bevorzugt Na-Kanäle frei. Vorgang wird durch das Enzym Acetylcholinesterase rückgängig gemacht.

Je nach Art des Transmitterstoffes werden Na-Kanäle (Acetylcholin) oder K-Kanäle (Dopamin) geöffnet und die Membran wird an bestimmten Stellen entladen ("erregendes" postsynaptisches Potential) oder noch stärker aufgeladen ("hemmendes" postsynaptisches Potential).

Das "Feuern" eines Neurons:

Zeitliche und räumliche Integration der postsynaptischen Potentiale (Summation der Erregungspotentiale, Subtraktion der Hemmungspotentiale). Ab einem gewissen Schwellenwert "feuert" das Neuron (die Na-Kanäle öffnen sich, Na^+ -Ionen dringen in die Zelle ein, was zu einer Entladung der Membran des Neurons führt). Die Entstehung eines Aktionspotentials läuft nach dem Alles-oder-nichts-Prinzip ab. Bereits nach ca. 5 ms hat sich das Ruhepotential wieder aufgebaut und die Zelle kann erneut entladen werden.

Informationskodierung in einem Axon:

Alle Aktionspotentiale sind gleichartig. Nicht wie eine Zelle feuert ist die relevante Information, sondern daß sie feuert. Die Frequenz und die Phase, mit der ein Neuron feuert, bilden die durch das Axon vermittelte Information.

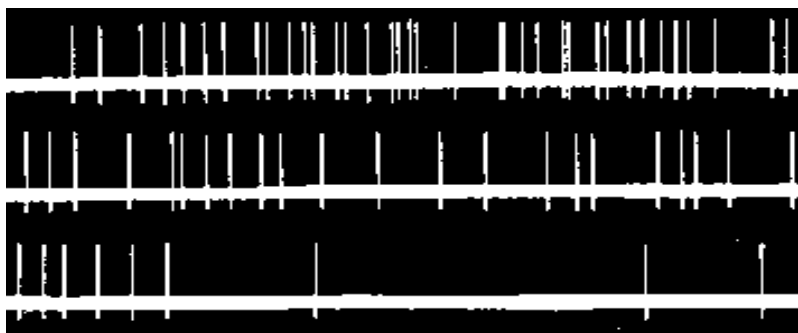


Abb 5: Aktionspotential-Registrierung von einer einzelnen Nervenzelle

Lernen

- (a) durch Veränderungen der neuronalen Vernetzung (s. Edelmans Theorie der neuronalen Selektion)
- (b) durch Veränderungen an den Synapsen und in deren Umgebung
 - Menge und Art des in der Synapse verfügbaren Neurotransmitters
 - Veränderung der Membranstabilität in Umgebung der Synapse
 - makroskopische Form und Größe der Synapse
 - bestimmte DNS- und RNS-Moleküle in Umgebung der Synapse

2.2 Künstliche Neurone

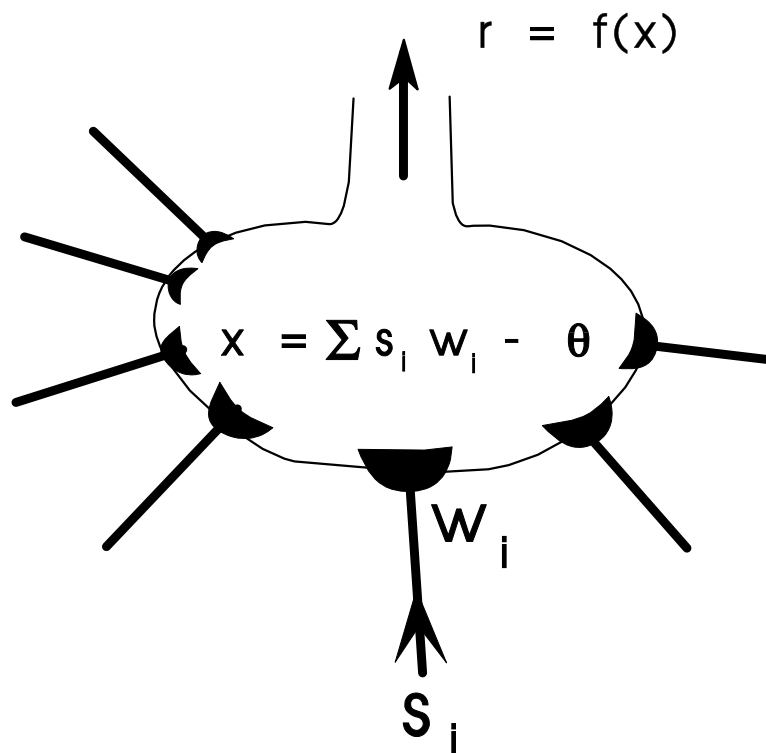


Abb. 6: Künstliches Neuron

Ein- und Ausgabemenge S

$S = [0, 1]$ (kontinuierlich)
 $S = [-1, 1]$

$S = \{0, 1\}$ (binär)
 $S = \{-1, 1\}$

Inputfunktion net (vereinfachte Beschreibung der Integration postsynaptischer Potentiale, liefert die Aktivierung des Neurons)

$$\text{net}(s_1, s_2, \dots) = \sum s_i w_i - \theta \quad (\text{Summe})$$

$$\text{net}(s_1, s_2, \dots) = \prod s_i w_i - \theta \quad (\text{Produkt})$$

$$\text{net}(s_1, s_2, \dots) = \max \{s_i w_i\} - \theta \quad (\text{Maximum})$$

