Neuronale Netze

Heinz Horner

Institut für Theoretische Physik

Ruprecht-Karls-Universität Heidelberg

WS 2000/2001

Inhalt

1 Einführung

Ein zweifaches Ziel:

— Verständnis elementarer Schritte der Datenverarbeitung im Gehirn.

— Neuronale Netze als algorithmische Struktur für Aufgaben der KI (künstliche Intelligenz) und technische Anwendungen.

Themen der Vorlesung:

— Anatomie des Hirns und der Großhirnrinde, Neuronen und ihre Funktion.

— Das Perceptron als einfaches Modell der Funktion eines Neurons. Überwachtes Lernen. Verallgemeinerungsfähigkeit.

— Geschichtete Netzwerke.

— Rückgekoppelte Netzwerke, das Hopfield Modell.

— Selbstorganisierte Karten.

Literatur

— Braitenberg V. and Schütz A. (1991): *Anatomy of the Cortex*, Springer: Berlin.

— Amit D.J. (1989):*Modeling Brain Function — The World of Attractor Neural Networks*, Cambridge University Press: Cambridge.

— Abeles M. (1991):*Corticonics: Neural Circuits of the Cerebral Cortex*, Cambridge University Press: Cambridge.

— H. Horner und R. Kühn (1997) *Neural Networks*, in: *Intelligence and Artificial Intelligence*, Hrsg.: U. Ratsch, M. Richter und I.O. Stamatescu (Springer, Heidelberg, 1997).

— Hertz J., Krogh A., and Palmer R.G. (1991):*Introduction to the Theory of Neural Computation*, Addison Wesley: Redwood City.

— Minsky M. and Papert, S. (1969):*Perceptrons*, MIT Press: Cambridge, Mass.; enlarged edition (1988).

— Kohonen T. (1989):*Self Organization and Associative Memory*, 3rd ed., Springer: Berlin.

— Vapnik V.N. (1995):*The Nature of Statistical Learning Theory*, Springer: Heidelberg.

2 Anatomie des Gehirns, Neuronen

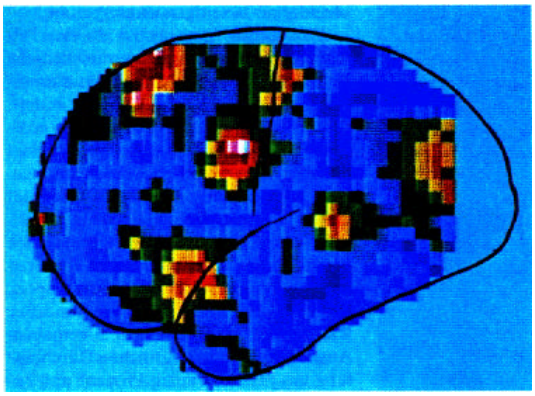
Gehirn

Großhirn:
Großhirnrinde (Cortex), ca. 2 bis 5 mm dick, enthält ca. 10^{10} Neuronen (graue Materie). Das Innere (weiße Materie) enthält Axonen (Nervenfasern) mit Myelinhüllen.
Es können verschiedene funktionelle (anatomische) Areale identifiziert werden, z.B. Sehzentrum, Hörzentrum, Riechzentrum, somatosensorische Rinde, motorische Rinde, Sprachzentrum, Rindenfelder mit komplexeren Funktionen.

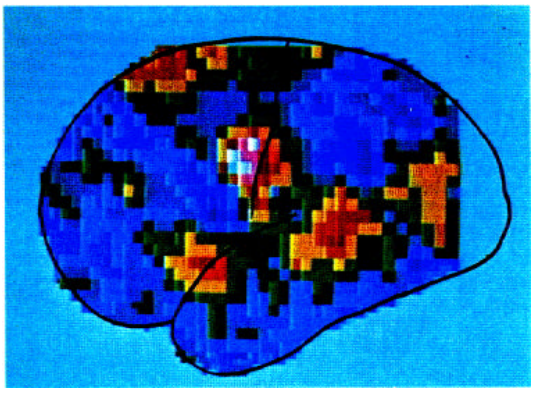
Kleinhirn:
Registerartige Strukturen in der Kleinhirnrinde. Bewegungsabläufe, Feinmotorik, e.t.c.

Stammhirn
Verbindung zum Rückenmark.

Aktivitätsmuster:
Beispiel: Verschiedene aktive Bereiche beim stillen Lesen:

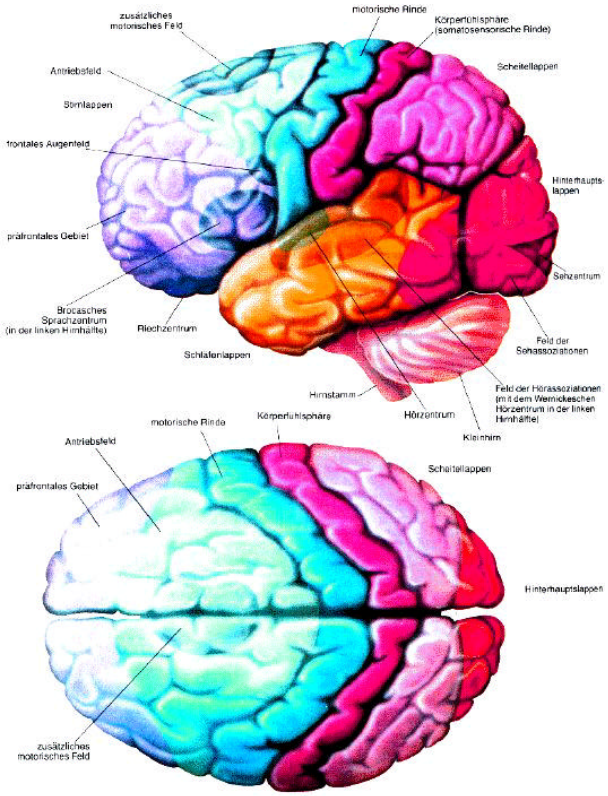


beim lauten Lesen:



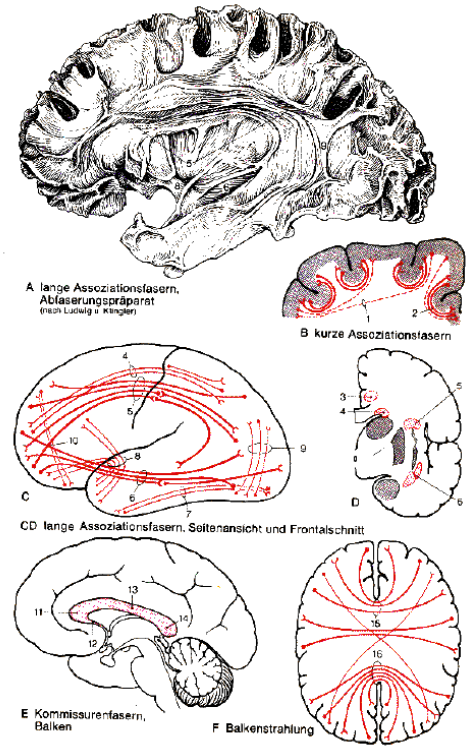
Zusätzlich ist Hörzentrum und motorisches Zentrum aktiv

Areale der Großhirnrinde:



Assoziationsfasern zwischen Arealen:

Axonale Verbindungen fast nur innerhalb des Hirnrinde.



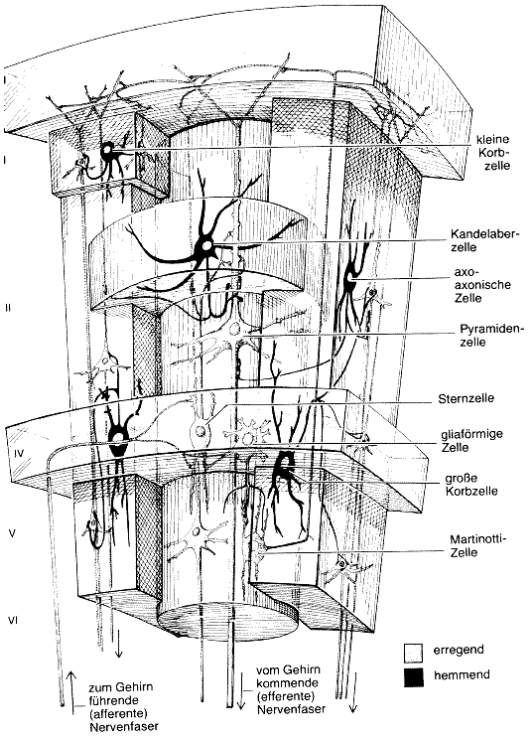
Neuronen

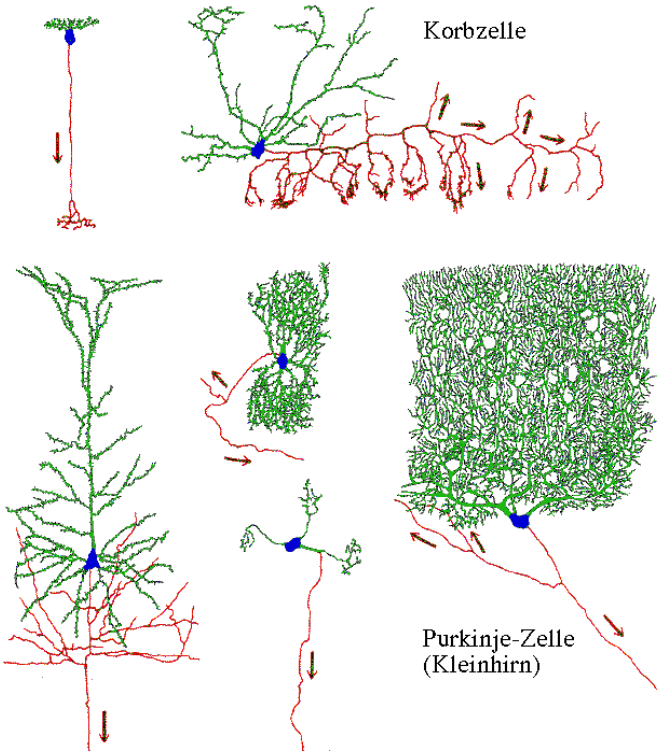
Bestandteile:
Dendritenbaum (grün)
Zellkörper, Soma (blau)
Nervenfaser, Axon (rot)

Pyramidenzellen:
Häufiger Typ in der Großhirnrinde. Erregend, lange Axonen, die teilweise in entfernte Bereiche der Großhirnrinde reichen.

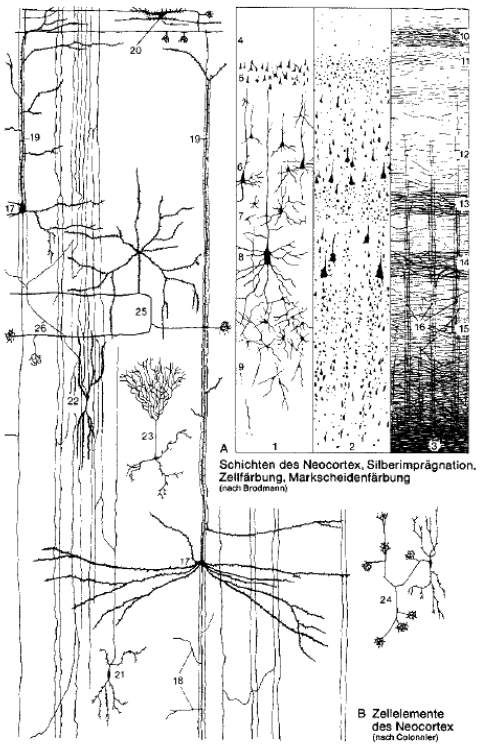
Glatte Sternzellen:
Hemmend, Axonen mit kurzer Reichweite.

Dornige Sternzellen:
Erregend, Axonen mit kurzer Reichweite.





Aufbau der Großhirnrinde
Schichten I bis VI
Pyramidenzellen (erregend) mit lokalen und langreichweitigen axonalen Verbindungen.
Andere Zelltypen mit kurzreichweitigen Axonen, zum Teil hemmend.

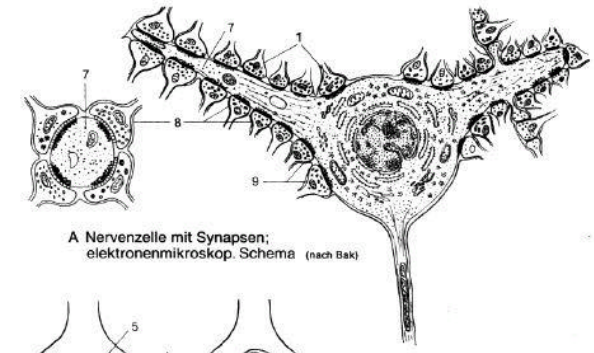


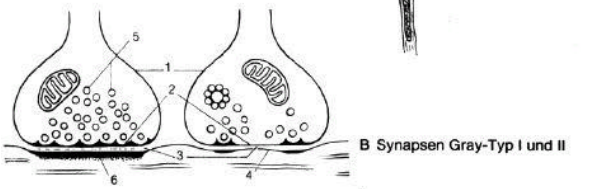
Färbungen:

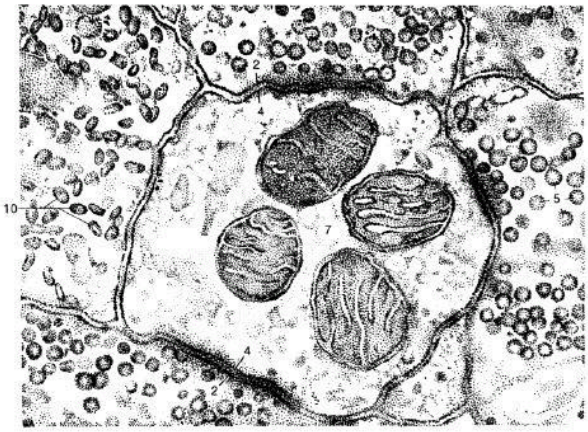
Bei Siberfärbung sind nur wenige Neuronen sichtbar.
Färbung der Zellkörper.
Färbung der Axonen (myelisiert).

Synapsen

Schaltstellen zwischen Axon eines Neurons und Dendrit eines ”nachgeschalteten” Neurons.
Dendriten und Soma eines Neurons sind dicht mit Synapsen ”vorgeschalteter” Neuronen belegt.
Die Synapsen erregender Neuronen sind vorwiegend an den Dendriten, die hemmnder Zellen an den Zellkörpern der ”nachgeschalteten” Neuronen zu finden.







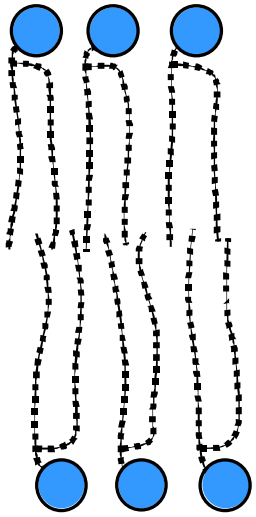
Einige Zahlen:

Dicke des Cortex:	3 - 4 mm
Fläche des Cortex:	0.5 qm
Neuronen:	$10^{10} - 10^{11}$
Neuronen pro qmm:	10^5
Synapsen:	$10^{14} - 10^{15}$
Synapsen pro Neuron:	10^4
Dendriten pro Neuron:	10 mm
Dendriten pro mm^3:	400 m
Axonenen pro mm^3:	3000 m

16.10.00

3 Neuronale Signalverarbeitung

Membran:



hydrophil

hydrophob

hydrophil

”Flüssige” Lipid Doppelschicht
Dicke ca. $5\,nm$

Zellinneres und Außenraum: Elektrolyth, wässrige Lösung von K^+, Na^+, Cl^- sowie Ca^{++} und organische Anionen$^-$.

Konzentration: $[m\,Mol/Liter]$

Ion	Radius $[\overset{o}{A}]$	Außen	Innen
Na^+	0.95	460	50
K^+	1.33	10	400
Cl^-	1.81	540	50
$org.An^-$	—	—	350

Gleichgewicht zwischen osmotischem Druck und Spannungsdifferenz: Für Ion ”x” mit Ladung Z_x, und Konzentration $C_x^{i/a}$ innen/außen.

$$V_x = \frac{R\,T}{F\;Z_x}\,\ln\frac{C_x^i}{C_x^a} \approx \frac{60}{Z_x}\,\log\frac{C_x^i}{C_x^a}\;\;[mV] \tag{3.1}$$

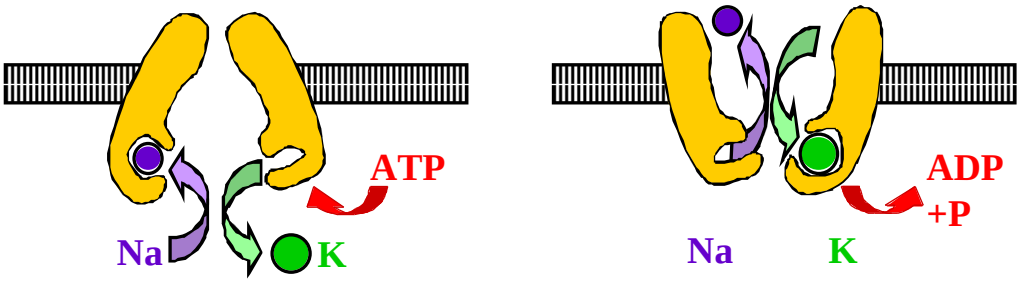
Strom J_x: g_x Leitfähigkeit der Membran für Ion ”x”

$$J_x = -Z_x\,g_x\,(V - V_x) \qquad\qquad V = \sum_x V_x \tag{3.2}$$

$V_{Na} \approx 50 \cdots 60$ mV $\quad V_{Ka} \approx -70 \cdots -100$ mV $\quad V_{Cl} \approx -50 \cdots -60$ mV
$V \approx -60 \cdots -80$ mV

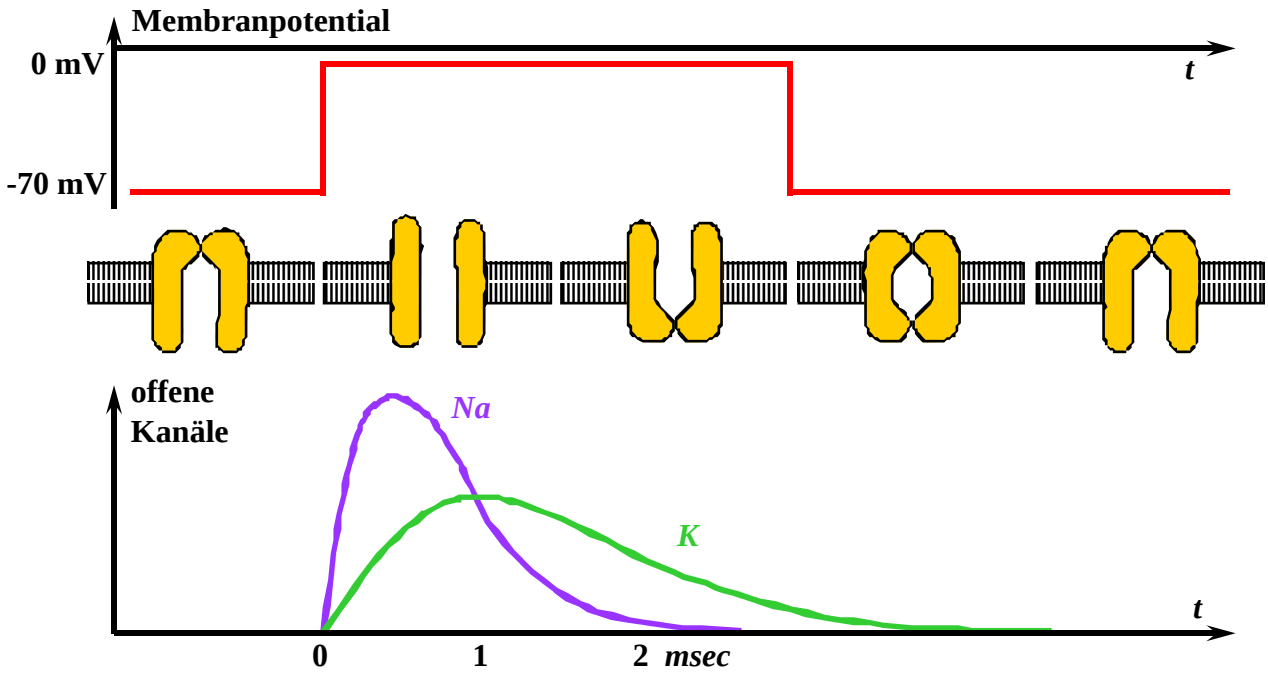
Elektrische Feldstärke: 10^5 V/cm

<u>K-Na-Pumpe</u>
Na wird nach außen, K nach innen gepumpt.

