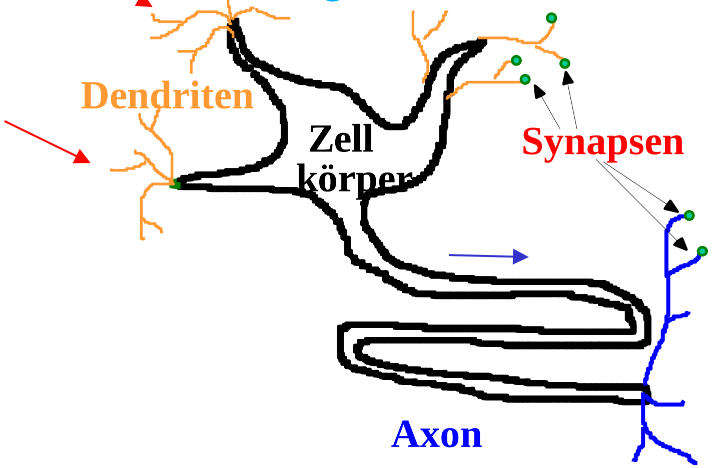
Single-Layer-Netze: Assoziativspeicher und Delta-Regel

Praktikum Adaptive Systeme

Was sind Neuronale Netze ?

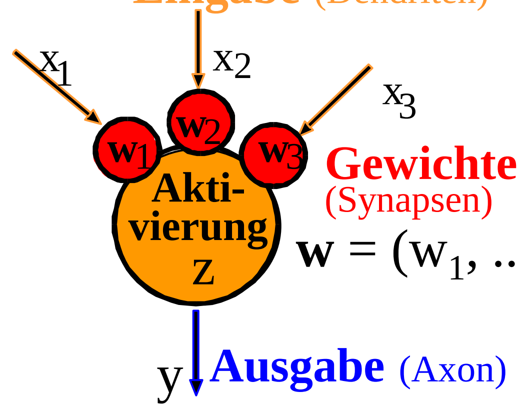
Biologische Neuronen



Abstrakte Neuronen

$\mathbf{x} = (x_1, \ldots, x_n)$

Eingabe (Dendriten)

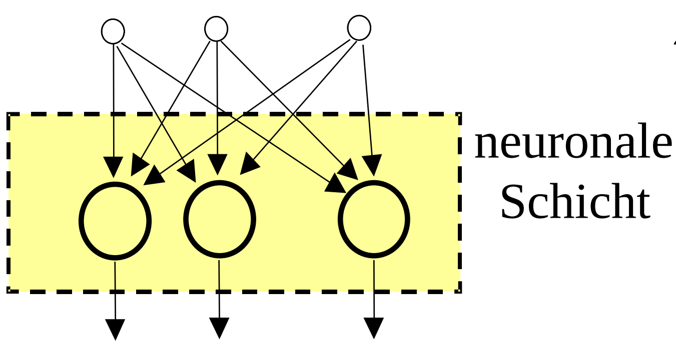


$\mathbf{w} = (w_1, \ldots, w_n)$

$$z = \sum_{i=1}^{n} w_i x_i = \mathbf{w}^T \mathbf{x}$$

$$y = S(z)$$

$\mathbf{x} = (x_1 \quad x_2 \quad \cdots \quad x_n)$



$\mathbf{y} = (y_1 \quad y_2 \quad \cdots \quad y_m)$

Formale Neuronen

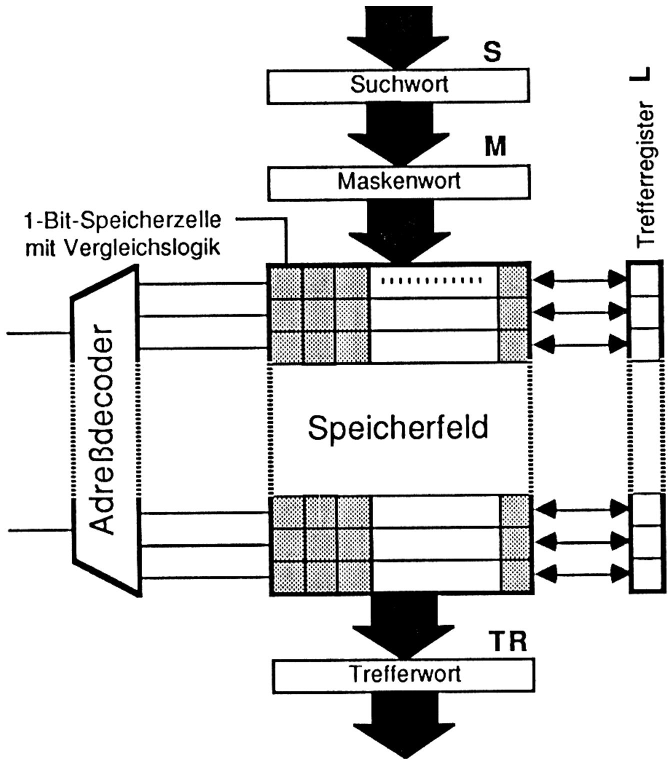
z.B. Lineare Schicht =Matrixmultiplikation

