Backprop

$\boxed{2}$

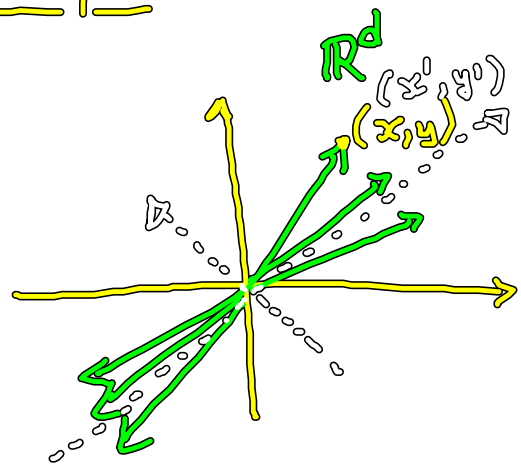
Merkmale

$\vec{x}$

$\vec{x}^{\,i}$

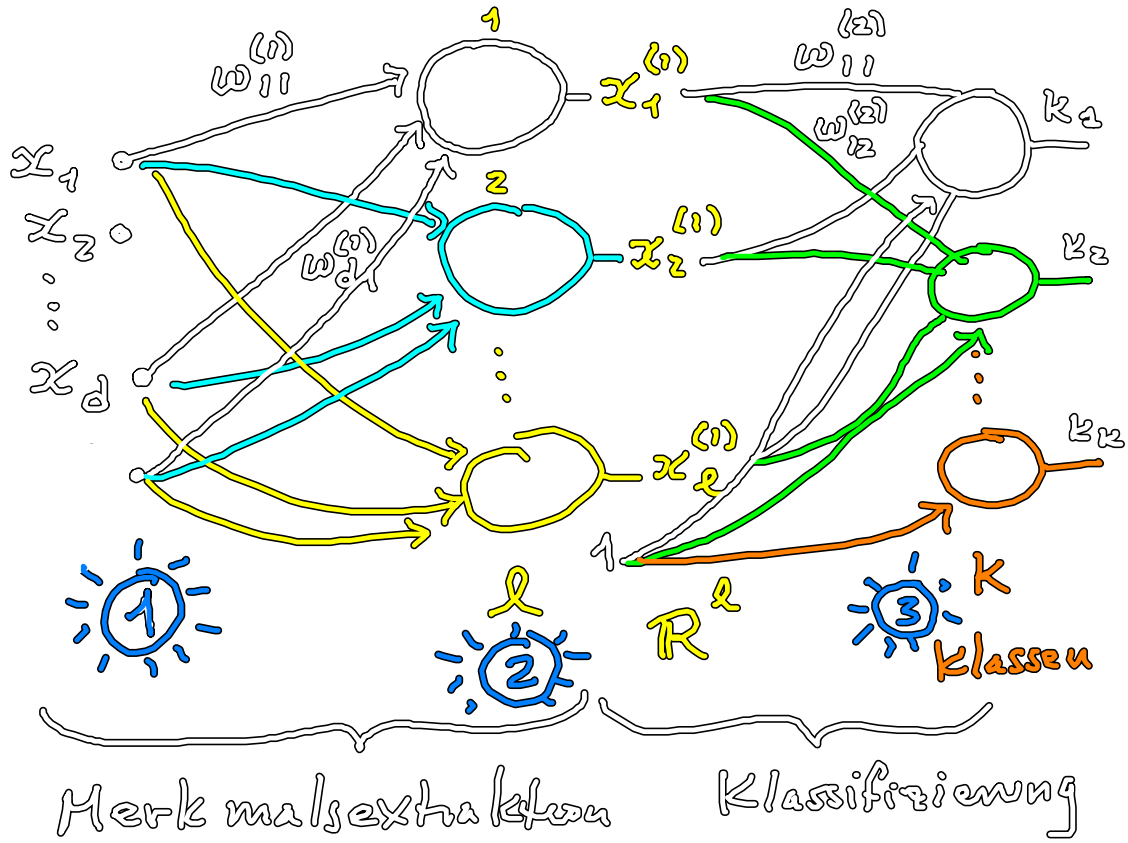
$(x_1^i, x_2^i, \dots, x_d^i)$

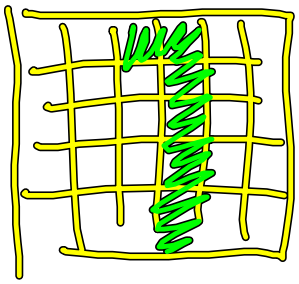
unabhängig



Mustererkennung

Mehrschichtigen Netz





$$\underbrace{0\,0\,1\,1\,0\,0\,0\,0\,0\,1\,0\,0\ldots}$$

$$\underbrace{1\;1\;1\;4}$$

Merkmals-extraktion

Lernproblem:

→ Eine Trainingsmenge ist gegeben

Muster $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_m$

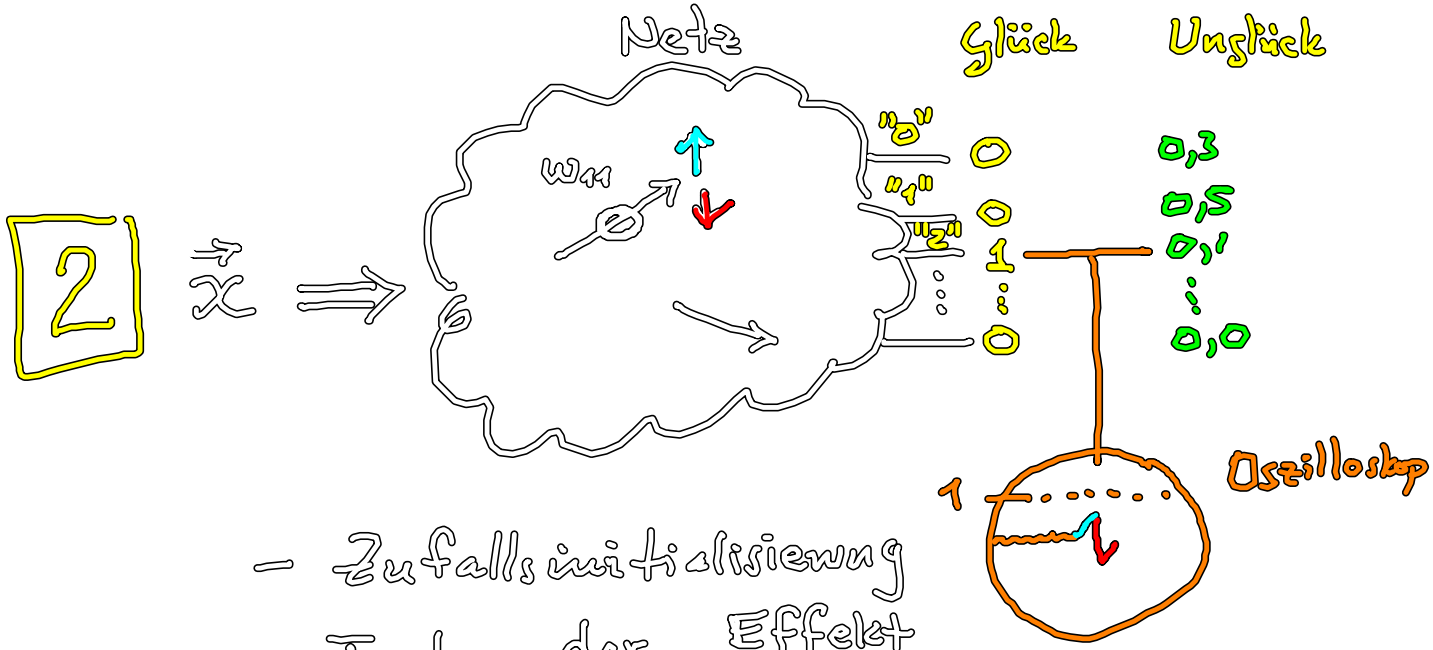
Klassifizierung $\vec{t}_1, \vec{t}_2, \ldots, \vec{t}_m$

→ Finde die Gewichte

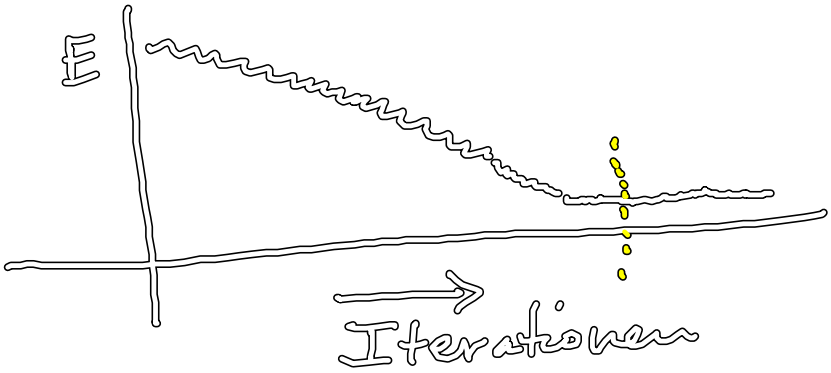
$$w_{11}^{(1)} \ldots w_{d+1,\ell}^{(1)}$$