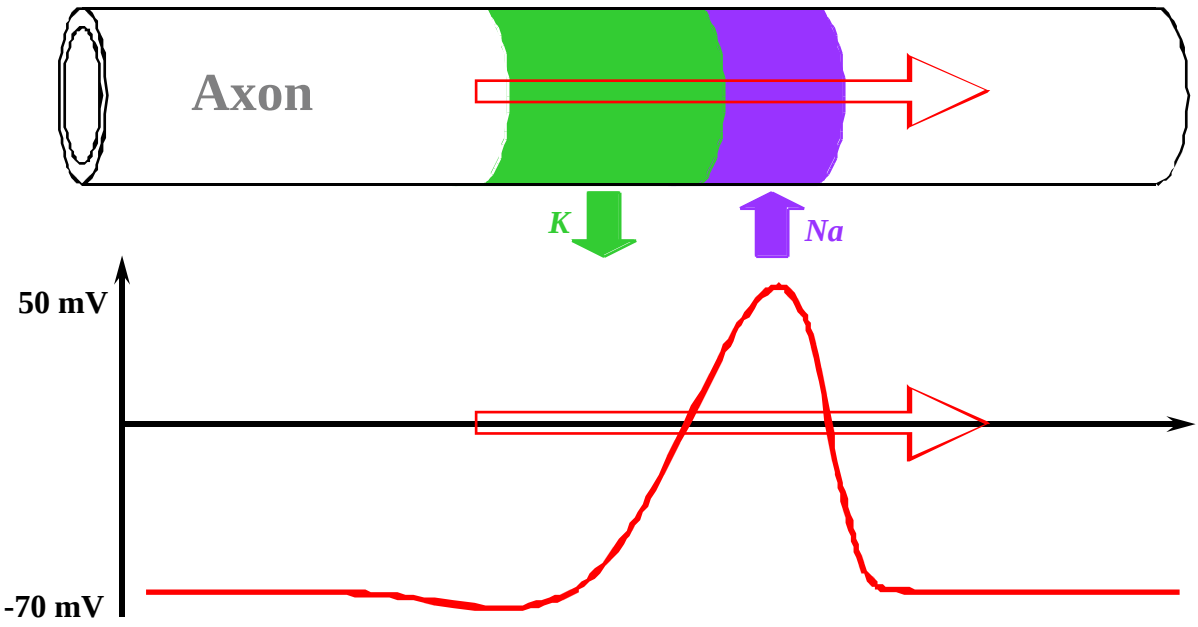


ATP: Adenosintriphosphat ADP: Adenosindiphosphat

Spannungsabhängige Kanäle

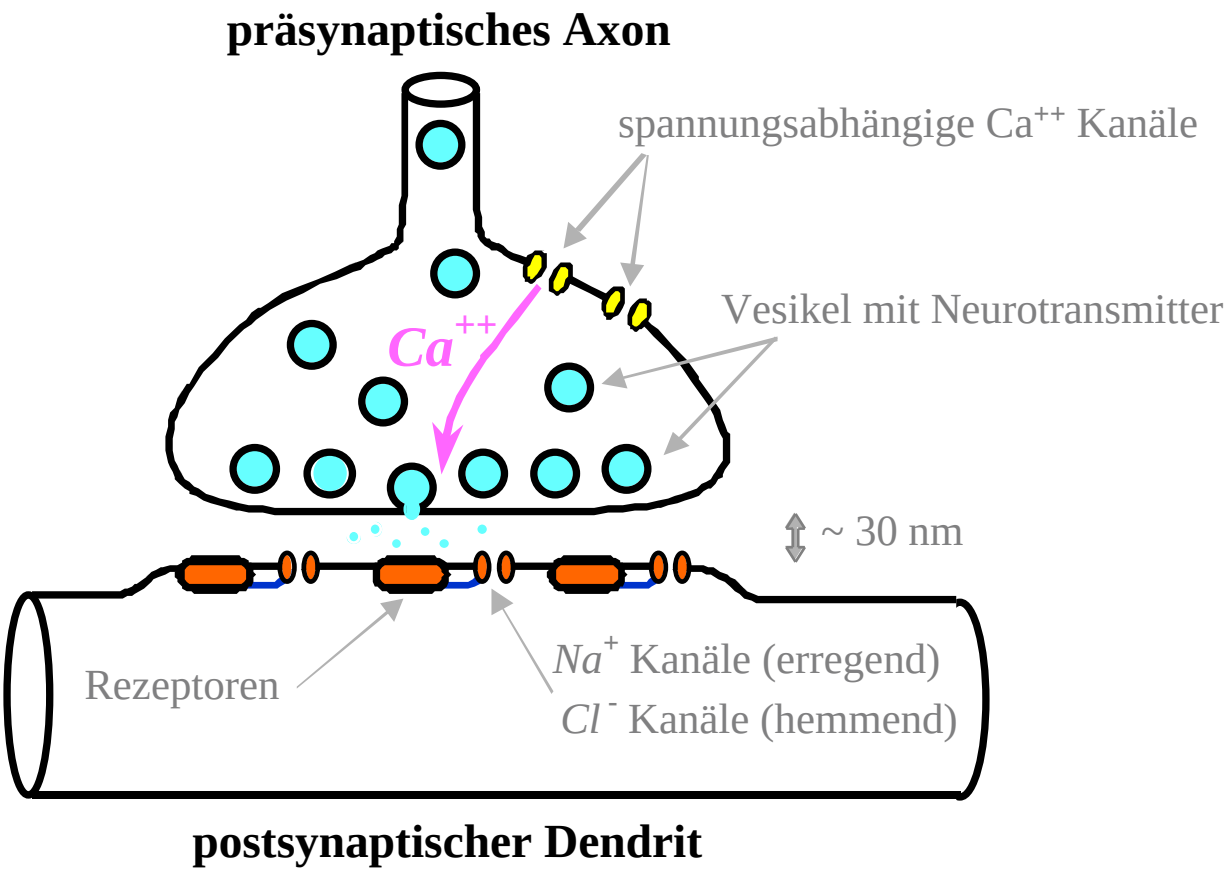


Aktionspotential, Spike



Ausbreitungsgeschwindigkeit: $\sim 3\,m/sec$ mit Myelinummantelung $\sim 20\,m/sec$

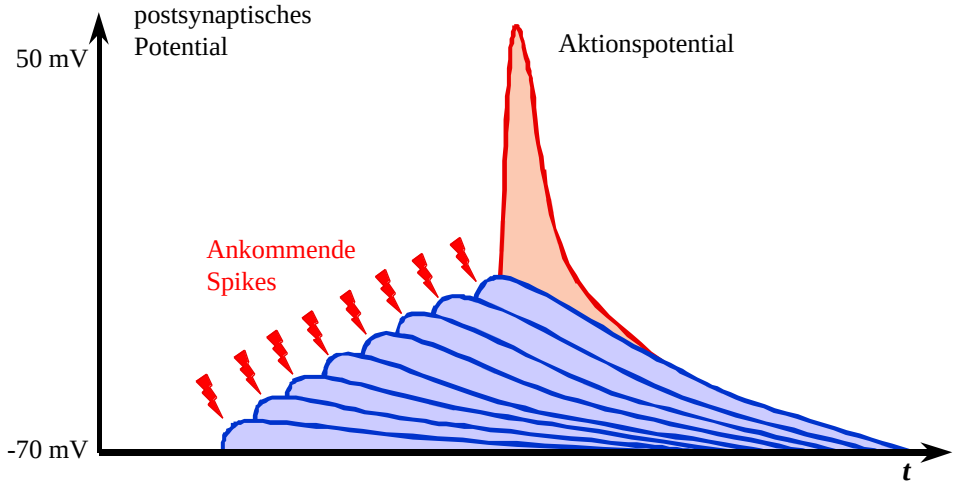
Synapsen



Ein ankommender Spike öffnet spannungsabhängige Ca^{++}–Kanäle. Das einströmende Ca^{++} bewirkt ein Öffnung eines Vesikels, der Neurotransmitter gelangt in den synaptischen Spalt und auf die Rezeptoren des postsynaptischen Dendrits (oder Soma). Diese bewirken ein Öffnen von Na^{+}–Kanälen (erregend) oder Cl^{-}–Kanälen (hemmend).

Neurotransmitter:
Erregend: Acetylcholin, Glutamat
Hemmend: GABA (gamma-aminobutyric acid)

Ankommende Spikes präsynaptischer Neuronen erhöhen das Membranpotential des postsynaptischen Neurons. Sobald eine Schwelle überschritten wird, wird ein Aktionspotential ausgelöst.



23.10.00

Modellneuronen

"Integrate and fire" Neuron"

Strom durch Membran des Neuron "i" (offene transmittergesteuerte Kanäle)

$$\partial_t J_i(t) = \sum_j W_{ij}\,\delta(t-\tau_j) - \Gamma J_i \tag{3.3}$$

W_{ij}: Effizienz einer Synapse mit präsynaptischem Neuron "j" und postsynaptischem Neuron "i". $W_{ij} > 0$: erregend, $W_{ij} < 0$: hemmend.
τ_j: Zeitpunkt des Feuerns des Neurons "j".
Γ beschreibt das Schließen der Kanäle $(1/\Gamma \sim 2$ msec).

Membranpotential:

$$\partial_t U_i(t) = \frac{1}{C} J_i(t) - \gamma\,(U_i(t) - \bar{U}) \tag{3.4}$$

C: Kapazität der Membran von Neuron "i".
γ: Inverse RC-Zeitkonstante (Wiederstand * Kapazität ~ 5 msec)
$\bar{U}$: Ruhepotential
Feuern des Neurons "i": $U_i(\tau_i) \geq \Theta$ und $U_i(\tau_i + \varepsilon) = \bar{U}$.

Ratengleichung

Feuerrate $F(U_i)$

$$\partial_t J(t) = \sum_j W_{ij}\,F(U_j(t)) - \Gamma\,J_i(t) \tag{3.5}$$

Membranpotential:

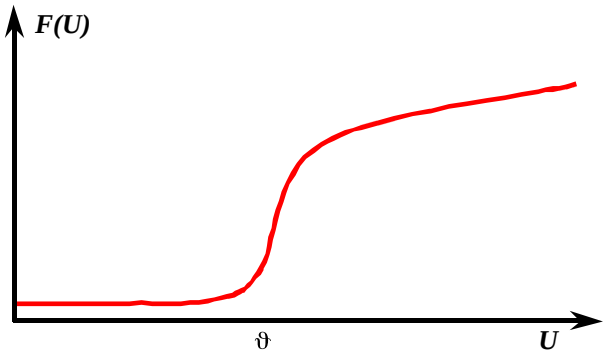
$$\partial_t U_i(t) = \frac{1}{C} J_i(t) - \gamma\,(U_i(t) - \bar{U}) - \lambda F(U_i(t)) \tag{3.6}$$

Quasistationär

$$J_i = \frac{1}{\Gamma} \sum_j W_{ij}\,F(U_j) \tag{3.7}$$

$$U_i = \bar{U} + \frac{1}{\gamma\,\Gamma C} \sum_j W_{ij}\,F(U_j)) \tag{3.8}$$

Feuerrate:



Schwelle

Lernen, Vergessen

I.P. Pawlow (1890): Bedingte Reflexe

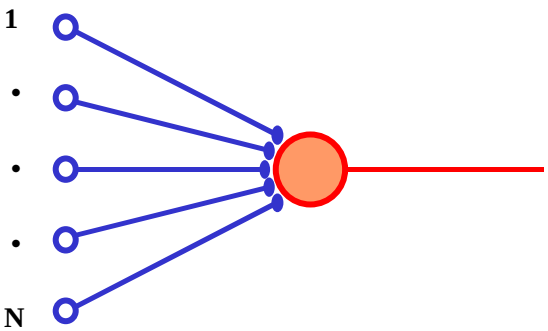
D.O. Hebb (1949): Hebb'sche Regel

Die Effizienz einer Synapse wird verstärkt falls prä- und postsynaptisches Neuron gleichzeitig aktiv ist.

$$\partial_t W_{ij}(t) = F(U_i(t))\,F(U_j(t)) \;-\; \eta\,W_{ij}(t) \tag{3.9}$$

4 Perceptron

McCulloch-Pitts Neuron (1943)



Einlaufender Reiz: f_i

Summierte Erregung: $U = \sum_i W_i f_i$

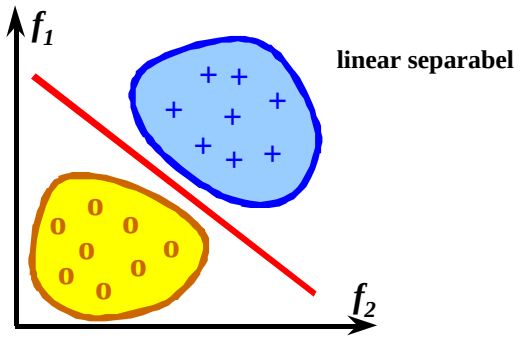
Zustand des Perceptrons: $F = 1$ falls $U > \vartheta$

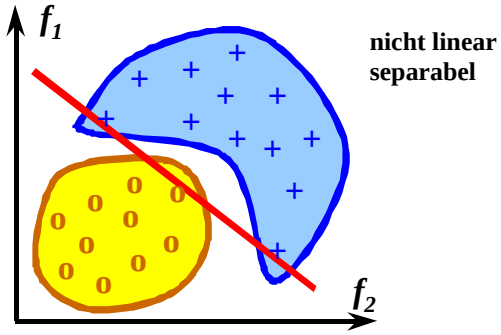
$F = 0$ falls $U < \vartheta$

$$f = \Theta\Big(\sum_{i=1}^{N} W_i f_i - \vartheta\Big) \tag{4.1}$$

Klassifizierung von Eingangsmustern

Linear separable Probleme können mit einem Perceptron behandelt werden





Beispiel:
Gegeben sei ein Prototyp Muster $\{\xi_i\}$.
Tatsächliches Muster $\{f_i\}$
Hamming Distanz (Fehler):

$$d = \sqrt{\frac{1}{N}\sum_i \big(f_i - \xi_i\big)^2} \tag{4.2}$$

Frage: Ist $d < d_o$?

$$\frac{1}{N}\sum_i \xi_i f_i > \frac{1}{2N}\sum_i \big(f_i^2 + \xi_i^2\big) - \frac{1}{2}d_o^2 \tag{4.3}$$

Perceptron mit Kopplungen

$$W_i = \frac{1}{N}\xi_i \tag{4.4}$$

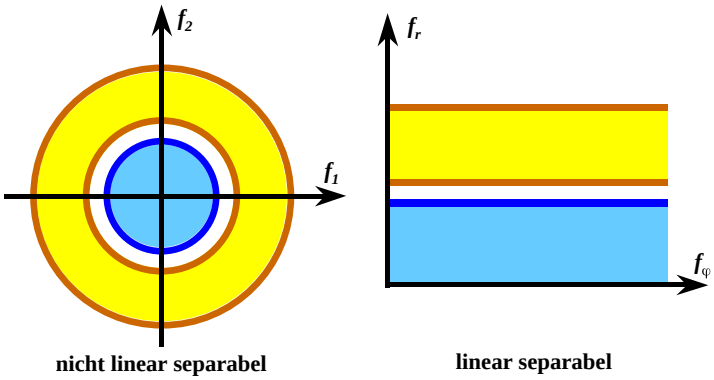
und Schwelle

$$\vartheta = \frac{1}{2N}\sum_i \big(f_i^2 + \xi_i^2\big) - \frac{1}{2}d_o^2 \tag{4.5}$$

Beispiel: Vorverarbeitung

Ohne Vorverarbeitung:
nicht linear separabel

Mit Vorverarbeitung (Polartransformation):
linear separabel



Weitere Beispiele nicht linear separabler Probleme:

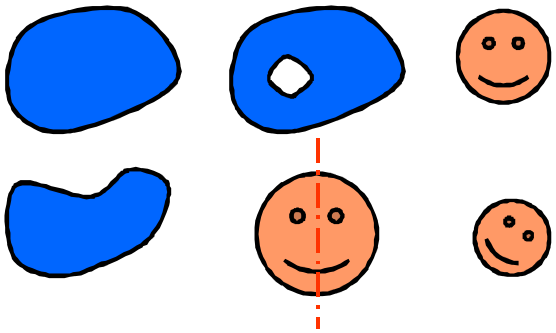
Figuren in einer Ebene
(Retina):

Ist die Figur konvex?

Hat die Figur Löcher?

Translatios-, rotations-,
skaleninvariantes Erkennen?

Symmetrien? $\cdots$?