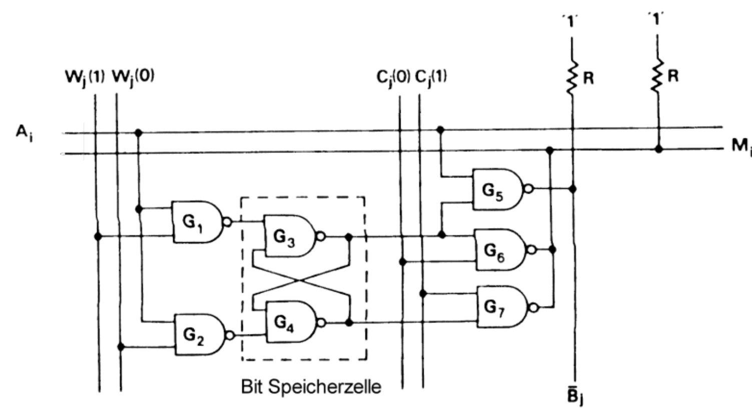
$$\mathbf{y} = \left(\sum_i w_{1i} x_i, \ldots, \sum_i w_{mi} x_i \right)^T$$

$$= \mathbf{W} \cdot \mathbf{x}$$

Lernziel: Gewichte **W** ermitteln

Konventionelle Assoziativspeicher

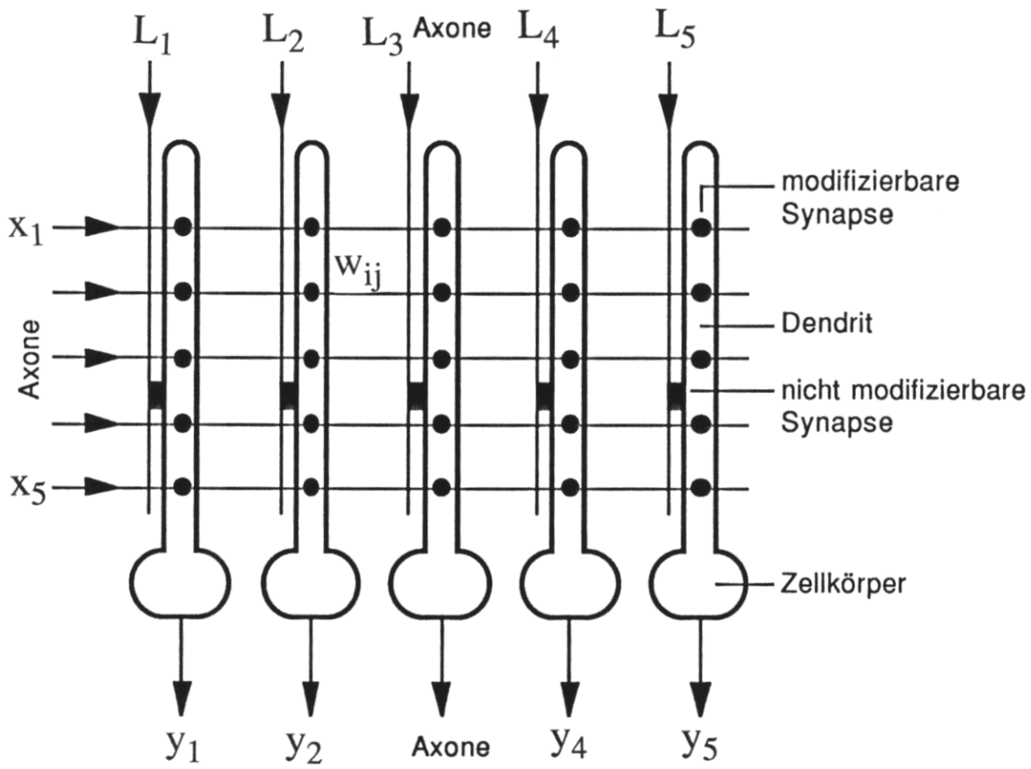




Eingabe: Suchwort, Ausgabe: Treffer in Daten (auch mehrere!)