Die "Gewichtsfaktoren" w_i bestimmen Zustand und Typus der Synapse; üblicherweise Variation zwischen -1 (stark hemmend), 0 (irrelevant) und +1 (stark erregend). Der Term θ verschiebt den 0-Punkt der Aktivierung. Da angenommen wird, daß das Neuron am 0-Punkt feuert, kann θ als Schwellenwert aufgefaßt werden.

Transferfunktion f (vereinfachte Beschreibung für das "Feuern" eines Neurons, liefert Ausgabe-signal)

$$f(x) = x \quad (\text{linear})$$

$$f(x) = \begin{cases} 1, & \text{falls } x \geq 0 \\ 0, & \text{sonst} \end{cases} \quad (\text{Schwellenfunktion})$$

$$f(x) = 1/(1 + \exp(-k \cdot x)) \quad (\text{Sigmoidfunktion})$$

Input-Output-Verhalten eines Neurons

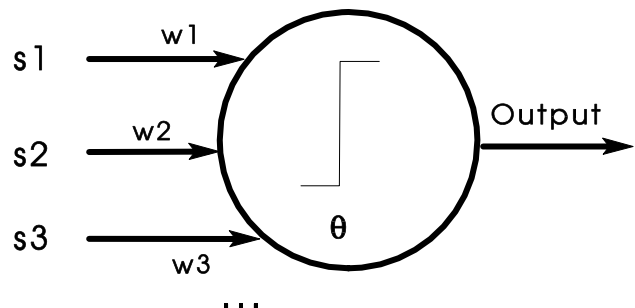
$$r(s_1, s_2, \dots) = f(\text{net}(s_1, s_2, \dots))$$

Beispiel: McCulloch-Pitts-Neuron

$$S = \{0, 1\}$$

$$\text{net}(s_1, s_2, \dots) = \sum s_i w_i - \theta$$

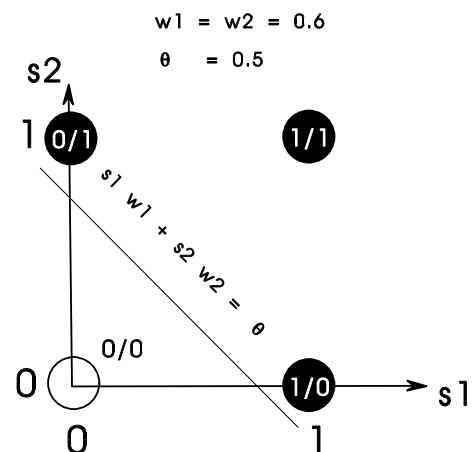
$$f(x) = \begin{cases} 1, & \text{falls } x \geq 0 \\ 0, & \text{sonst} \end{cases}$$



$$r(s_1, s_2) = \begin{cases} 1, & \text{falls } s_1 w_1 + s_2 w_2 - \theta \geq 0 \\ 0, & \text{sonst} \end{cases}$$

Fall 1: $w_1 = 0.6, w_2 = 0.6, \theta = 0.5$

s_1	s_2	net	r
0	0	-0.5	0
1	0	0.1	1
0	1	0.1	1
1	1	0.7	1



Fall 2: $w_1 = 0.3, w_2 = 0.3, \theta = 0.5$

s_1	s_2	net	r
0	0	-0.5	0
1	0	-0.2	0
0	1	-0.2	0
1	1	0.1	1

Fall 3: $w_1 = ? w_2 = ? \theta = ?$

s_1	s_2	net	r
0	0	?	0
1	0	?	1
0	1	?	1
1	1	?	0

2.3 Lernregeln für künstliche Neurone

Allgemeine Zielstellung bei der Konstruktion neuronaler Netze:

- (i) Plastizität (System paßt sich an neue Umgebung an, Lernfähigkeit)
- (ii) Stabilität (gespeicherte Information bleibt weitgehend erhalten)

A. Hebbsche Regel

When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency as one of the cells firing B, is increased. (Hebb 1949, p. 62)

Einfachste Formulierung

Simultane Erregungen am Eingang und Ausgang eines Neurons führen zu einer Erhöhung des entsprechenden Gewichtungsfaktors.

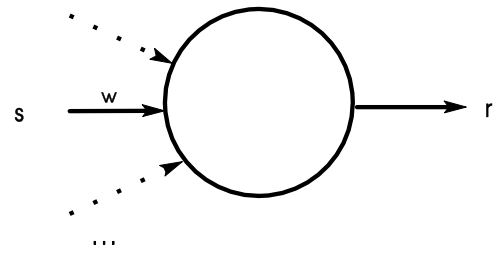
Konsequenz der Regel: Vorhandene (oder durch einen "Lehrer" herbeigeführte) positive s-r-Korrelationen werden verstärkt.