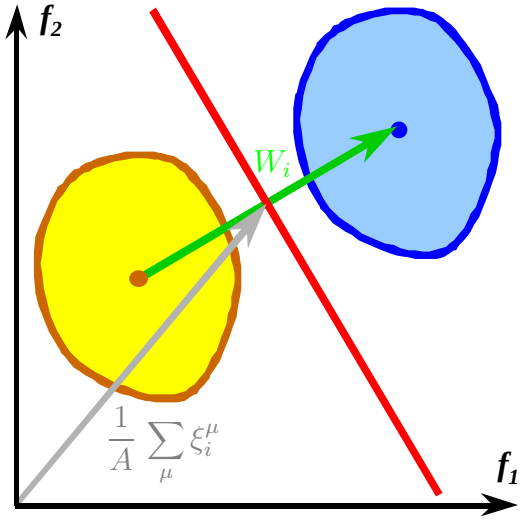


30.10.00

Lernen: Einteilung von Mustern in zwei Klasses

A Muster $\{\xi_i^\mu\}$ mit $\mu = 1 \cdots A$ und $i = 1 \cdots N$.

Klasse $\eta_\mu = \{0, 1\}$ oder $\eta_\mu = \{-1, 1\}$



Hebb'sche Lernregel

Klasse "η": A_η Muster mit $\eta_\mu = \eta = \pm 1$

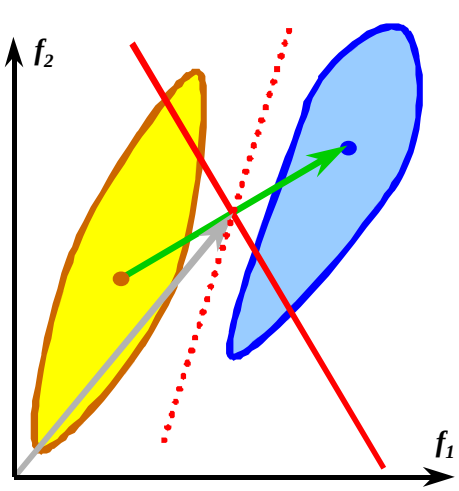
$$W_i = \frac{1}{A} \sum_\mu \xi_i^\mu \, \eta_\mu \qquad (4.6)$$

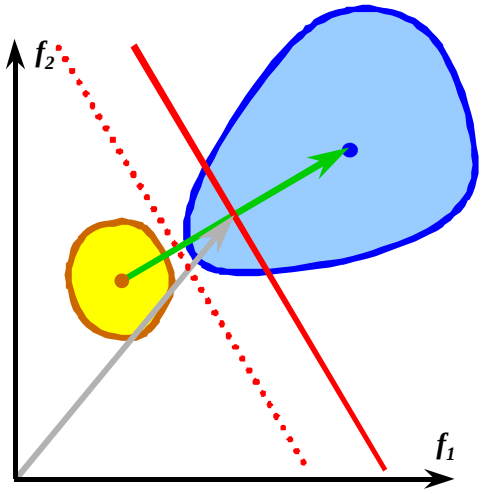
$$\vartheta = \sum_i W_i \, \frac{1}{A} \sum_\mu \xi_i^\mu \qquad (4.7)$$

$$f = \text{sign}\Big(\sum_{i=1}^{N} W_i \, f_i - \vartheta \Big) \qquad (4.8)$$

Beispiele

für linear separable Probleme, die nicht mit Hebb'scher Lernregel gelöst werden können:





Zufallsmuster, Hebb'sche Lernregel

Eingabemuster: $\xi_i^\mu = \pm 1$ unabhängig zufällig gleichverteilt $\langle \xi_i^\mu \rangle = 0$

Klassifizierung: $\eta_\mu = \pm 1$ mit Wahrscheinlichkeit $P_+ = a$ und $P_- = 1 - a$

Kopplungsstärken

$$W_i = \frac{1}{N} \sum_\mu \xi_i^\mu \, \eta_\mu \tag{4.9}$$

Potential U_λ für Muster "λ" in Klasse η_λ

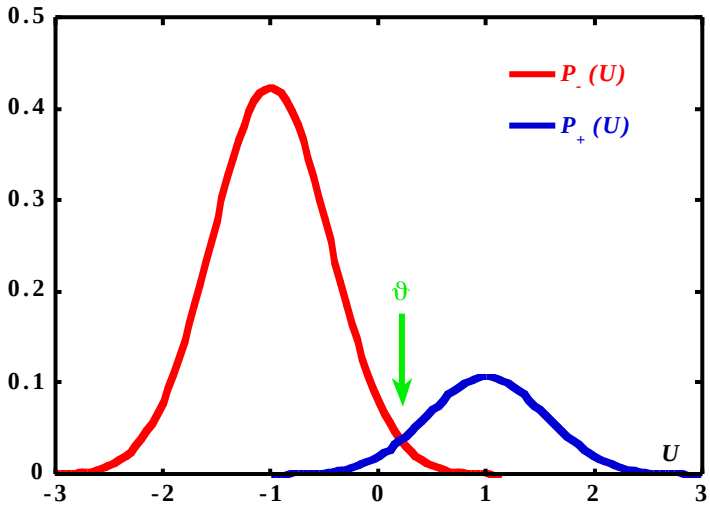
$$U_\lambda = \sum_i W_i \, \xi_i^\lambda = \frac{1}{N} \sum_i \xi_i^\lambda \sum_\mu \xi_i^\mu \, \eta_\mu = \eta_\lambda + \frac{1}{N} \sum_i \xi_i^\lambda \sum_{\mu \neq \lambda} \xi_i^\mu \, \eta_\mu \tag{4.10}$$

$$\langle U_\lambda \rangle = \eta_\lambda \tag{4.11}$$

$$U_\lambda^2 = \frac{1}{N^2} \sum_{i,j} \sum_{\mu,\nu} \xi_i^\lambda \, \xi_i^\mu \, \eta_\mu \, \xi_j^\lambda \, \xi_j^\nu \, \eta_\nu \tag{4.12}$$

Nach Mittelung nicht verschwindende Beiträge nur für $\mu = \nu = \lambda$ oder $i = j$ und $\mu = \nu$

$$\langle U_\lambda^2 \rangle - \langle U_\lambda \rangle^2 = \frac{A}{N} \tag{4.13}$$



Verteilungsfunktion für große N, mit $\alpha = A/N$

$$P_+(U) = \frac{a}{\sqrt{2\pi\alpha}} \, \mathrm{e}^{-(U-1)^2/2\alpha} \qquad P_-(U) = \frac{1-a}{\sqrt{2\pi\alpha}} \, \mathrm{e}^{-(U+1)^2/2\alpha} \tag{4.14}$$

Fehlerrate für $\eta = \pm 1$ hängt von ϑ ab

$$n_+(\vartheta) = \tfrac{1}{2} \, \mathrm{erfc}\Big(\frac{1-\vartheta}{\sqrt{2\,\alpha}}\Big) \qquad n_-(\vartheta) = \tfrac{1}{2} \, \mathrm{erfc}\Big(\frac{1+\vartheta}{\sqrt{2\,\alpha}}\Big) \tag{4.15}$$

mit

$$\mathrm{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty \mathrm{d}y \, \mathrm{e}^{-y^2} \tag{4.16}$$

Mittlere Fehlerrate $n(\vartheta) = a \, n_+(\vartheta) \; + \; (1-a) \, n_-(\vartheta)$

Beispiel:

a	α	ϑ	n_+	n_-	n
0.5	0.2	0	1.3 %	1.3 %	1.3 %
0.5	0.3	0	3.4 %	3.4 %	3.4 %
0.1	0.2	0	1.3 %	1.3 %	1.3 %
0.1	0.2	0.2	4.0 %	0.3 %	0.7 %

Speicherkapazität für Zufallsmuster

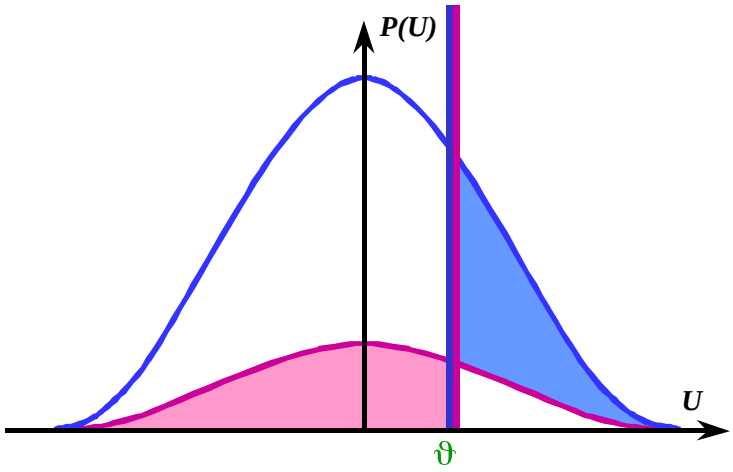
Wieviele Zufallsmuster kann man mit einem Perceptron lernen?

N Eingänge, $A = \alpha N$ Muster, $N \gg 1$

Für zufällige W_i mit $\langle W_i \rangle = 0$ und $\langle W_i^2 \rangle = 1/N$

$$P_+(U) = \frac{a}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}U^2} \qquad (4.17)$$

$$P_-(U) = \frac{1-a}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}U^2}$$



Zahl der falsch klassifizierten Mustern

$$A_f = \frac{A}{2}\Big\{a\,\mathrm{erfc}\Big(\frac{-\vartheta}{\sqrt{2}}\Big) + (1-a)\,\mathrm{erfc}\Big(\frac{\vartheta}{\sqrt{2}}\Big)\Big\} \qquad (4.18)$$

Einbettung:
Wähle A' Muster ξ_i^μ mit $\mu = 1 \cdots A'$. Konstruiere W_i so daß

$$\sum_{i=1}^{N} W_i\,\xi_i^\mu = \vartheta \qquad \text{für} \quad \mu = 1 \cdots A' \qquad (4.19)$$

Lösung existiert falls die $N \times A'$ Matrix mit Elementen ξ_i^μ von Rang A' ist.
Dies ist für Zufallsmatrizen und $A' < N$ fast immer erfüllt.

Fehlerfreies Lernen ist möglich für $A_f < A' = N$.

Für $a = \frac{1}{2}$: $\alpha_c = 2$ für $a \to 0$: $\alpha_c \to 1/a$

Perceptron Lernregel

Hebb'sches Lernen nur falls Ergebnis falsch ist:

$$\delta_\mu W_i = \frac{\Delta}{N}\Big\{\eta_\mu\,\xi_i^\mu - \gamma\,W_i\Big\} \qquad \text{falls} \qquad \eta_\mu\Big\{\sum_i W_i\,\xi_i^\mu - \vartheta\Big\} < \kappa \qquad (4.20)$$

mit $\kappa \geq 0$ und γ so daß $\sum_i W_i^2 = 1$.

Beweis für $a = \frac{1}{2}$ und $\vartheta = 0$:

Für $\alpha < 2$ existiert W_i^* so daß

$$\eta_\mu \sum_i W_i^*\,\xi_i^\mu > \kappa^* \qquad \text{und} \qquad \sum_i W_i^{*2} = 1 \qquad (4.21)$$

mit $\kappa^* > \kappa > 0$. Betrachte

$$\Omega = \sum_i W_i^*\,W_i \leq 1 \qquad (4.22)$$

Lernen von Muster "μ":

$$\delta_\mu \sum_i W_i^2 = 0 = 2\frac{\Delta}{N}\Big\{\eta_\mu \sum_i W_i\,\zeta_i^\mu - \gamma\Big\} + \sum_i (\delta_\mu W_i)^2 \tag{4.23}$$

Mit (4.20)

$$\gamma < \kappa + \mathcal{O}(\Delta) \tag{4.24}$$

Mit (4.21) und (4.24)

$$\delta_\mu \Omega = \frac{\Delta}{N}\Big\{\eta_\mu \sum_i W_i^*\,\xi_i^\mu - \gamma\,\Omega\Big\} > \frac{\Delta}{N}\Big\{\kappa^* - \kappa\,\Omega - \mathcal{O}(\Delta)\Big\} \tag{4.25}$$

Damit wächst Ω für hinreichend kleine Lerngeschwindigkeit Δ. Da $\Omega \le 1$ ist, kann nach einer endlichen Zahl von Lernschritten kein Muster mehr existieren, für das $\eta_\mu\Big\{\sum_i W_i\,\xi_i^\mu - \vartheta\Big\} < \kappa$ erfüllt ist.

13.11.00

Adaptives Lernen mit maximaler Stabilität

$$\delta_\mu W_i = \frac{\Delta_\mu}{N}\Big\{\eta_\mu\,\xi_i^\mu - \gamma_\mu\,W_i\Big\} \qquad \text{falls} \qquad \eta_\mu\Big\{\sum_i W_i\,\xi_i^\mu - \vartheta\Big\} < \kappa \tag{4.26}$$

Lernen so daß

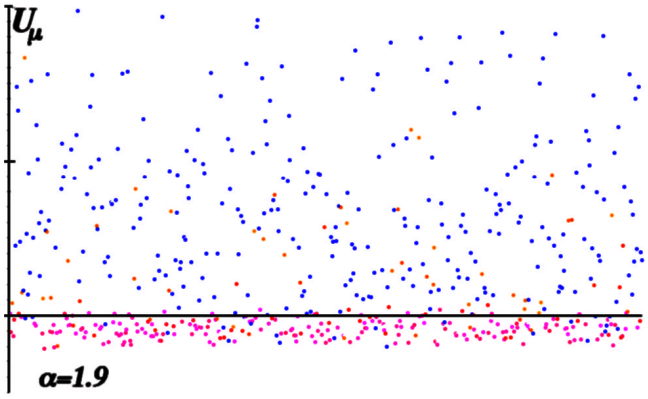
$$\eta_\mu\Big\{\sum_i (W_i + \delta_\mu W_i)\,\xi_i^\mu - \vartheta\Big\} = \kappa \tag{4.27}$$

$$\Delta_\mu = \kappa - \eta_\mu\Big\{\sum_i W_i\,\xi_i^\mu - \vartheta\Big\} \qquad \gamma_\mu = \kappa - \tfrac{1}{2}\Delta_\mu + \eta_\mu\vartheta \tag{4.28}$$

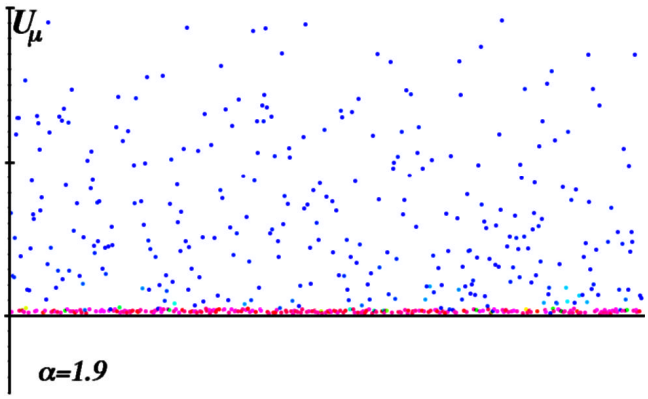
Auswahl von μ: Muster mit minimalem $\eta_\mu\{\sum_i W_i\,\xi_i^\mu - \vartheta\}$.
Bestimmung von κ: So daß für $c\,N$ Muster $\eta_\lambda\{\sum_i W_i\,\xi_i^\lambda - \vartheta\} < \kappa$ mit $c < 1$ z.B. $c = 0.7$.

Beispiel: $N = 300$, $\alpha = 1.9$, $a = 0$: Eingebettete Muster •, nicht eingebettete Muster •

$3\,N$ Lernschritte:



$30\,N$ Lernschritte:

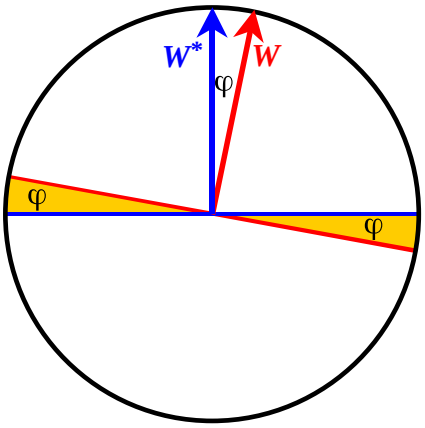


Lernen und Generalisieren

Regel nach der Muster klassifiziert werden:

"Lehrer": Perceptron mit Kopplungen $\{ W_i^* \}$

"Schüler": Perceptron mit Kopplungen $\{ W_i(\alpha) \}$ nach αN Lernschritten.



$$\sum_i \xi_i^{\mu\,2} = N \qquad \sum_i W_i^{*\,2} = 1 \qquad \sum_i W_i^2(\alpha) = 1 \tag{4.29}$$

Winkel $\varphi(\alpha)$:

$$\Omega(\alpha) = \sum_i W_i^*\, W_i = \cos\,\varphi(\alpha) \tag{4.30}$$

Fehlerwahrscheinlichkeit für ein neues Zufallsmuster, Generalisierungsfehler:

$$\varepsilon_G(\alpha) = \frac{\varphi(\alpha)}{\pi} \tag{4.31}$$

Lernen, Training:

Präsentieren und Klassifizieren eines neuen Zufallsmusters "μ": $\mu = \alpha\, N + 1$

$$\eta_\mu = \text{sign}\Big(\sum_i W_i^*\, \xi_i^\mu \Big) \tag{4.32}$$

Lernen:

$$\delta_\mu\, W_i(\alpha) = \frac{\Delta_\mu}{N} \Big\{ \eta_\mu\, \xi_i^\mu - \gamma_\mu\, W_i(\alpha) \Big\} \tag{4.33}$$

mit γ_μ so daß $\delta_\mu \sum_i W_i^2 = 0$

Für $N \gg 1$ ist

$$U_\mu^* = \sum_i W_i^*\, \xi_i^\mu \qquad \text{und} \qquad U_\mu = \sum_i W_i\, \xi_i^\mu \tag{4.34}$$

Gauss-verteilt mit $\langle U_\mu^* \rangle = 0$ und $\langle U_\mu^{*2} \rangle = 1$

$$\langle \eta_\mu\, U_\mu^* \rangle = \frac{1}{\sqrt{2\pi}} \qquad \langle \eta_\mu\, U_\mu \rangle = \frac{1 - \varepsilon_G^2(\alpha)}{\sqrt{2\pi}} \tag{4.35}$$

$\delta_\mu \sum_i W_i^2 = 0$:

$$\langle \Delta\, \gamma \rangle = \langle \Delta \eta U \rangle + \tfrac{1}{2} \langle \Delta^2 \rangle \tag{4.36}$$

$$\frac{\mathrm{d}\Omega}{\mathrm{d}\alpha} = \langle \Delta\, \eta\, U^* \rangle - \langle \Delta\, \eta\, U \rangle\, \Omega + \tfrac{1}{2} \langle \Delta^2 \rangle\, \Omega \tag{4.37}$$