Problem: Teile des Suchworts unbekannt oder falsch (unbekannte Maske)

Neuro-Modell des Assoziativspeichers



Funktion:

Jede Komp.ist lin. Summe

$$\mathbf{z = Wx}$$

Nichtlin. Ausgabe:

$$y_i = S_B(z_i) = \begin{cases} 1 & z > \theta \\ 0 & z \leq \theta \end{cases}$$

Spalte *i* der Synapsengewichte = $\mathbf{w}_i$

(Alle Spalten $\mathbf{w}_i$) = Matrix **W**

Lernen: Hebbsche Regel

Beobachtung des Physiologen Hebb (1949):

"Wenn ein Axon der Zelle A nahe genug ist, um eine Zelle B zu erregen und wiederholt oder dauerhaft sich am Feuern beteiligt, geschieht ein Wachstumsprozess oder metabolische Änderung in einer oder beiden Zellen dergestalt, dass A's Effizienz, als eine der auf B feuernden Zellen, anwächst."

Also: $w_{AB}(t) - w_{AB}(t\text{-}1) =: \Delta w \sim x_A\, y_B$

oder $\mathbf{w}_i(t) = \mathbf{w}_i(t\text{-}1) + \gamma_i(t)\, y_i \mathbf{x}$ Iterative Hebb'sche Lernregel

$$\mathbf{W}(t) = \mathbf{W}(t\text{-}1) + \gamma(t)\, \mathbf{y}\mathbf{x}^T$$

Lernen im Assoziativspeicher

- **Speichern der Muster L mit Schlüssel x**

$$w_{ij} = \sum_k \Delta w_{ij} = \sum_k \gamma_k L_i^k x_j^k \qquad w_{ij}(0) := 0$$

$y_i\, x_j$ Hebb

- **Abrufen der Muster L mittels Schlüssel $\mathbf{x}^r$**

$$\mathbf{y} = \mathbf{W}\mathbf{x}^r = \mathbf{z} = \gamma_r\, \mathbf{L}^r(\mathbf{x}^r)^T\mathbf{x}^r + \sum_{k \neq r} \gamma_k \mathbf{L}^k \left(\mathbf{x}^k\right)^T \mathbf{x}^r$$

assoziierte Antwort + Übersprechen von anderen Mustern

Orthogonale Muster **x**: Übersprechen = 0, exakte Reproduktion.
Nichtorthogonale Muster: Schwellwerte nötig zum Unterdrücken.

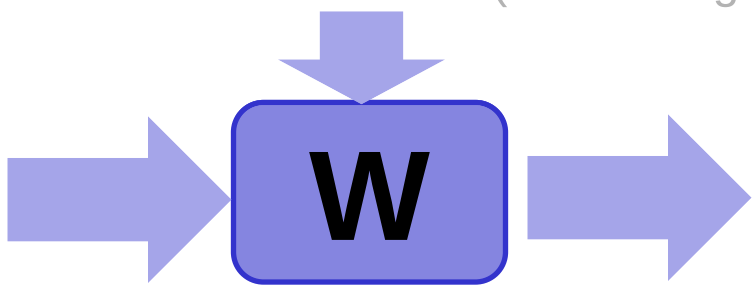
Speicherarten

- **Heteroassoziativer Speicher**

Inhalt **L** (Lehrervorgabe)

Schlüssel = **x**

(„Adresse“)