Exakte Formulierung

$$S = \{0, 1\}$$

$$w_{n+1} = w_n + \Delta w$$

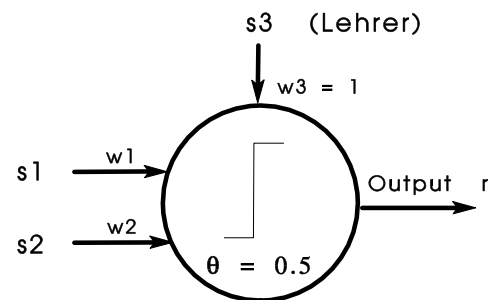
$$\Delta w = \begin{cases} \mu, & \text{falls } s=r=1 \\ 0, & \text{sonst} \end{cases} \quad (\mu \text{ heißt Lernkonstante})$$



Beispiel

AUSGANGSZUSTAND: $\theta = 0.5$; $w_1 = w_2 = 0$

s_1	s_2	s_3	r
0	0	0	0
1	0	0	0
0	1	0	0
1	1	0	0



BELEHRUNG:

s_1	s_2	s_3	r	w_1	w_2
0	0	0	0	0	0
1	0	1	1	0.3	0
0	1	1	1	0.3	0.3
1	1	1	1	0.6	0.6

ERREICHTER ZUSTAND: $\theta = 0.5$; $w_1 = w_2 = 0.6$

s_1	s_2	s_3	r
0	0	0	0
1	0	0	1
0	1	0	1
1	1	0	1

- Plastizität: System nur anfangs lernfähig; schnell stellt sich "Sättigung" ein (Gewichte können nur erhöht werden). "Umlernen" nicht möglich
- Stabilität: Bei Wahl kleiner Werte für die Lernkonstante μ verhält sich System anfangs relativ stabil. Durch "Übersättigung" der Gewichte geht leicht Information verloren, wenn Lernphase nicht rechtzeitig abgebrochen wird.

B. Generalisierte Hebbsche Regel

$$\Delta w = \mu \cdot s \cdot r$$

- $S = \{0, 1\} \Rightarrow$ einfache Hebbsche Regel
- Plastizität: S enthält auch negative Werte, z.B. $S = \{-1, +1\} \Rightarrow$ negatives Lernen: auch die negativen Korrelationen werden verstärkt, und zwar durch Verringerung des entsprechenden Gewichtungsfaktors. Dadurch wird "Umlernen" möglich.
- Stabilität: s. einfache Hebbsche Regel.

C. Delta-Regel

$$\Delta w = \mu \cdot s \cdot (r' - r) \quad (r' = \text{vom Lehrer gesetzter Sollwert, } r = \text{Outputwert})$$

Beispiel

AUSGANGSZUSTAND: $\theta = 0.5$; $w_1 = w_2 = 0$

s_1	s_2	r	r' (Soll)
0	0	0	0
1	0	0	1
0	1	0	1
1	1	0	1

BELEHRUNG:

s_1	s_2	r	r' (Soll)	w_1	w_2
0	0	0	0	0	0
1	0	0	1	0.3	0
0	1	0	1	0.3	0.3
1	1	1	1	0.3	0.3

0	0	0	0	0.3	0.3
1	0	0	1	0.6	0.3
0	1	0	1	0.6	0.6
1	1	1	1	0.6	0.6
0	0	0	0	0.6	0.6
1	0	1	1	0.6	0.6
0	1	1	1	0.6	0.6
1	1	1	1	0.6	0.6

- Plastizität: optimal — s. Konvergenztheorem für Perzeptrons: Alles was an Input-Output-Funktionen realisiert werden kann, kann durch entsprechende Belehrung erworben werden.
Problem: In Netzen mit "versteckten Einheiten" ist Sollwert für diese Einheiten nicht bekannt (\Rightarrow *Backpropagation*).
- Stabilität: Übersättigung der Gewichte wird vermieden, da die Gewichtsänderung gegen Null konvergiert, je näher der Sollwert r' dem Istwert r kommt.

Übung 1: Belehre ein Neuron entsprechend der Delta-Regel zunächst mit der AND-Funktion und ausgehend von den erreichten Gewichten anschließend mit der OR-Funktion.

AND		OR	
00	0	00	0
10	0	10	1
01	0	01	1
11	1	11	1

Gehe von folgenden Parametern aus: $\theta = 0.5$; $w_1 = w_2 = 0$; benutze die Lernkonstante $\mu = 0.3$! Was geschieht, wenn zuerst die OR-Funktion gelehrt wird und anschließend die AND-Funktion?

Übung 2: Was geschieht, wenn ein Neuron mit der XOR-Funktion

00	0
10	1
01	1
11	0

entsprechend der Delta-Regel belehrt wird? Gehe von folgenden Parametern aus: $\theta = 0.5$; $w_1 = w_2 = 0$; benutze die Lernkonstante $\mu = 0.3$!