Für $\varepsilon_G \ll 1$: $\Omega \approx 1 - \frac{1}{2}\, \pi^2\, \varepsilon_G^2$

$$\frac{\mathrm{d}\varepsilon_G}{\mathrm{d}\alpha} \approx -\frac{1}{\pi^2\, \varepsilon_G} \Big\{ \langle \Delta\, \eta\, (U^* - U) \rangle + \tfrac{1}{2}\, \pi^2 \varepsilon_G^2 \langle \Delta\, \eta\, U \rangle - \tfrac{1}{2} \langle \Delta^2 \rangle \Big\} \tag{4.38}$$

Hebb'sches Lernen: $\Delta_\mu = \Delta$:

$$\frac{\mathrm{d}\varepsilon_G}{\mathrm{d}\alpha} \approx -\frac{\Delta}{\pi^2\,\varepsilon_G}\Big\{\frac{1+\pi^2}{\sqrt{2\pi}}\,\varepsilon_G^2 - \tfrac{1}{2}\Delta\Big\} \tag{4.39}$$

Optimales Lernen: $\Delta = \frac{1+\pi^2}{\sqrt{2\pi}}\,\varepsilon_G^2$:

$$\varepsilon_G \sim \alpha^{-1/2} \tag{4.40}$$

Perceptron Lernen nur für falsch klassisfizierte Muster: Extra Faktor ε_G in (4.39):

$$\varepsilon_G \sim \alpha^{-1/3} \tag{4.41}$$

Schüler stellt Fragen (wählt Muster) so daß $U_\mu(\alpha) = 0$, d.h. "er ist nicht voreingenommen":

Für Muster so daß $U_\mu = 0$ ist $U_\mu^* \approx \varepsilon_G$ i.e.:

$$\frac{\mathrm{d}\varepsilon_G}{\mathrm{d}\alpha} \sim -c\,\varepsilon_G \qquad\qquad \varepsilon_G \sim \mathrm{e}^{-c\,\alpha} \tag{4.42}$$

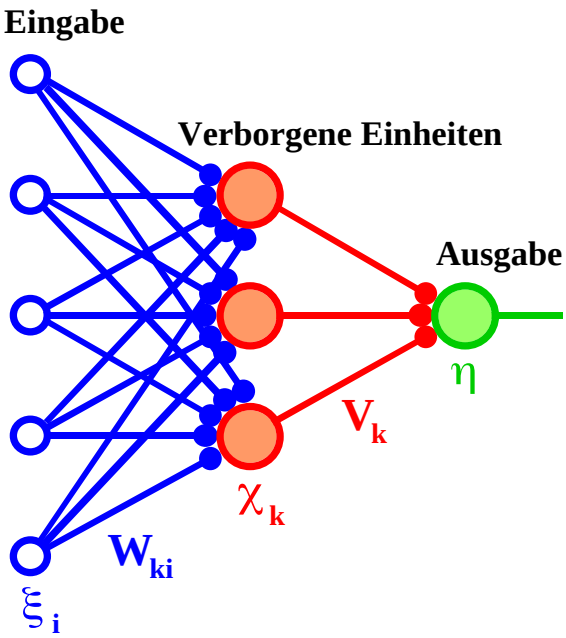
Optimales Lernen: Schüler stellt geeignete Fragen!

Perceptron Lernen, i.e. nur Lernen falls Antwort unerwartet ist, ist schlecht!

20.11.00

5 Geschichtete Netzwerke mit verborgenen Einheiten

Architektur



N Eingabeneuronen

K Verborgene Einheiten

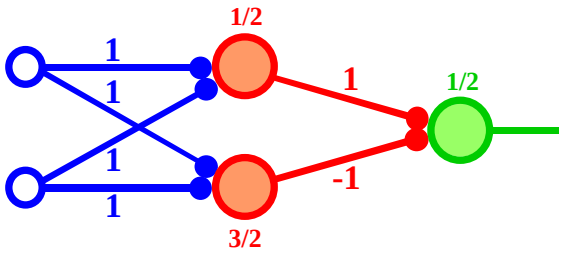
1 Ausgabeneuron

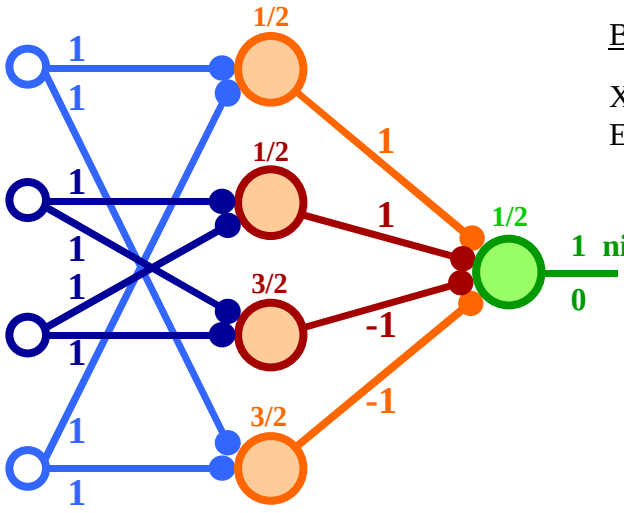
Jede Bool'sche Funktion kann mit hinreichend vielen verborgenen Einheiten dargestell werden

Eingabe: $\{1,0\}_1 \ \{1,0\}_2 \cdots \{1,0\}_N$

Ausgabe: $\{1,0\}$

Beispiel: X-or





Beispiel: Symmetrisch?

X-or Einheiten mit symmetrisch angeordneten Eingängen

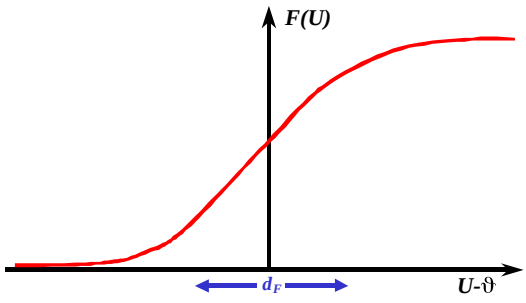
Lernen: Backpropagation

Rummelhart, Hinton, Williams (1986)
Le Chun (1985)

Neuron-Kennlinie $F(U)$

$$\chi_k = F(U_k - \vartheta_k) \qquad \eta = F_o(U_o - \vartheta_o) \qquad (5.1)$$

$$U_k = \sum_i W_{k,i}\,\xi_i \qquad U_o = \sum_k V_k\,\chi_k \qquad (5.2)$$



Lernziel:

Gewünschte Ausgabe $\bar{\eta}$, tatsächliche Ausgabe η

Kostenfunktion

$$E(\{W\},\vartheta,\{V\},\vartheta_o) = \tfrac{1}{2}\sum_\mu \left(\eta_\mu - \bar{\eta}_\mu\right)^2 \tag{5.3}$$

Lernziel: Kostenfunktion minimal, Gradienten-Abstieg

$$\begin{aligned}
\frac{\partial E}{\partial V_k} &= (\eta-\bar{\eta})\,F_o'(U_o-\vartheta_o)\,\chi_k \\
\frac{\partial E}{\partial \vartheta_o} &= -(\eta-\bar{\eta})\,F_o'(U_o-\vartheta_o) \\
\frac{\partial E}{\partial W_{ki}} &= (\eta-\bar{\eta})\,F_o'(U_o-\vartheta_o)\,V_k\,F'(U_k-\vartheta_k)\,\xi_i \\
\frac{\partial E}{\partial \vartheta_k} &= -(\eta-\bar{\eta})\,F_o'(U_o-\vartheta_o)\,V_k\,F'(U_k-\vartheta_k)
\end{aligned} \tag{5.4}$$

$$\begin{aligned}
\delta V_k &= -\Delta \sum_\mu \frac{\partial E}{\partial V_k}\Big|_{\xi_\mu} \\
\delta \vartheta_o &= -\Delta \sum_\mu \frac{\partial E}{\partial \vartheta_o}\Big|_{\xi_\mu} \\
\delta W_{ki} &= -\Delta \sum_\mu \frac{\partial E}{\partial W_{ki}}\Big|_{\xi_\mu} \\
\delta \vartheta_k &= -\Delta \sum_\mu \frac{\partial E}{\partial \vartheta_k}\Big|_{\xi_\mu}
\end{aligned} \tag{5.5}$$

$$\begin{aligned}
\delta E &= -\Delta \sum_\mu \Big\{ \sum_k \Big(\frac{\partial E}{\partial V_k}\Big|_{\xi_\mu}\Big)^2 + \Big(\frac{\partial E}{\partial \vartheta_o}\Big|_{\xi_\mu}\Big)^2 + \sum_{ik}\Big(\frac{\partial E}{\partial W_{ki}}\Big|_{\xi_\mu}\Big)^2 + \sum_k \Big(\frac{\partial E}{\partial \vartheta_k}\Big|_{\xi_\mu}\Big)^2\Big\} \\
&\quad +\mathcal{O}(\Delta^2)
\end{aligned} \tag{5.6}$$

Für hinreichend kleine Lernschritte Δ nimmt E ab.

Mögliche Probleme:

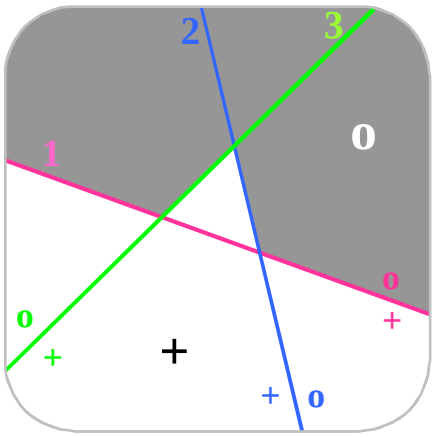
Es können lokale Minima mit $E > 0$ existieren. Auch wenn fehlerfreie Lösungen existieren, werden sie nicht gefunden.
Schlechte Konvergenz, es sind viele Lernschritte notwendig.
Kein Lernen falls $U_* - \vartheta_* \gg d_F$
In biologischen Netzwerken vermutlich nicht realisiert, nichtlokale Lernregel.

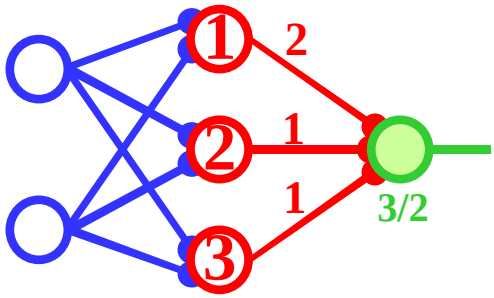
Verbesserte Lernverfahren:

Adaptive Steuerung von d_F, d_{F_o} und Δ
Konjugiertes Gradientenverfahren: Aufeinanderfolgende Schritte im Parameterraum $\{V_k, \vartheta_o, W_{ki}, \vartheta_k\}$ senkrecht zueinander.
”Simulated annealing” als universelle Optimierungsmethode.

27.11.00

Beispiel: Zweischichtiges Perceptron mit Kennlinie $F(x) = \Theta(x)$.
$N = 2,\ K = 3$

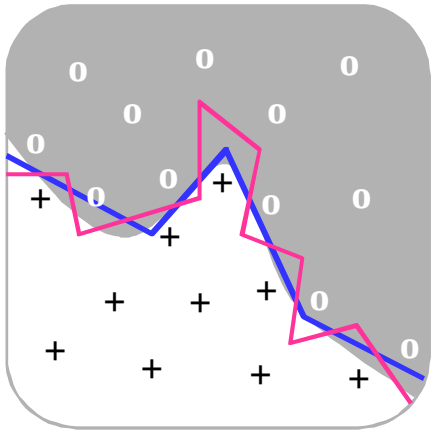




"Overfitting":

Möglichst wenig verborgene Einheiten zur Erhöhung der Generalisierungsfähigkeit.

Eliminieren schwach gekoppelter verborgener Einheiten während des Lernprozesses, "Pruning".



Vorverarbeitung, Merkmale

N Eingabeeinheiten, $i = 1 \cdots N$
K verborgene Einheiten, Merkmale $k = 1 \cdots K$
1 Ausgabeeinheit $\eta = \{1, 0\}$ oder $\{1, -1\}$
A Muster $\{\xi_i^\mu\},\ \eta_\mu$

1. Schicht: Merkmale

$$\Psi_k(\xi) = F\Big(\sum_i W_{ki}\,\xi_i\Big) \tag{5.7}$$

2. Schicht: Perceptron mit K Eingabeeinheiten.

$$\eta = F_o\Big(\sum_k V_k\,\Psi_k(\xi)\Big) \tag{5.8}$$

Kapazität für Muster mit $a = \frac{1}{2}$ und Perceptron-Lernen: $A \leq 2\,K$ falls die Merkmale $\Psi_k(\xi_\mu)$ unkorreliert sind.
Für $K > N$ ist die Kapazität größer als die eines Perceptrons mit N Eingabeeinheiten.
Bei geeigneter Vorverarbeitung können linear nicht separable Probleme bearbeitet werden.

Beispiel: Coding-Maschine

BethgeA.,KuhnR.,andHornerH.(1994): *StorageCapacityofaTwoLayerPerceptronwithFixedPreprocessingintheFirstLayer* ,J.Phys.A271929.

Muster $\xi_i^\mu = \{1, 0\}$
Gruppierung der Eingabeeinheiten in Gruppen von n Einheiten.
Es sei für die erste Gruppe $b(\xi_1 \cdots \xi_n)$ die n-stellige Binärzahl $\xi_1 \xi_2 \cdots \xi_n$, entsprechend für die weiteren Gruppen.
Jeder Gruppe werden 2^n Merkmale zugeordnet, so daß

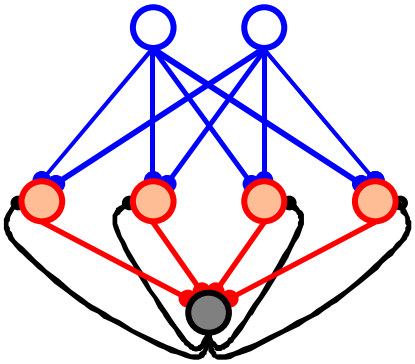
$$\begin{aligned} \Psi_k(\xi_1 \cdots \xi_n) &= 1 \qquad \text{für} \qquad k = b(\xi_1 \cdots \xi_n) \\ \Psi_k(\xi_1 \cdots \xi_n) &= 0 \qquad \text{für} \qquad k \neq b(\xi_1 \cdots \xi_n) \end{aligned} \tag{5.9}$$

entsprechend für die weiteren Gruppen.

Damit ist die Zahl der Merkmale

$$K = \frac{N}{n} 2^n \tag{5.10}$$

Nicht überwachtes Lernen, Merkmale



”Winner takes all” Funktion:
Inhibitorisches Neuron ”0” zur Kontrolle der Aktivität

$$\begin{aligned} U_k &= \sum_i W_{ki}\, \xi_i - \eta_0 - \vartheta_k \\ \chi_k &= F(U_k) \\ \eta_0 &= F_0\Big(\sum_k \chi_k\Big) \end{aligned} \tag{5.11}$$

F_0 so daß maximal ein Neuron in der Zwischenschicht aktiv ist.

Hebb’sches Lernen:

$$\delta_\mu\, W_{ki} = \Delta_W\, \xi_i^\mu\, \chi_k^\mu \qquad \sum_i W_{ki}^2 = 1 \tag{5.12}$$

Adaption, Ermüdung:

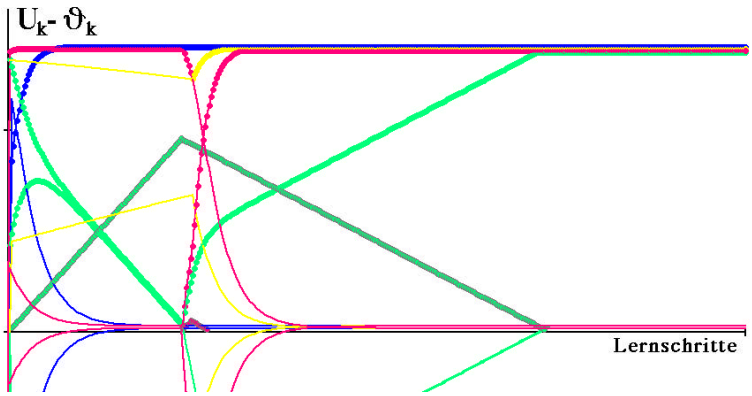
$$\delta_\mu\, \vartheta_k = \Delta_+\, \chi_k^\mu - \Delta_o \qquad \vartheta_k > 0 \tag{5.13}$$

Beispiel: $N = 2,\ K = 4, \quad \xi^\mu = \{\pm 1, \pm 1\}$ W_{1*} W_{2*} W_{3*} W_{4*}

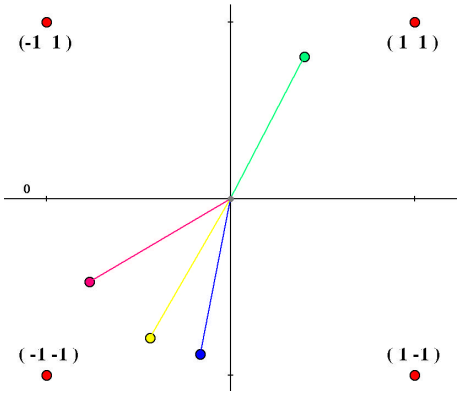
Signal:

$U_k - \vartheta_k = \sum_i W_{ki}\, \xi_i^\mu - \vartheta_k$

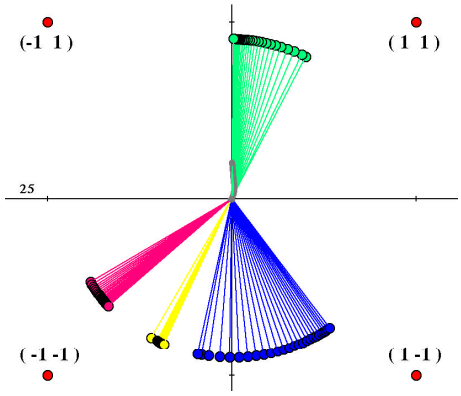
Für kurze Lernzeiten repräsentiert Neuron 2 (grün) zwei Muster: (1,1) und (-1,1)



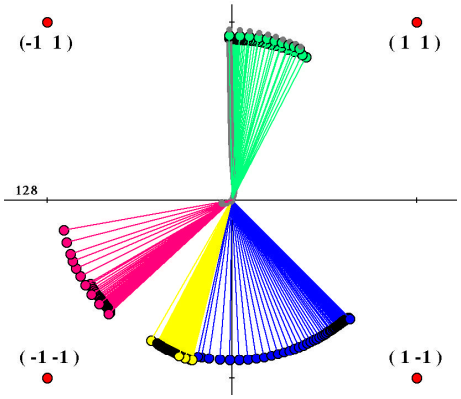
W_{ki} für verschiedene Lernschritte: $k = 1$ $k = 2$ $k = 3$ $k = 4$