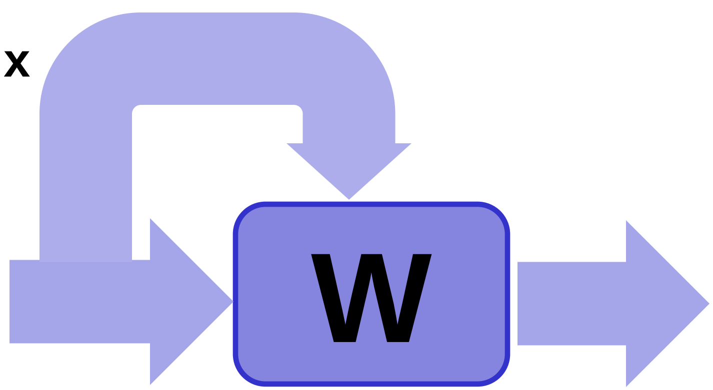
Assoziation

y = **L**

- **Autoassoziativer Speicher**

Inhalt **L** = **x**

Schlüssel = **x**



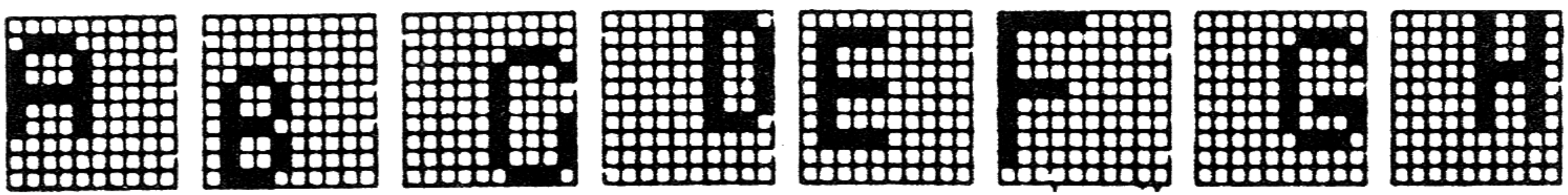
Assoziation

y = **x**

Autoassoziative Ergänzung

Setze **L**(**x**) = **x**, lerne alle Muster (symm. Matrix).

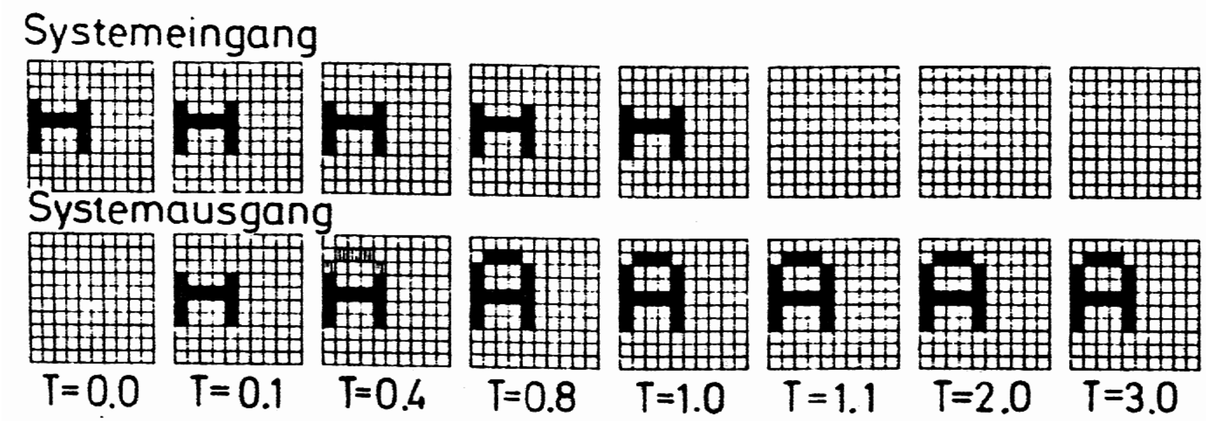
Beispiel: Buchstaben, kodiert mit 0 und 1