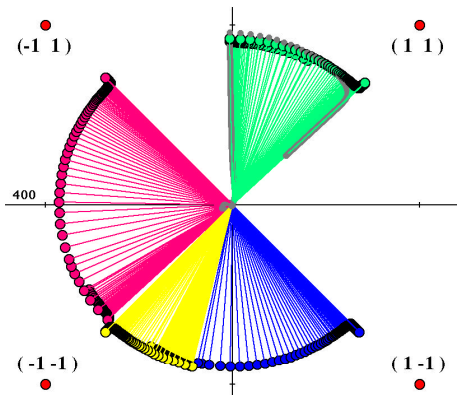
Zufälliger Anfangszustand



25 Lernschritte:
Neuron 1: $(1, -1)$ $\vartheta_1 = 0$
Neuron 2: $(1, 1)$ und $(-1, 1)$ $\vartheta_2 > 0$
Neuron 3: — $\vartheta_3 = 0$
Neuron 4: $(-1, -1)$ $\vartheta_4 = 0$



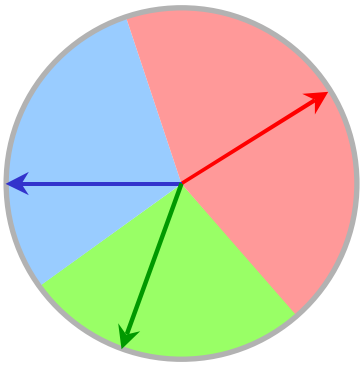
120 Lernschritte:
Neuron 1: $(1, -1)$ $\vartheta_1 = 0$
Neuron 2: $(1, 1)$ und $(-1, 1)$ $\vartheta_2 \approx 1$
Neuron 3: — $\vartheta_3 = 0$
Neuron 4: $(-1, -1)$ $\vartheta_4 = 0$



200 Lernschritte:
Neuron 1: $(1, -1)$ $\vartheta_1 = 0$
Neuron 1: $(1, 1)$ $\vartheta_2 = 0$
Neuron 3: $(-1, -1)$ $\vartheta_3 = 0$
Neuron 4: $(-1, 1)$ $\vartheta_4 = 0$

4.12.00

Merkmale für ”Winner takes all”



6 Lernen aus Beispielen, Supportvektor Maschine

Allgemeine Überlegungen

Vapnik (1995)

Wahrscheinlichkeit für Daten (Muster $\cdots$): $P(\boldsymbol{\xi},\eta)$ $\quad \boldsymbol{\xi}=\{\xi_1\cdots\xi_N\}$

Parameter (Kopplungen, Schwellen $\cdots$): $\quad \boldsymbol{\alpha}=\{\alpha_1\cdots\alpha_D\}$

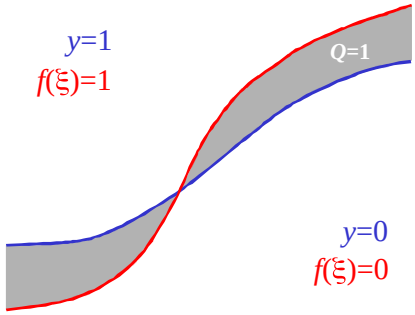
Kostenfunktion:

$$E(\boldsymbol{\alpha})=\int \mathrm{d}\xi\mathrm{d}\eta\, P(\boldsymbol{\xi},\eta)\, Q(\boldsymbol{\xi},\eta,\boldsymbol{\alpha}) \tag{6.1}$$

Beispiel: Klassifizierung $\eta=\{1,0\}$

Testfunktionen $f(\boldsymbol{\xi},\boldsymbol{\alpha})=\{1,0\}$

$$Q(\boldsymbol{\xi},\eta,\boldsymbol{\alpha})=\begin{cases}0 & \text{für } f(\boldsymbol{\xi},\boldsymbol{\alpha})=\eta\\ 1 & \text{für } f(\boldsymbol{\xi},\boldsymbol{\alpha})\neq\eta\end{cases} \tag{6.2}$$



Beispiel: Regressionsanalyse:

$$P(\boldsymbol{\xi},\eta)=P(\boldsymbol{\xi})\,P(\eta|\boldsymbol{\xi}) \qquad P(\boldsymbol{\xi})=\int \mathrm{d}\eta\, P(\boldsymbol{\xi},\eta) \tag{6.3}$$

$$f_0(\boldsymbol{\xi})=\int \mathrm{d}\eta\,\eta\, P(\eta|\boldsymbol{\xi}) \qquad Q(\boldsymbol{\xi},\eta,\boldsymbol{\alpha})=\Big(\eta-f(\boldsymbol{\xi},\boldsymbol{\alpha})\Big)^2 \tag{6.4}$$

$E(\boldsymbol{\alpha})$ minimal für $f(\boldsymbol{\xi},\boldsymbol{\alpha}_0)=f_0(\boldsymbol{\xi})$; $\quad E(\boldsymbol{\alpha}_0)=0$

Beispiel: Schätzung einer Verteilung $P(\boldsymbol{\xi})$: Testfunktion $p(\boldsymbol{\xi},\boldsymbol{\alpha})$

$$Q(\boldsymbol{\xi},\boldsymbol{\alpha})=-\ln\Big(p(\boldsymbol{\xi},\boldsymbol{\alpha})\Big) \tag{6.5}$$

mit

$$p(\boldsymbol{\xi},\boldsymbol{\alpha})=P(\boldsymbol{\xi})\Big(1+g(\boldsymbol{\xi},\boldsymbol{\alpha})\Big) \quad \text{und} \quad \int \mathrm{d}\xi\, P(\boldsymbol{\xi})\, g(\boldsymbol{\xi},\boldsymbol{\alpha})=0 \tag{6.6}$$

und $-\ln(1+x)\geq -x$: $\quad E(\boldsymbol{\alpha})$ minimal für $p(\boldsymbol{\xi},\boldsymbol{\alpha}_0)=P(\boldsymbol{\xi})$

Empirische Kostenfunktion

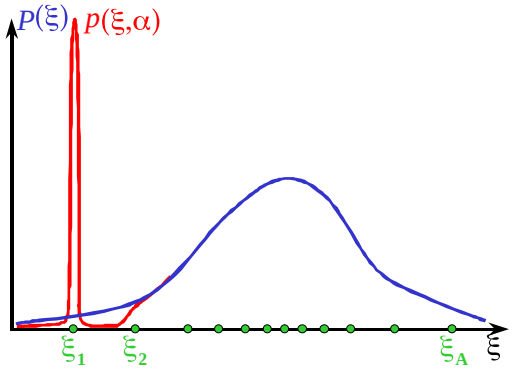
A Beispiele $(\boldsymbol{\xi}_1,\eta_1)\cdots(\boldsymbol{\xi}_\mu,\eta_\mu)\cdots(\boldsymbol{\xi}_A,\eta_A)$ gezogen mit Wahrscheinlichkeit $P(\boldsymbol{\xi},\eta)$

$$E_A(\boldsymbol{\alpha})=\frac{1}{A}\sum_{\mu=1}^{A} Q(\boldsymbol{\xi}_\mu,\eta_\mu,\boldsymbol{\alpha}) \tag{6.7}$$

Lernen aus Beispielen: $\quad E_A(\boldsymbol{\alpha})$ minimal?

Im Allgemeinen nicht wohldefiniert!

Beispiel: Schätzen einer Verteilung



Für

$$p(\boldsymbol{\xi}_1, \boldsymbol{\alpha}) = p_1 \to \infty \tag{6.8}$$

$$E_A(\boldsymbol{\alpha}) \to -\frac{1}{A} \log p_1 \to -\infty \tag{6.9}$$

d.h. das Problem ist nicht wohl definiert falls $E(\boldsymbol{\alpha})$ durch $E_A(\boldsymbol{\alpha})$ ersetzt wird ("Overfitting")

Entsprechend für andere Lernprobleme.

Forderung:

Mit

$$E_A(\boldsymbol{\alpha}_A) = \inf_{\alpha} E_A(\boldsymbol{\alpha}) \tag{6.10}$$

$$E(\boldsymbol{\alpha}_A) \underset{A\to\infty}{\longrightarrow} \inf_{\alpha} E(\boldsymbol{\alpha}) \quad \text{und} \quad E_A(\boldsymbol{\alpha}_A) \underset{A\to\infty}{\longrightarrow} \inf_{\alpha} E(\boldsymbol{\alpha}) \tag{6.11}$$

für fast alle $(\boldsymbol{\xi}_1, \eta_1) \cdots (\boldsymbol{\xi}_A, \eta_A)$.

Äquivalent:

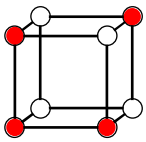
$$\sup_{\alpha} \Big| \int \mathrm{d}\xi \mathrm{d}\eta \, P(\boldsymbol{\xi}, \eta) \, Q(\boldsymbol{\xi}, \eta, \boldsymbol{\alpha}) - \frac{1}{A} \sum_{\mu} Q(\boldsymbol{\xi}_\mu, \eta_\mu, \boldsymbol{\alpha}) \Big| \underset{A\to\infty}{\longrightarrow} 0 \tag{6.12}$$

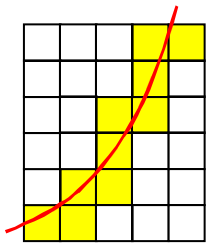
Dies ist eine Einschränkung an die Testfunktionen $f(\boldsymbol{\xi}, \boldsymbol{\alpha})$ b.z.w. $Q(\boldsymbol{\xi}, \eta, \boldsymbol{\alpha})$.

11.12.00

VC-Entropie:

Vapnik, Chervonenkis (1981)





Betrachte A-dimensionalen Raum $q_1 \cdots q_A$

und die Manigfaltigkeit $\{q_\mu = Q(\boldsymbol{\xi}_\mu, \eta_\mu, \boldsymbol{\alpha})\}$

Der Raum werde in Zellen der Größe ϵ eingeteilt.

Es sei $N(A, \epsilon)$ die Zahl der Zellen, die die Manifaltigkeit enthalten.

VC-Entropie $H_{VC}(A, \epsilon)$:

$$H_{VC}(A, \epsilon) = \ln N(A, \epsilon) \tag{6.13}$$

Gl.(6.12) gilt für

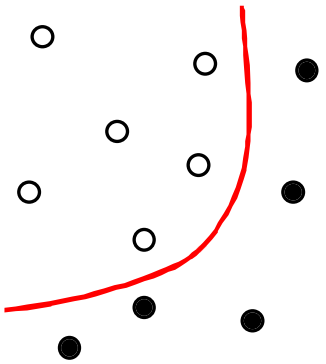
$$\frac{1}{A} H_{VC}(A, \epsilon) \underset{A\to\infty}{\longrightarrow} 0 \tag{6.14}$$

Beispiel: Klassifizierung $\eta = f_0(\xi) = \{1, -1\}$ $f(\boldsymbol{\xi}, \boldsymbol{\alpha}) = \{1, -1\}$

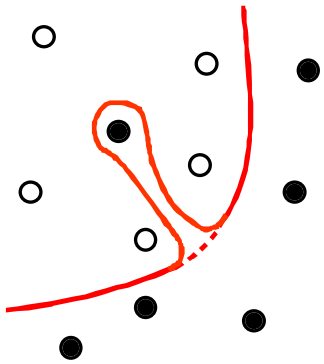
$$Q(\boldsymbol{\xi}, \boldsymbol{\alpha}) = \mathrm{sign}\Big(f(\boldsymbol{\xi}, \boldsymbol{\alpha}) f_0(\xi)\Big) \tag{6.15}$$

$\lim_{A\to\infty} H_{VC}(A) > 0$

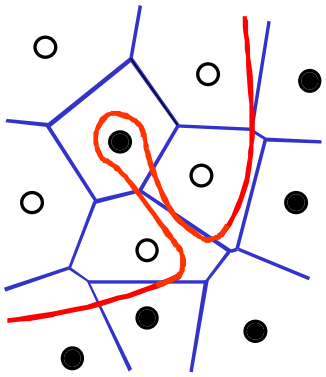
falls Konfigurationen der Art $Q(\boldsymbol{\xi}, \boldsymbol{\alpha}')$ zugelassen sind



Konfiguration $\boldsymbol{\alpha}$



Konfiguration $\boldsymbol{\alpha}'$



Konfiguration $\boldsymbol{\alpha}'$

Konstruiere Zellen V_μ um $\boldsymbol{\xi}_\mu$ so daß

$$\int_{V_\mu} \mathrm{d}\xi\, P(\xi) = \frac{1}{A} \tag{6.16}$$

Argument von Gl.(6.12):

$$\int \mathrm{d}\xi\, P(\boldsymbol{\xi})\, Q(\boldsymbol{\xi}, \boldsymbol{\alpha}) - \frac{1}{A}\sum_\mu Q(\boldsymbol{\xi}_\mu, \boldsymbol{\alpha})$$

$$= \sum_\mu \int_{V_\mu} \mathrm{d}\xi\, P(\xi)\Big\{Q(\boldsymbol{\xi}, \boldsymbol{\alpha}) - Q(\boldsymbol{\xi}_\mu, \boldsymbol{\alpha})\Big\} \tag{6.17}$$

Konfigurationen der Art $Q(\boldsymbol{\xi}, \boldsymbol{\alpha}')$ liefern endlichen Beitrag.

Lernen durch Beispiele, Gl.(6.11), ist möglich falls Zahl der Beispiele (Muster) groß gegen Zahl der Parameter α ist: $A \gg D$

Perceptron Lernen mit maximaler Stabilität

Äquivalente Formulierung (Kap. 10): Stabilitätsparameter $\kappa = 1$ und Summe über Kopplungen $\sum_i W_i^2$ minimal.

Kostenfunktion mit Lagrange-Parametern für Nebenbedingungen:

$$L(\boldsymbol{W}, \vartheta, \boldsymbol{\alpha}) = \tfrac{1}{2}\sum_k W_k^2 - \sum_\mu \alpha_\mu \Big\{\Big(\sum_k W_k\, \chi_k^\mu - \vartheta\Big)\eta_\mu - 1\Big\} \tag{6.18}$$

$\partial L/\partial W_i = 0$:

$$W_k = \sum_\mu \alpha_\mu\, \chi_k^\mu\, \eta_\mu \tag{6.19}$$

$\partial L/\partial \vartheta = 0$:

$$\sum_\mu \alpha_\mu\, \eta_\mu = 0 \tag{6.20}$$

mit

$$\begin{aligned} \alpha_\mu &> 0 \quad \text{falls} \quad \Big(\sum_k W_k\, \chi_k^\mu - \vartheta\Big)\eta_\mu = 1 \\ \alpha_\mu &= 0 \quad \text{sonst} \end{aligned} \tag{6.21}$$

Mit Gl.(6.18):

$$L(\boldsymbol{\alpha}) = -\tfrac{1}{2}{\sum_{\mu,\nu}}' \alpha_\mu\, \eta_\mu\, X_{\mu,\nu} \eta_\nu\, \alpha_\nu + {\sum_{\mu}}' \alpha_\mu \tag{6.22}$$

wobei die Summe nur über die "Support-Vektoren" mit $\alpha_\mu > 0$ läuft.

Bestimmung von α_μ durch Minimisierung von $L(\boldsymbol{\alpha})$ mit Nebenbedingungen $\alpha_\mu \geq 0$ und Gl.(6.20) und

$$X_{\mu\nu} = \sum_k \chi_k^\mu\, \chi_k^\nu \tag{6.23}$$

Mit Vorverarbeitung:

$$\chi_k^\mu = \Psi_k(\boldsymbol{\xi}_\mu) \qquad X_{\mu\nu} = \sum_k \Psi_k(\boldsymbol{\xi}_\mu)\, \Psi_k(\boldsymbol{\xi}_\nu) \qquad W_k = \sum_\mu \alpha_\mu\, \eta_\mu\, \Psi_k(\boldsymbol{\xi}_\mu) \tag{6.24}$$

Support-Vektor Maschine

$$\eta = \Theta\Big(\sum_k W_k\, \Psi_k(\boldsymbol{\xi}) - \vartheta \Big) \tag{6.25}$$

"Entscheidungsfläche": $\sum_k W_k\, \Psi_k(\boldsymbol{\xi}) - \vartheta = 0$

Beispiel: Quadratische Entscheidungsfläche

$$\sum_i W_i\, \xi_i + \tfrac{1}{2} \sum_{i\,j} W_{ij}\, \xi_i\, \xi_j - \vartheta = 0 \tag{6.26}$$

$$\sum_\mu \alpha_\mu\, \eta_\mu \Big\{ \sum_i \xi_i^\mu\, \xi_i + \tfrac{1}{2} \sum_{i\,j} \xi_i^\mu\, \xi_j^\mu\, \xi_i\, \xi_j \Big\} - \vartheta = 0 \tag{6.27}$$

$$X_{\mu\nu} = \sum_i \xi_i^\mu\, \xi_i^\nu + \tfrac{1}{2} \sum_{i\,j} \xi_i^\mu\, \xi_j^\mu\, \xi_i^\nu\, \xi_j^\nu \tag{6.28}$$

N Eingänge $\xi_1 \cdots \xi_N$

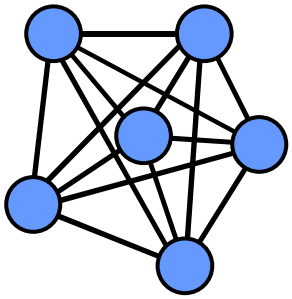
$K = \frac{1}{2}\, N\,(N+3)$ verborgene Einheiten

$D \leq K$ Parameter $\alpha_1 \cdots \alpha_D > 0$

18.12.00

7 Assoziativer Speicher – Attraktor Netzwerk

Assoziativer Speicher

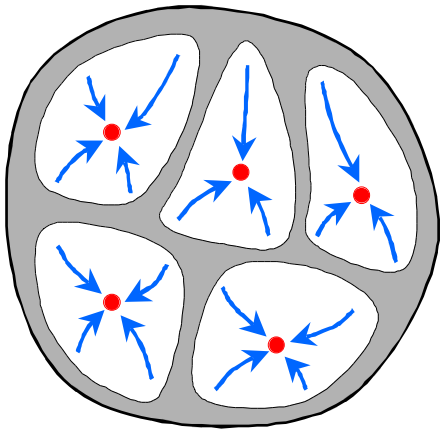


”Adresse”: Teil eines Musters;

”Ausgabe”: Vollständiges Muster

Realisierung durch ein Neuronales Netz mit Rückkopplung
Hopfield Modell (Hopfield 1982)

Dynamisches Neuronales Netz mit Fixpunkt Attraktoren (Mustern)



Dynamik

Gl.(3.5) und Gl.(3.6) mit $\Gamma = 1/C \gg \gamma = 1,\ \lambda = 0$ und Feuerrate $\nu = F(u)$ mit $u = U - \bar{U}$:

$$\partial_t\, u_i(t) = \sum_j W_{ij}\, F\Big(u_j(t)\Big) - u_i(t) \qquad (7.1)$$

Fixpunkt:

$$\begin{aligned} u_i^* &= \sum_j W_{ij} F(u_j^*) \\ \nu_i^* &= F\Big(\sum_j W_{ij}\, \nu_j^*\Big) \end{aligned} \qquad (7.2)$$

Modifizierte Gradientendynamik für $W_{ij} = W_{ji}$: ”Energie”:

$$E(\boldsymbol{u}) = -\tfrac{1}{2} \sum_{ij} F(u_i)\, W_{ij}\, F(u_j) + \sum_i G(u_i) \qquad (7.3)$$

mit $G'(u) = u\, F'(u)$:

$$\begin{aligned} \partial_t\, E(\boldsymbol{u}(t)) &= -\sum_i F'(u_i(t)) \Big\{ \sum_j W_{ij}\, F(u_j(t)) - u_i(t) \Big\}\, \partial_t u_i(t) \\ &= -\sum_i F'(u_i(t)) \Big(\partial_t u_i(t)\Big)^2 \\ &\le 0 \qquad \text{für} \qquad F'(u) > 0 \end{aligned} \qquad (7.4)$$

Fixpunkte sind Minima von $E(\boldsymbol{u})$. Die einzigen Attraktoren sind Fixpunkte.

<u>Beispiel:</u>

$$\nu = F(u) = \tanh(\beta u) \qquad (7.5)$$

$$E(\boldsymbol{\nu}) = -\frac{1}{2} \sum_{ij} \nu_i\, W_{ij}\, \nu_j + \frac{1}{\beta} \sum_i \Big[\nu_i\, \text{Artanh}\, \nu_i - \tfrac{1}{2}\, \ln(1 - \nu_i^2)\Big] \qquad (7.6)$$

$$E(\boldsymbol{\nu}) \underset{\beta\to\infty}{\longrightarrow} -\frac{1}{2} \sum_{ij} \nu_i\, W_{ij}\, \nu_j \qquad |\nu_i| \le 1 \qquad (7.7)$$

Hopfield Modell

HopfieldJ.J.(1982): *NeuralNetworks andPhysical SystemswithEmergent CollectiveComputational Abilities*,Proc. Nat. Acad. Sci. U.S.A.792554.

Zufallsmuster $\xi_i^\mu = \pm 1$ mit $i = 1 \cdots N, \quad \mu = 1 \cdots A$ und $\langle \xi_i^\mu \, \xi_j^\nu \rangle = \delta_{ij} \, \delta_{\mu\nu}, \quad \alpha = A/N$ und Hebb-Lernen

$$W_{ij} = W_{ji} = \frac{1}{N} \sum_\mu \xi_i^\mu \, \xi_j^\mu - \alpha \, \delta_{ij} \qquad\qquad W_{ii} = 0 \tag{7.8}$$

Überlapp

$$m_\mu = \frac{1}{N} \sum_i \xi_i^\mu \, \nu_i \tag{7.9}$$

Zeitabhängigkeit mit Gl.(7.1) und

$$u_i = F^{-1}(\nu_i) \tag{7.10}$$

$$\begin{aligned} \partial_t \, \nu_i &= F'\Big(F^{-1}(\nu_i)\Big)\Big\{ \sum_j W_{ij} \nu_j - F^{-1}(\nu_i)\Big\} \\ &\approx F\Big(\sum_j W_{ij} \, \nu_j\Big) - \nu_i \qquad \text{für} \quad u_i \approx \sum_j W_{ij} \nu_j \end{aligned} \tag{7.11}$$

Überlappdynamik mit Gl.(7.8)

$$\partial_t \, m_\mu(t) = \frac{1}{N} \sum_i \xi_i^\mu \, F\Big(\sum_\nu \xi_i^\nu \, m_\nu - \alpha \nu_i\Big) - m_\mu \tag{7.12}$$

Hopfield Modell für gerige Zahl gespeicherter Muster: $N \to \infty \quad \alpha \to 0$

Amit D.J.(1989): *ModelingBrainFunction–TheWorldofAttractor Neural Networks*,CambridgeUniversityPress: Cambridge.

$$\partial_t \, m_\mu(t) = -\frac{\partial \mathcal{E}(\boldsymbol{m}(t))}{\partial \, m_\mu(t)} \tag{7.13}$$

mit

$$\mathcal{E}(\boldsymbol{m}) = \frac{1}{2} \sum_\mu m_\mu^2 + \frac{1}{N} \sum_i \mathcal{F}\Big(\sum_\nu \xi_i^\nu \, m_\nu\Big) \tag{7.14}$$

und

$$\mathcal{F}'(u) = -F(u) \tag{7.15}$$

8.1.01

Mit Gl.(7.5)

$$\mathcal{F}(u) = -\frac{1}{\beta} \ln \, \cosh(\beta u) \underset{\beta \to \infty}{\longrightarrow} -|u| \tag{7.16}$$

Dynamik für $\beta \to \infty$:

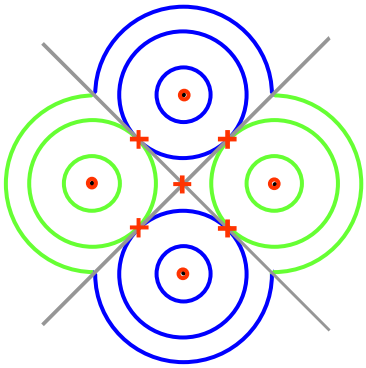
$$\partial_t \, m_\mu = \frac{1}{N} \sum_i \text{sign}\Big(\sum_\nu \xi_i^\mu \, \xi_i^\nu \, m_\nu\Big) - m_\mu \tag{7.17}$$

Fixpunkte:

$$m_\mu^* = \frac{1}{N} \sum_i \text{sign}\Big(\sum_\nu \xi_i^\mu \, \xi_i^\nu \, m_\nu^*\Big) \tag{7.18}$$

Stabil falls

$$\sum_{\nu} \xi_i^{\mu}\, \xi_i^{\nu}\, m_{\nu}^{*} \neq 0 \qquad \text{für alle } i \qquad (7.19)$$

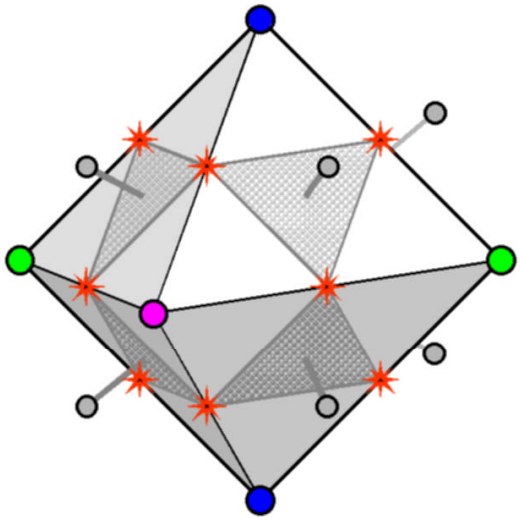


Reines Muster "κ":

$$m_{\kappa} = 1 \qquad m_{\mu} = 0 \quad \text{für } \mu \neq \kappa \qquad (7.20)$$

Mischzustand "1,2,3" e.t.z.:

$$\begin{aligned} m_1 &= \pm m_2 = \pm m_3 = \pm\tfrac{1}{2} \\ m_{\mu} &= 0 \quad \text{für } \mu \geq 4 \end{aligned} \qquad (7.21)$$



Ungerade Mischzustände höherer Ordnung.

Instabile Fixpunkte (mit $\text{sign}(0) = 0$):

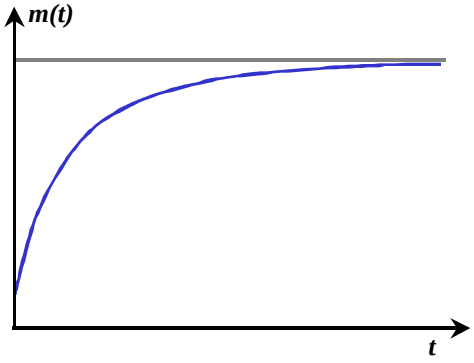
Symmetrischer Mischzustand "1,2" e.t.z.:

$$\begin{aligned} m_1 &= \pm m_2 = \pm\tfrac{1}{2} \\ m_{\mu} &= 0 \quad \text{für } \mu \geq 3 \end{aligned} \qquad (7.22)$$

Gerade Mischzustände höherer Ordnung.

<u>Dynamik für reinen Musterzustand</u>

$$\begin{aligned} \partial_t\, m_1(t) &= 1 - m_1(t) \\ m_1(t) &= 1 - \big(1 - m_1(0)\big)\, \mathrm{e}^{-t} \end{aligned} \qquad (7.23)$$



<u>Dynamik eines Hopfield Modell mit verdünnten Kopplungen:</u>

DerridaB.,GardnerE.,andZippeliusA.(1987): *AnExactlySolubleAsymmetricNeural NetworkModel* ,Europhys.Lett4167.

Kopplungen

$$W_{ij} = \frac{C_{ij}}{c\,N} \sum_{\mu} \xi_i^{\mu}\, \xi_j^{\mu} \qquad W_{ii} = 0 \qquad (7.24)$$

$$C_{ij} = \{1, 0\} \qquad P(C_{ij} = 1) = c \qquad P(C_{ij} = 0) = 1 - c \qquad (7.25)$$

mit C_{ij} und C_{ij} statistisch unabhängig, $c \ll 1$ und $N \to \infty$.

Dynamik für Mittelwerte, gemittelt über ξ_i^{μ} und C_{ij}

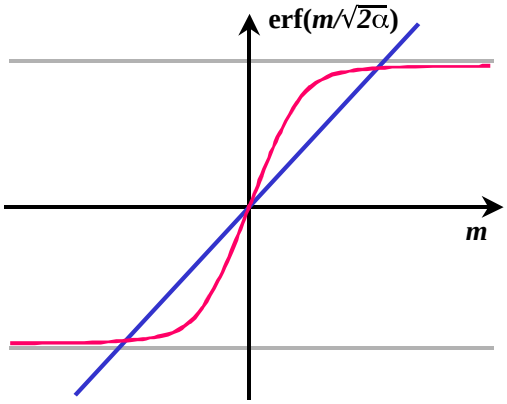
$$\partial_t\, m_1(t) = \sum_i \Big\langle \xi_i^{(1)}\, F\big(\sum_j W_{ij}\, \nu_j(t)\big) \Big\rangle - m_1(t) \qquad (7.26)$$

Mit Gl.(7.9), $\nu_i(t)^2 = 1$ und $\alpha = A/cN$

$$\Big\langle \sum_j W_{ij}\,\nu_j(t)\Big\rangle = \xi_i^{(1)}\,m_1(t) \qquad \Big\langle \Big(\sum_j W_{ij}\,\nu_j(t)\Big)^2\Big\rangle = m_1(t)^2 + \alpha \tag{7.27}$$

Für $\beta \to \infty$

$$\partial_t\, m_1(t) = \mathrm{erf}\Big(\frac{m_1(t)}{\sqrt{2\alpha}}\Big) - m_1(t) \tag{7.28}$$

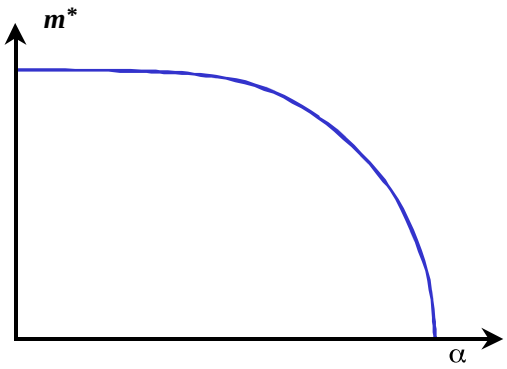


mit

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x \mathrm{d}x\,\mathrm{e}^{-x^2} \tag{7.29}$$

Trivialer Fixpunkt $m_i^* = 0$ instabil für

$$\alpha < \frac{2}{\pi} \tag{7.30}$$



Nichttrivialer Fixpunkt (stabil)

$$m_1^* = \mathrm{erf}\Big(\frac{m_1^*}{\sqrt{2\alpha}}\Big) \tag{7.31}$$

Attraktionsbereich $m_1(0) > 0$

15.1.01

Dynamik des voll vernetzten Hopfield Modells

Horner H.,BormannD.,FrickM.,KinzelbachH.,andSchmidt A.(1989): *TransientsandBasins of AttractioninNeural NetworkModels* ,Z.Phys.B76,381.

Zeitabhängigkeit für $m_\mu(t)$ mit $\mu \neq 1$: Gl.(7.12)

$$\begin{aligned}\partial_t\, m_\mu(t) &= \frac{1}{N}\sum_i \xi_i^\mu\, F\Big(\xi_i^{(1)}\,m_1(t) + \xi_i^\mu\,m_\mu(t) + \frac{1}{N}\sum_{\kappa\neq\mu}\xi_i^\kappa \sum_{j\neq i}\xi_i^\kappa\,\nu_i(t)\Big) - m_\mu(t)\\ &\approx \Big\{\frac{1}{N}\sum_i F'\Big(\xi_i^{(1)}\,m_1(t) + \frac{1}{N}\sum_{\kappa\neq\mu}\xi_i^\kappa\sum_{j\neq i}\xi_i^\kappa\,\nu_i(t)\Big) - 1\Big\}\,m_\mu(t)\end{aligned} \tag{7.32}$$

Beachte: $F'(\cdots)$ hängt nicht von $\xi_{\cdots}^\mu$ ab.

Näherung: Vernachlässigung der Zeitabhängigkeit in $F'(\cdots)$

Mit

$$\partial_t\,\hat m_\mu(t) = -\hat m_\mu(t) \tag{7.33}$$

$$m_\mu(t) = \hat m_\mu(t) + \int_0^t \mathrm{d}s\,\mathrm{e}^{(F'-1)\,(t-s)}\,F'\,\hat m_\mu(s) \tag{7.34}$$

Beachte: $\hat m_\mu(t) = \frac{1}{N}\sum_j \xi_j^\mu\,\hat\nu_j(t)$ ist linear in $\xi_{\cdots}^\mu$, wobei $\hat\nu_j(t)$ nicht von $\xi_{\cdots}^\mu$ abhängt.

Gl.(7.11) und Gl.(7.12) gelten für Mittelwerte. Die tatsächlichen Neuronenzustände im Hopfield

Modell sind $\nu_i(t) = \pm 1$ oder $\langle \nu_i(t)\, \nu_i(t) \rangle = 1$.

Berechnung von $\langle m_\mu(t)\, m_\mu(t) \rangle$ mit Näherung $\langle \nu_i(t)\, \nu_i(t') \rangle \approx 1$:

$$\langle m_\mu^2(t) \rangle = \frac{1}{N}\Big(1 + \int_0^t \mathrm{d}s\, \mathrm{e}^{(F'-1)(t-s)} F'\Big)^2 = \frac{1}{N}\Big(1 + g(t)\Big)^2 = \frac{1}{N} D(t) \tag{7.35}$$

mit $g(0) = 0$ und

$$\partial_t g(t) = F'(t) - \Big(1 - F'(t)\Big)\, g(t) \tag{7.36}$$

Zeitabhängigkeit für $m_1(t)$ ist durch Gl.(7.26) gegeben, wobei $\langle \cdots \rangle$ Gaussmittelung mit Breite αD meint.

Für $\beta \to \infty$

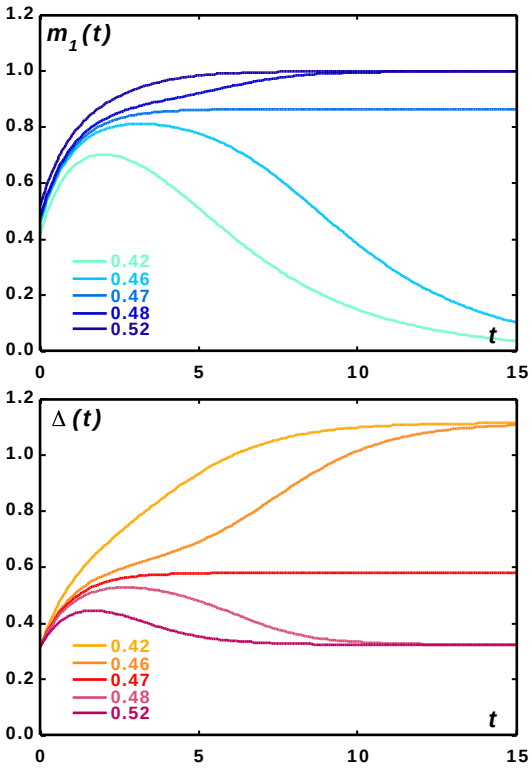
$$\partial_t\, m_1(t) = \mathrm{erf}\Big(\frac{m_1(t)}{\sqrt{2\alpha\, D(t)}}\Big) - m_1(t) \tag{7.37}$$

und

$$F'(t) = \sqrt{\frac{2}{\pi\, \alpha\, D(t)}}\; \mathrm{e}^{\frac{m_1(t)^2}{2\alpha D}} \tag{7.38}$$

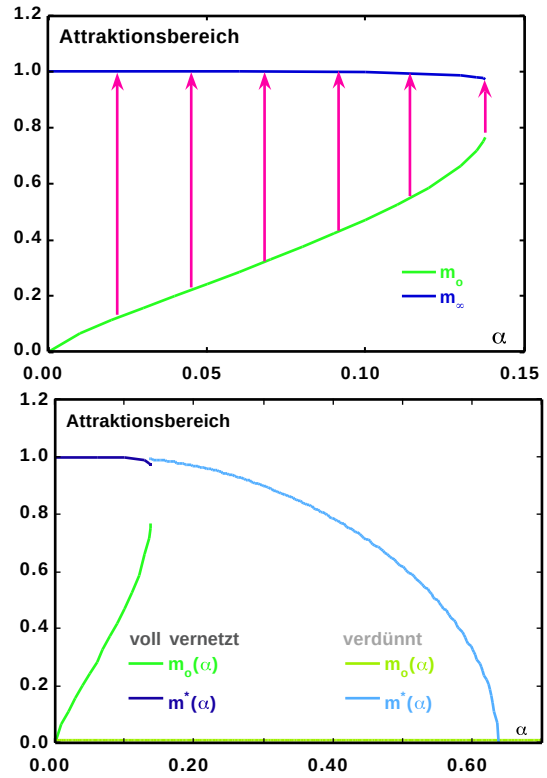
Numerische Integration von Gl.(7.35 – 7.38):

$\alpha = 0.1$: $m(t)$ und $\Delta(t) = \sqrt{\alpha D(t)}$



Konkurrenz zwischen Verstärkung des Musters und des Rauschens

Attraktionsbereich,
Vergleich mit verdünntem Netzwerk



$\alpha = A/c\,N$: $c\,N$ Synapsen pro Neuron.