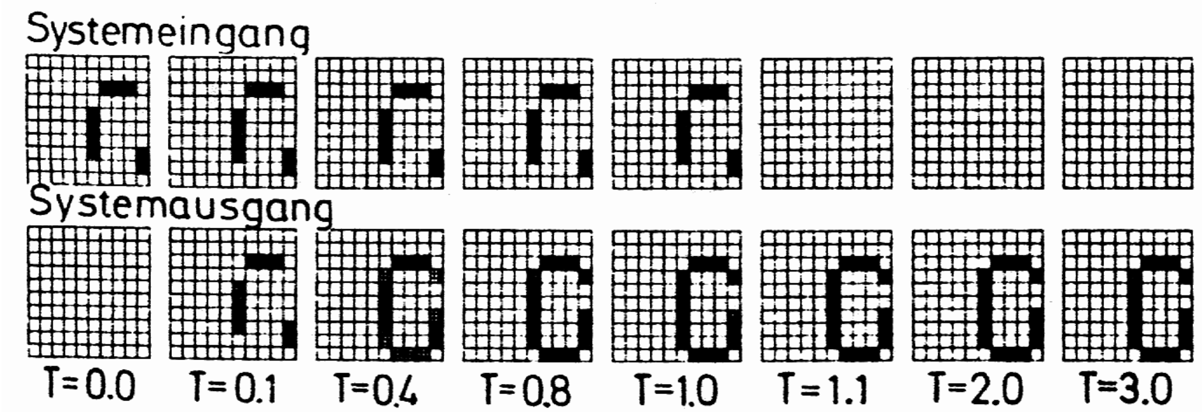
Autoassoziative Ergänzung

Gebe **Teil**muster von x ein ⇒ erhalte **Gesamt**muster x

A



C



Lernen mit Zielfunktionen

- **Lernziel: minimaler quadratischer Fehler**

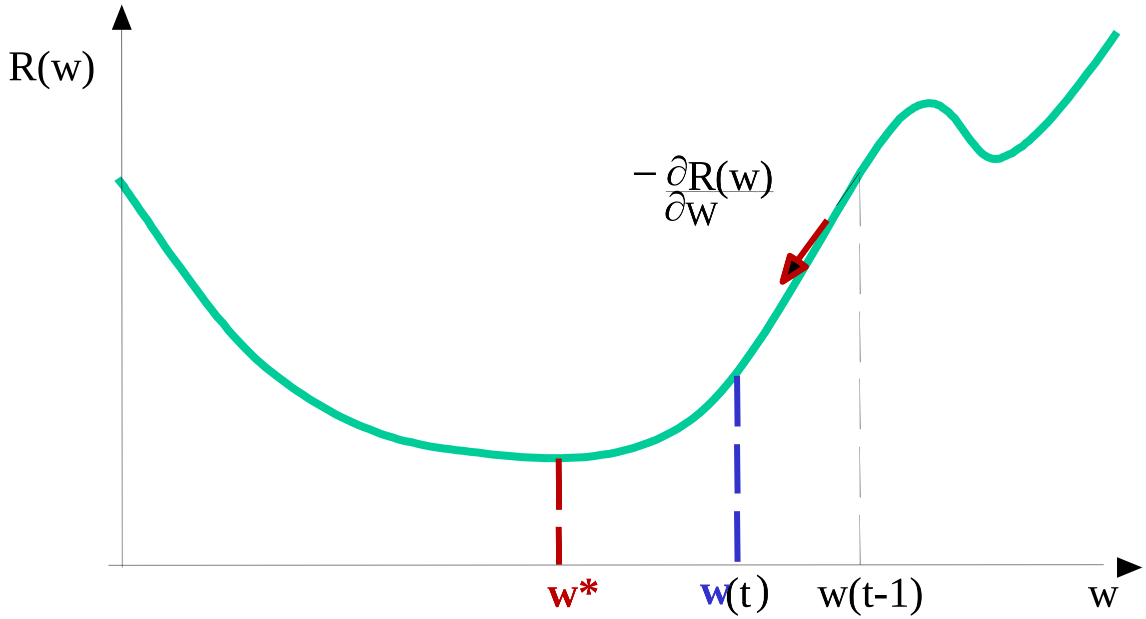
  Zielfunktion R(**w**) = $\langle (y-L)^2 \rangle = min$ mit y = $\mathbf{w}^T\mathbf{x}$, **L** = Lehrervorgabe

  wird erreicht durch schrittweise Approximation zum besten Wert **w***.

WIE ?