Voll vernetzt: $\alpha_c = 0.138$
verdünnt: $\alpha_c = 0.637$

22.1.01

Assoziativer Speicher mit geringer Aktivität

Horner H.(1989): *Neural Networks withlow levels of activity: Isingvs. McCulloch–Pitts neurons*,Z.Phys. B75,133

Repräsentation: aktiv $\xi = 1$ inaktiv $\xi = -a/(1-a)$

Aktivität: $P(\xi_i^\mu = 1) = a \qquad P(\xi_i^\mu = -a/(1-a)) = 1-a$

Hebb'sches Lernen, Gl.(7.8):

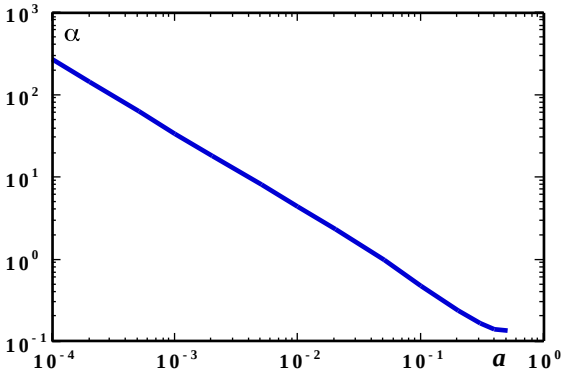
$$W_{ij} = \frac{1-a}{a\,N} \sum_\mu \xi_i^\mu\, \xi_j^\mu \quad \text{für} \quad i \neq j \qquad\qquad W_{ii} = 0 \tag{7.39}$$

Beachte: für $a \to 0$ werden nur Kopplungen zwischen gleichzeitig aktiven Neuronen verstärkt.

Entsprechende Rechnung:

Maximale Speicherkapazität für $a \to 0$

$$\alpha_c = \frac{1 - \sqrt{\frac{\ln \ln 1/a}{\ln 1/a}}}{2\,a \ln 1/a} \tag{7.40}$$

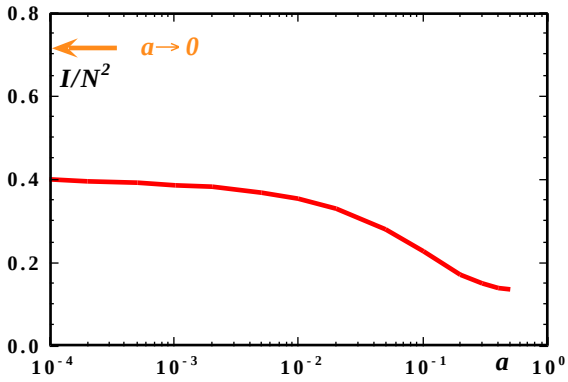


Information pro Muster und pro Neuron

$$i = -a\,\log_2(a) - (1-a)\,\log_2(1-a) \tag{7.41}$$

Information

$$I/N^2 = \alpha i \underset{a\to 0}{\longrightarrow} \frac{1}{2\ln 2} \approx 0.72 \tag{7.42}$$



Hopfield Modell mit $a = \frac{1}{2}$: $I/N^2 \approx 0.138$

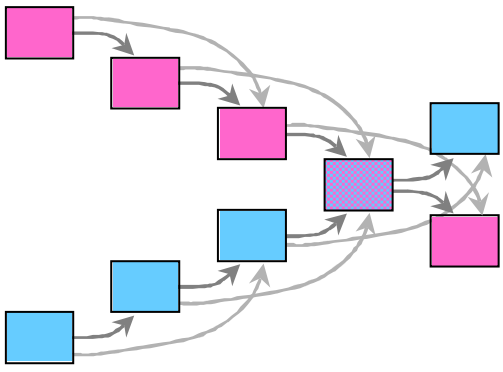
Assoziativer Speicher für Mustersequenzen

HerzA.,Sulzer B.,KuhnR.,andvanHemmenJ.L.(1989): *HebbianLearningReconsidered: RepresentationofStaticandDynamicObjectsin AssociativeNeural Nets*,Biol. Cybernetics60457.

Muster $\xi_i^\mu(t_n)$

Retardierte Kopplungen $W_{ij}^{(m)}$:
Erregung

$$U_i(t_n) = \sum_j \sum_m W_{ij}^{(m)}\, \nu_j(t_n - \tau_m) \tag{7.43}$$



Hebb'sches Lernen:

$$\delta W_{ij}^{(m)} = \Delta_m \sum_n \sum_\mu \xi_i^\mu(t_n)\, \xi_j^\mu(t_n - \tau_m) \tag{7.44}$$

8 Nicht überwachtes Lernen

Hauptachsen einer Verteilung, Oja's Regel

Hebb'sches Lernen mit linearer Ausgabeeinheit

Signale $\boldsymbol{\xi}_\mu$ mit Verteilung $P(\boldsymbol{\xi})$

Lernregel:

$$\delta W_i = \Delta \left(U \xi_i - \gamma W_i \right) \qquad U = \sum_j W_j \, \xi_j \tag{8.1}$$

mit γ so daß $\sum_i W_i^2 = 1$: $\quad \gamma = U^2$

$$\delta \sum_i W_i^2 = 2 \, \Delta \, U_\mu^2 \left(1 - \sum_i W_i^2 \right) \tag{8.2}$$

Korrelationsmatrix:

$$C_{ij} = \sum_{\boldsymbol{\xi}} P(\boldsymbol{\xi}) \, \xi_i \, \xi_j = \sum_\lambda C_\lambda \, \xi_i^\lambda \, \xi_j^\lambda \qquad \sum_i \xi_i^\lambda \, \xi_i^\kappa = \delta_{\lambda\kappa} \tag{8.3}$$

mit $C_\lambda \geq 0$ und $\lambda = 1 \cdots N$.

Es sei

$$W_\lambda = \sum_i W_i \, \xi_i^\lambda \qquad W_i = \sum_\lambda W_\lambda \, \xi_i^\lambda \qquad \gamma = \sum_\lambda C_\lambda \, W_\lambda^2 \tag{8.4}$$

Lernen:

$$\delta W_\lambda = \Delta \sum_\lambda \left(C_\lambda - \gamma \right) W_\lambda \tag{8.5}$$

Stabile stationäre Lösung: Maximaler Eigenwert $C_{\bar{\lambda}}$

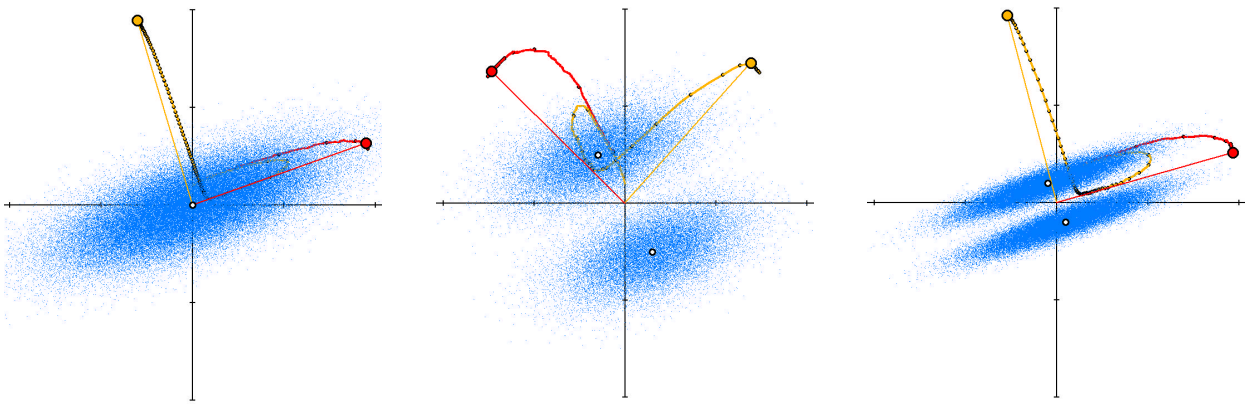
$$W_\lambda = \delta_{\lambda\bar{\lambda}} \qquad \gamma = C_{\bar{\lambda}} \tag{8.6}$$

Verallgemeinerung für K größte Hauptachsen: K Ausgabeeinheiten $k = 1 \cdots K$

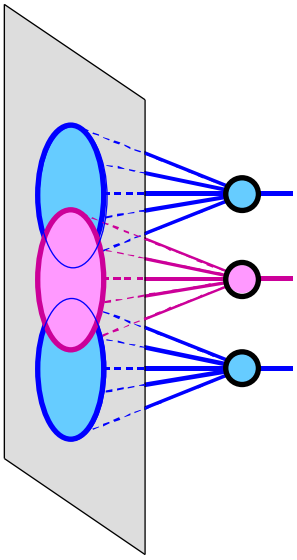
$$\delta W_{ik} = \Delta U_k \Big(\xi_i - \sum_{\ell=1}^{k} U_\ell \, W_{i\ell} \Big) \qquad U_k = \sum_j W_{jk} \xi_j \tag{8.7}$$

29.1.01

Beispiele: **Größte Hauptachse** **Zweite Hauptachse**



Lernen mit korrelierten Signalen, Rezeptive Felder



Eingabe-Ebene (-Raum): $\boldsymbol{x}$

Rezeptives Feld eine Neurons:
$\boldsymbol{x}$ innerhalb: $\Theta_k(\boldsymbol{x}) = 1$ $\boldsymbol{x}$ außerhalb: $\Theta_k(\boldsymbol{x}) = 0$

Erregung bei Signal $\xi(\boldsymbol{x})$

$$U_k = \sum_{\boldsymbol{x}} W_k(\boldsymbol{x})\, \xi(\boldsymbol{x}) \tag{8.8}$$

Normierung

$$n = \sum_{\boldsymbol{x}} W_k(\boldsymbol{x})^2 \to 1 \tag{8.9}$$

Nebenbedingung

$$a = \sum_{\boldsymbol{x}} \alpha(\boldsymbol{x})\, W_k(\boldsymbol{x}) \to \bar{a} \qquad \text{mit} \quad \sum_{\boldsymbol{x}} \alpha(\boldsymbol{x})^2 = 1 \tag{8.10}$$

Lernregel: (Index k unterdrückt)

$$\delta\, W(\boldsymbol{x}) = \Delta\, \Theta(\boldsymbol{x}) \left\{ U\, \xi(\boldsymbol{x}) - \gamma\, W(\boldsymbol{x}) - \mu\, \alpha(\boldsymbol{x}) \right\} \tag{8.11}$$

$$\delta n = 2\Delta \left\{ U^2 - \gamma\, n - \mu\, a \right\} \qquad \gamma = U^2 + \mu\, \bar{a} \tag{8.12}$$

$$\delta a = \Delta \left\{ U \sum_{\boldsymbol{x}} \alpha(\boldsymbol{x})\, \xi(\boldsymbol{x}) - \gamma\, a - \mu \right\} \qquad \mu\,(1 + \bar{a}^2) = U \sum_{\boldsymbol{x}} \alpha(\boldsymbol{x})\, \xi(\boldsymbol{x}) - U^2\, \bar{a} \tag{8.13}$$

Mittelwerte:

Korrelationsfunktion

$$C(\boldsymbol{x}\, \boldsymbol{y}) = \Theta(\boldsymbol{x})\, \Theta(\boldsymbol{y}) \left\langle \xi(\boldsymbol{x})\, \xi(\boldsymbol{y}) \right\rangle = \sum_{\lambda} C_\lambda\, \eta_\lambda(\boldsymbol{x})\, \eta_\lambda(\boldsymbol{y}) \tag{8.14}$$

mit

$$\sum_{\boldsymbol{x}} \eta_\kappa(\boldsymbol{x})\, \eta_\lambda(\boldsymbol{x}) = \delta_{\kappa\lambda} \qquad \sum_{\lambda} \eta_\lambda(\boldsymbol{x})\, \eta_\lambda(\boldsymbol{y}) = \delta(\boldsymbol{x} - \boldsymbol{y}). \tag{8.15}$$

Mit

$$W_\lambda = \sum_{\boldsymbol{x}} \eta_\lambda(\boldsymbol{x})\, W(\boldsymbol{x}) \qquad W(\boldsymbol{x}) = \sum_{\lambda} W_\lambda\, \eta_\lambda(\boldsymbol{x}) \tag{8.16}$$

und

$$\alpha_\lambda = \sum_{\boldsymbol{x}} \eta_\lambda(\boldsymbol{x})\, \alpha(\boldsymbol{x}) \tag{8.17}$$

Lernregel:

$$\delta W_\lambda = \Delta \left\{ \left(C_\lambda - \gamma \right) W_\lambda - \mu\, \alpha_\lambda \right\} \tag{8.18}$$

$$\gamma = \sum_{\lambda} C_\lambda\, W_\lambda^2 \qquad (1 + \bar{a}^2)\mu = \sum_{\lambda} C_\lambda\, W_\lambda\, (\alpha_\lambda - W_\lambda) \tag{8.19}$$

Beispiel: Kreisförmiges Feld $\Theta(\boldsymbol{x}) = \Theta(d - |\boldsymbol{x}|)$

Radiale- und Winkelquantenzahl $\lambda = \{n, m\}$

$$\eta_{nm}(\boldsymbol{x}) = u_{nm}(r)\, \mathrm{e}^{im\varphi} \tag{8.20}$$

Größter Eigenwert: $\lambda = \{0, 0\}$ $\eta_{00}(\boldsymbol{x}) \approx$ const

Ohne Nebenbedingung: $\mu = 0$

$$W(\boldsymbol{x}) \to \eta_{00}(\boldsymbol{x}) \tag{8.21}$$

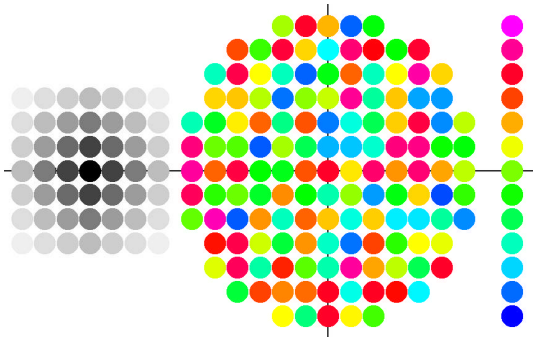
Mit Nebenbedingung

$$\sum_{\boldsymbol{x}} \eta_{00}(\boldsymbol{x})\, W(\boldsymbol{x}) \;=\; 0 \qquad \sum_{\boldsymbol{x}} W(\boldsymbol{x}) \;\approx\; 0 \tag{8.22}$$

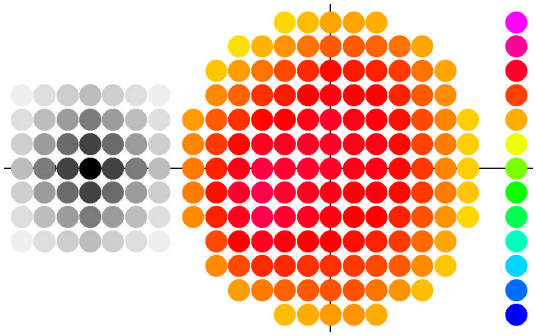
und $m = 0$:

$$W(\boldsymbol{x}) \to \eta_{10}(\boldsymbol{x}) \tag{8.23}$$

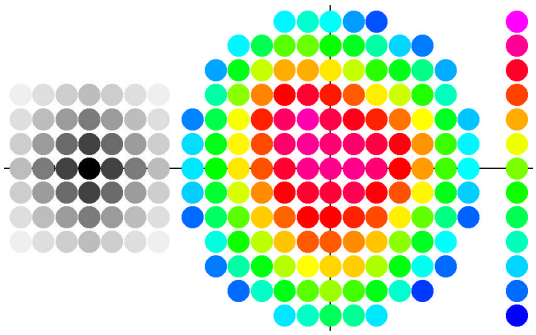
Simulation



Signal zufälliger Anfangszustand

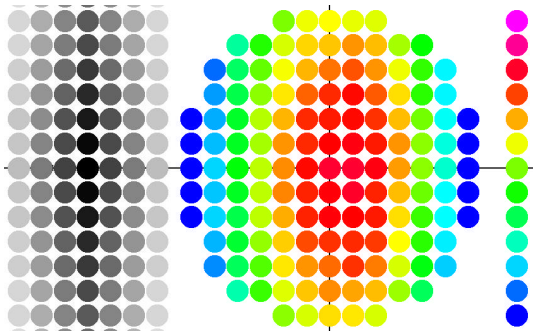


keine Nebenbedingung: $W \to \eta_{00}$



Nebenbedingung: $W \to \eta_{10}$

Zentrum exzitatorisch
Umgebung inhibitorisch



Anisotropes Signal $W \to \eta_{10} + \eta_{02}$

Sensitiv auf Linien bestimmter Orientierung

Retina Kantenverstärkung

Visueller Cortex

Vektorquantisierung, kompetitives Lernen

N-dimensionaler Musterraum mit Mustern ξ_i. Wahrscheinlichkeit $P(\boldsymbol{\xi})$

K Neuronen mit Kopplungen W_{ik}

Aufgabe: Aufteilung des Eingaberaums in K Zellen V_k, so daß

$$\int_{V_k} \mathrm{d}\boldsymbol{\xi}\, P(\boldsymbol{\xi}) = \frac{1}{K} \tag{8.24}$$

und

$$\sum_i (W_{ik} - \xi_i)^2 \quad \text{minimal für } k = k_0 \quad \text{falls} \quad \boldsymbol{\xi} \in V_{k_0} \tag{8.25}$$

Lernregel:
Distanz

$$D_k = \sqrt{\sum_i (W_{ik} - \xi_i)^2} + \vartheta_k \qquad \text{minimal für } k = k_0 \tag{8.26}$$