Lernen durch Iteration

Gradientenabstieg

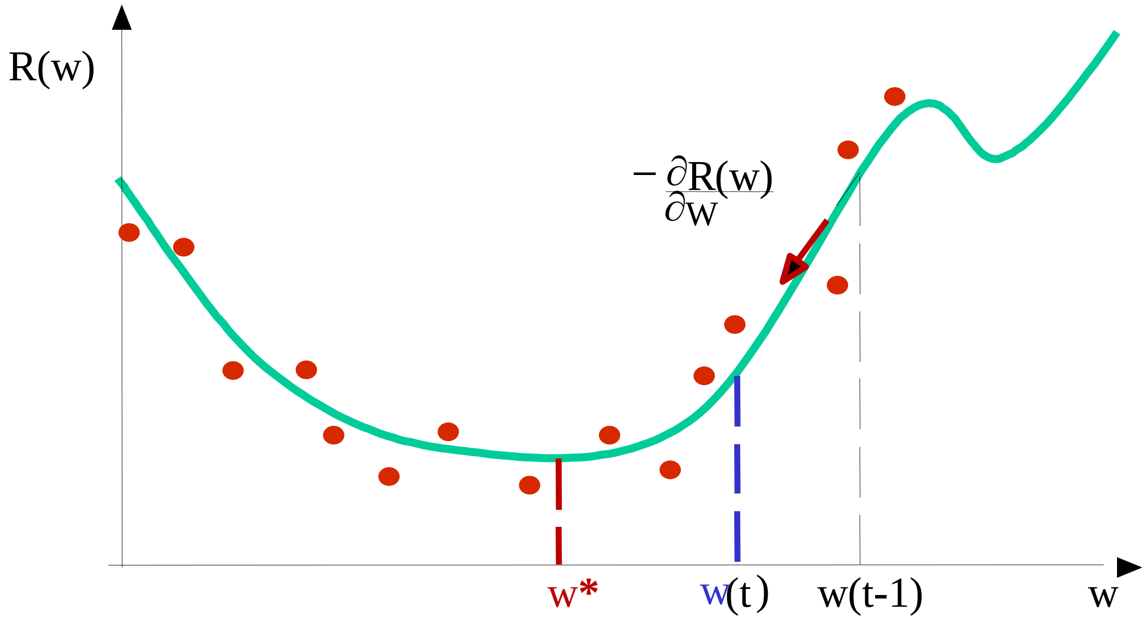


$$\Delta \mathbf{w} := (\mathbf{w}(t\text{-}1) - \mathbf{w}(t)) \sim -\frac{\partial}{\partial w} R(\mathbf{w}(t-1))$$

$$\mathbf{w}(t) = \mathbf{w}(t-1) - \gamma(t) \frac{\partial}{\partial w} R(\mathbf{w}(t-1))$$

Lernen durch Iteration

Problem: stochastischer Gradientenabstieg



$$R(w) = \langle r(w,x) \rangle_{\mathbf{x}}$$

Zielfunktion abhängig von stochastischer Beobachtung x(t) !

Stochastisches Lernen

Lernen mit Zielfunktion $R(w) = \langle r(w,x) \rangle_{\mathbf{x}}$

$$\mathbf{w}(t) = \mathbf{w}(t\text{-}1) - \gamma(t)\, \nabla_{\mathbf{w}}\, R(\mathbf{w}(t\text{-}1))$$

$$\mathbf{w}(t) = \mathbf{w}(t\text{-}1) - \gamma(t)\, \nabla_{\mathbf{w}}\, \langle r(w,x) \rangle_{x}$$

wird ersetzt durch

Lernen mit stochast. Zielfunktion $r(w,x)$

$$\mathbf{w}(t) = \mathbf{w}(t\text{-}1) - \gamma(t)\, \nabla_{\mathbf{w}}\, r(\mathbf{w}(t\text{-}1), \mathbf{x}(t))$$

stochastisches Lernen

Lernen mit Zielfunktionen

- **Lernziel: minimaler quadratischer Fehler**

  Zielfunktion R(**w**) = $\langle (y-L)^2 \rangle = min$ mit y = $\mathbf{w}^T\mathbf{x}$, **L** = Lehrervorgabe

  wird erreicht durch schrittweise Approximation zum besten Wert **w***.

$$\mathbf{w}(t) = \mathbf{w}(t\text{-}1) - \gamma(t)\,\nabla_w\, r(\mathbf{w}(t\text{-}1), \mathbf{x}(t))$$

$$= \mathbf{w}(t\text{-}1) - \gamma(t)\frac{\partial}{\partial w}(y-L)^2$$

$$= \mathbf{w}(t\text{-}1) - \gamma(t)\,2(y\text{-}L)\,\mathbf{x} \qquad \delta := (y\text{-}L)$$

$$\mathbf{w}(t) = \mathbf{w}(t\text{-}1) - \alpha(t)\,\delta\,\mathbf{x}$$

stochast. Approximation

Delta-Regel

Fragen ?