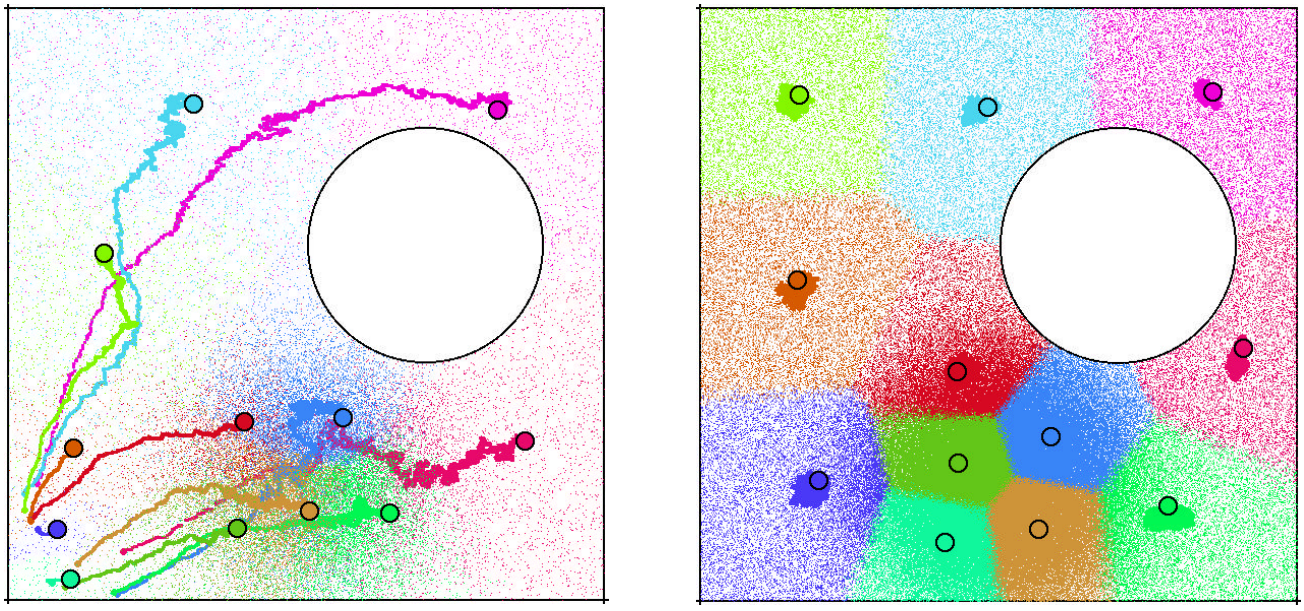
Lernen der Kopplungen $W_{i\,k_0}$

$$\delta W_{i\,k_0} = -\Delta \left\{ W_{i\,k_0} - \xi_i \right\} \tag{8.27}$$

Änderung der Schwelle (Ermüdung) ϑ_k:

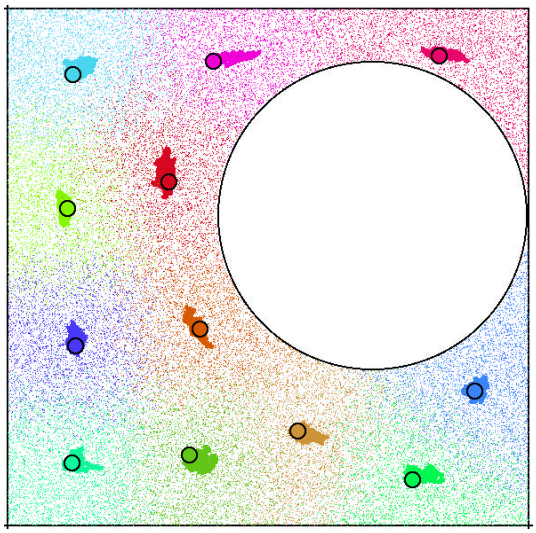
$$\delta\vartheta_{k_0} = (K-1)\Delta_\vartheta \qquad \delta\vartheta_k = -\Delta_\vartheta \quad \text{für} \quad k \neq k_0 \tag{8.28}$$

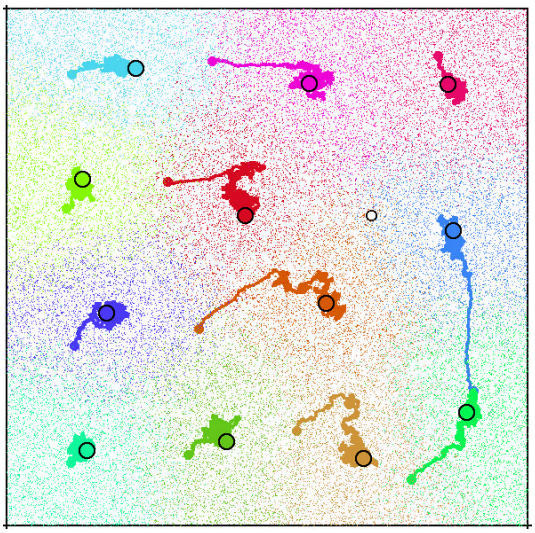
Beispiel: $N = 2$

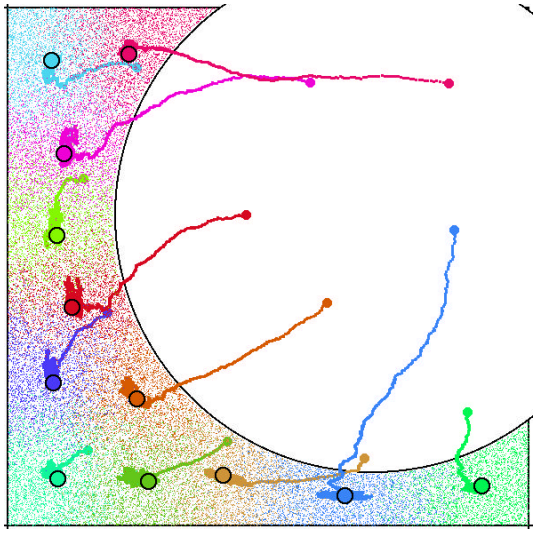


Die Muster $\boldsymbol{\xi}$ sind in der Farbe der zugehörigen $\boldsymbol{W}_{k_0}$ gezeigt.

5.2.01







Reversible Änderung der $\boldsymbol{W}_k$ bei Änderung
des Eingaberaums

<u>Topologische Ordnung</u>

Lernen entsprechend Gl.(8.27), modifiziert:

"Gewinner" k_0

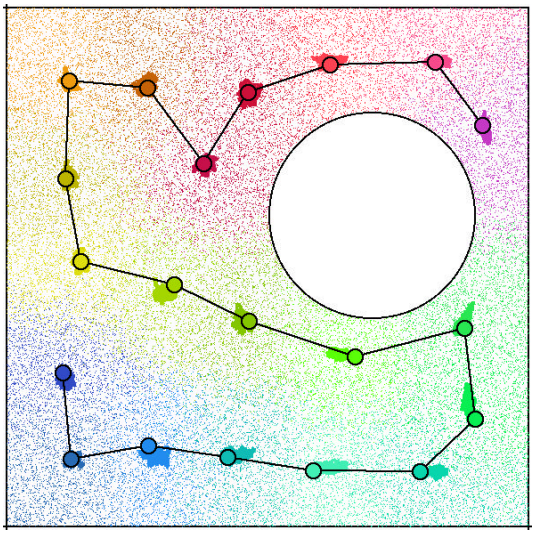
$$\delta W_{i\,k} = -\Delta\,\varphi(k-k_0)\left\{W_{i\,k_0} - \xi_i\right\} \qquad (8.29)$$

mit

$$\varphi(k) = \mathrm{e}^{-(k/d)^2} \qquad (8.30)$$

Im Ausgaberaum benachbarte Neuronen lernen korreliert.

Die Dimensionen des Ein- und Ausgaberaums können verschieden sein



Topologieerhaltende Karten

Korrelierte Muster ξ_i auf der "Retina", i_0 zufällig

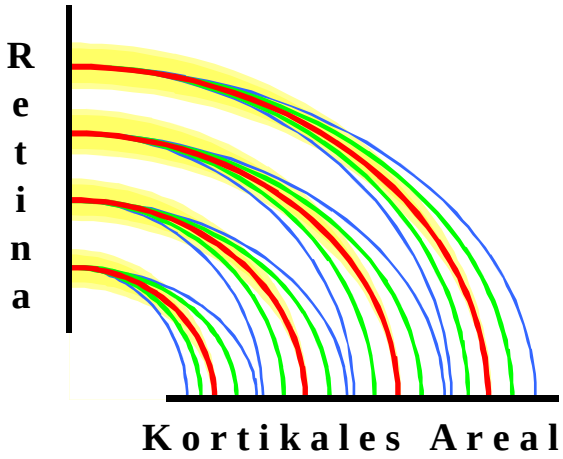
$$\xi_i = \mathrm{e}^{-(i-i_0)^2/d^2} \tag{8.31}$$

Erregung des Neurons K im "kortikalen Areal"

$$U_k = \sum_i \xi_i \, w_{ik} \tag{8.32}$$

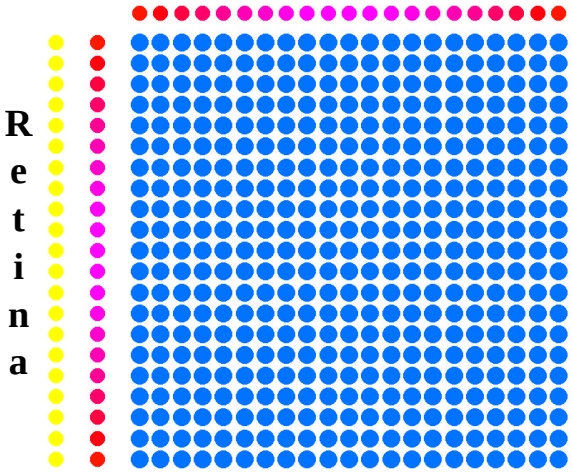
$U_k - \vartheta_k$ maximal für $k = k_0$

Lernen mit φ, Gl.(8.30) $d(t)$ veränderlich

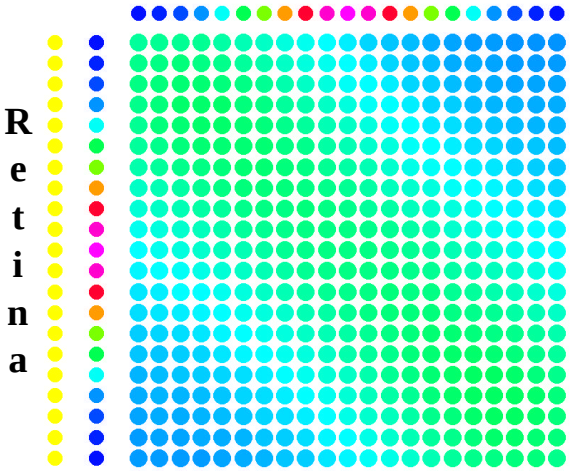


$$\delta W_{ik} = \Delta \, \varphi(k - k_0) \Big\{ \xi_i \, U_{k_0} - U_k \, U_{k_0} \, W_{ik} \Big\} \tag{8.33}$$

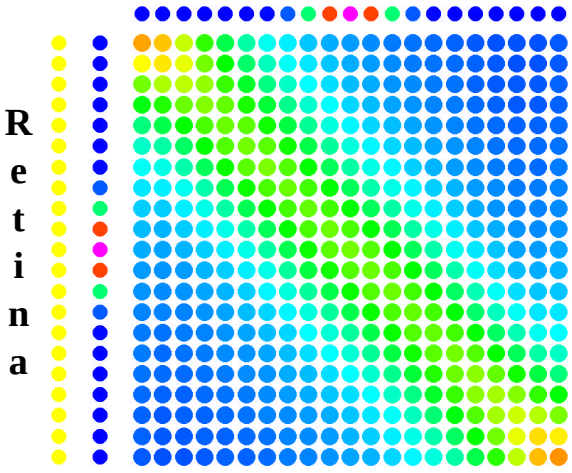
$$\delta \vartheta_{k_0} = K \, \Delta_\vartheta \qquad \vartheta_k = -\Delta_\vartheta \quad \text{für} \quad k \neq k_0 \tag{8.34}$$



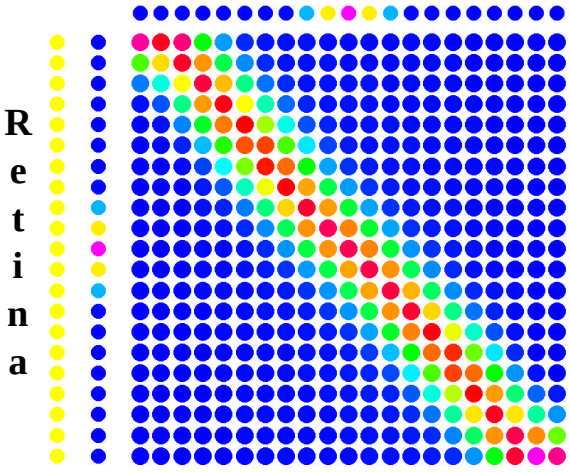
$t = 0 \qquad d = 20$



$t = 50 \qquad d = 5$

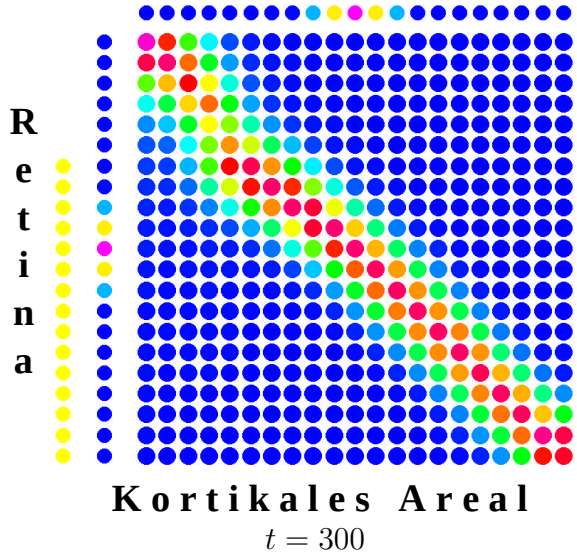


$t = 110 \qquad d = 2$

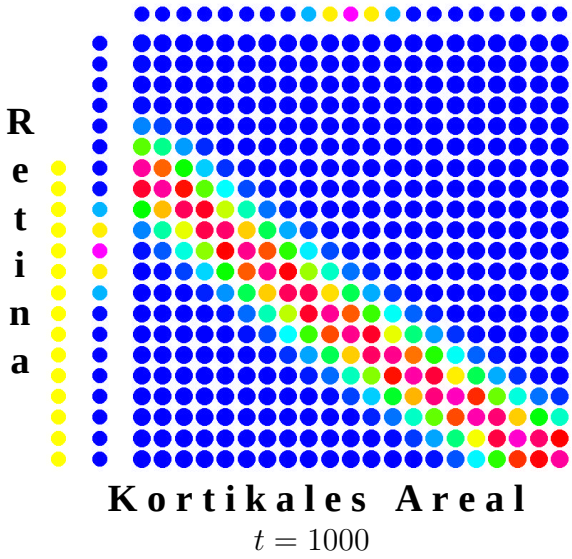


$t = 300 \qquad d = 1.5$

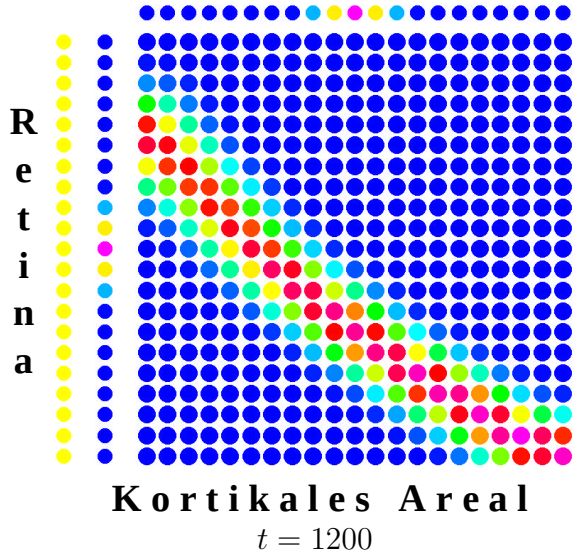
Plastizität bei Änderung des Bereichs der Eingangssignale



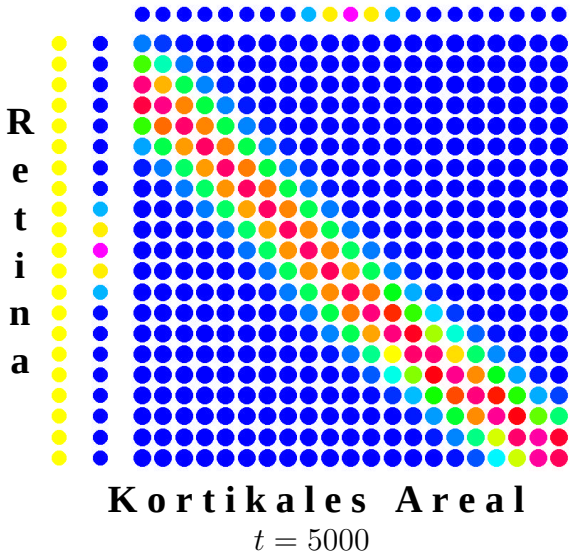
$t = 300$



$t = 1000$

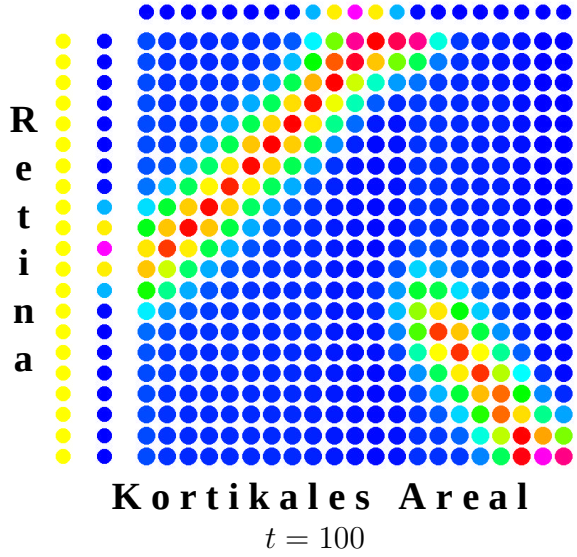


$t = 1200$



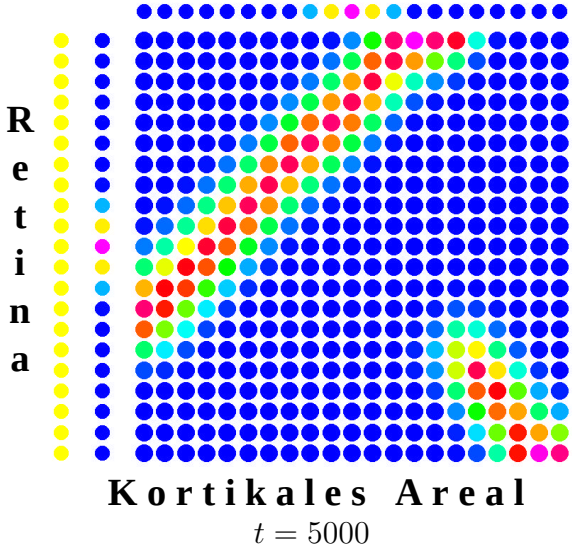
$t = 5000$

Diskontinuität bei zu kleiner Reichweite $d(t=0)$



$t = 100$

Ausheilen der Diskontinuität



$t = 5000$

2.2.01

$\mathcal{FINIS}$