$$w_{11}^{(2)} \ldots w_{\ell+1,k}$$

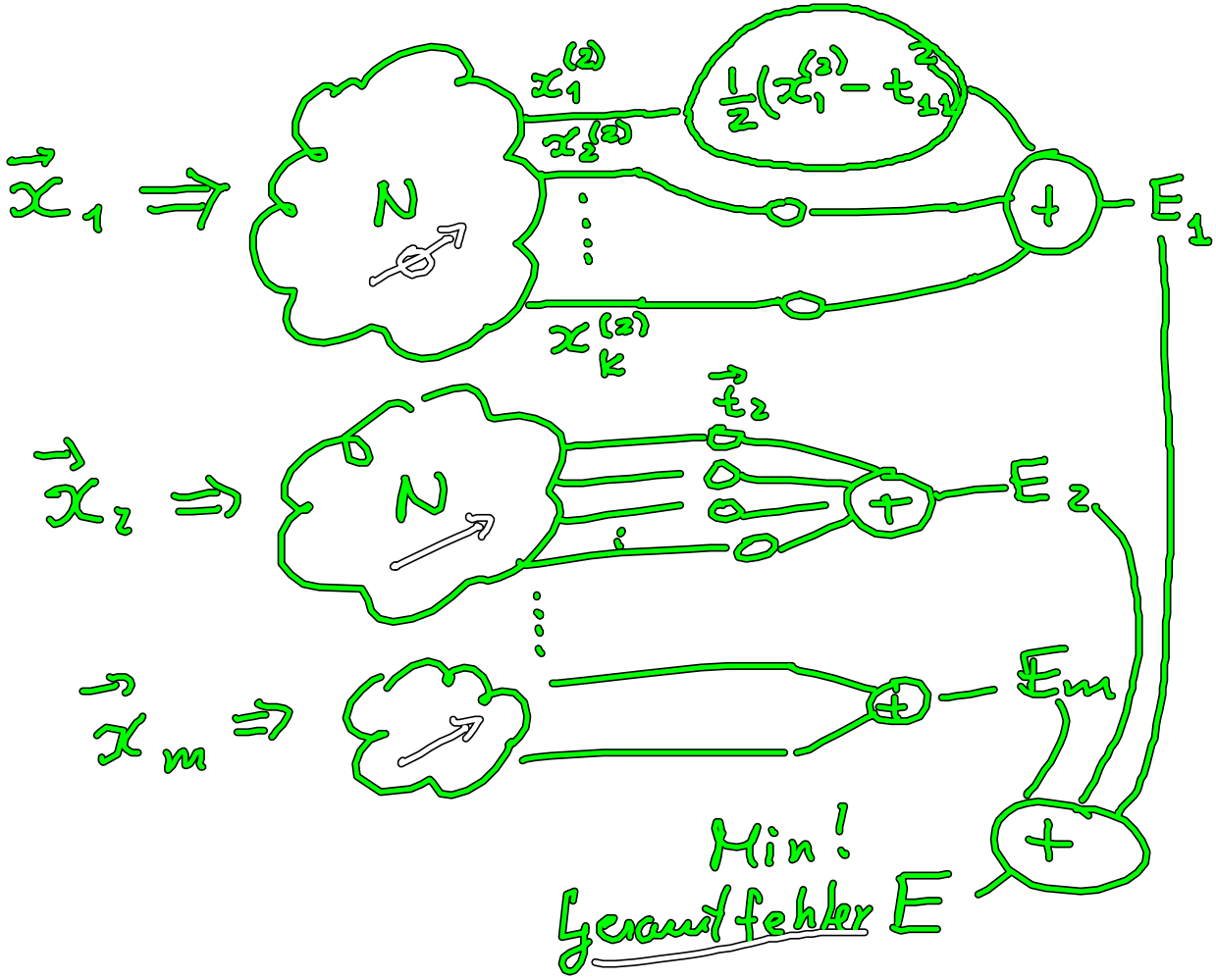
<u>Gradientenabstieg</u>



- Zufallsinitialisierung
- Teste der Effekt von δw_{ij} auf Fehler δE, d.h. $\frac{\partial E}{\partial w_{ij}}$ $\quad \partial E \overset{?}{>} 0$, $\quad \nu_{ij} \overset{?}{>} 0$
- Für alle Gewichte
- Iteriere mehrmals über die Trainingsmenge, bis der Klassifizierungsfehler klein ist.



Fehlerfunktion



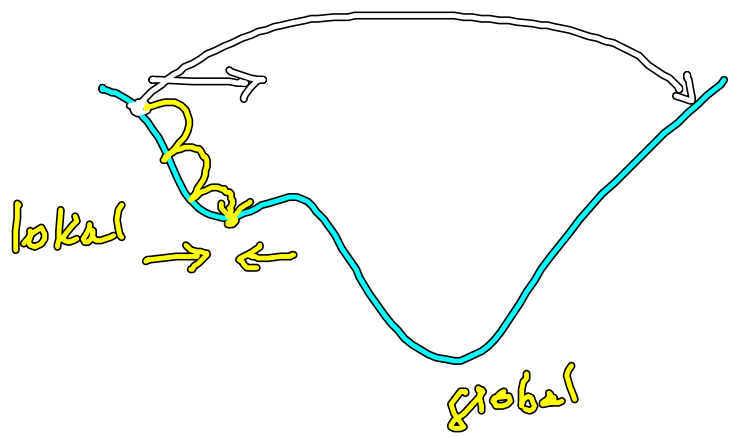
$$\nabla \Rightarrow \vec{\nabla} E = \left(\frac{\partial E}{\partial w_{11}^{(1)}}, \frac{\partial E}{\partial w_{12}^{(1)}} \cdots\cdots \frac{\partial E}{\partial w_{k+1,l}^{(2)}} \right)$$

1 2 1000

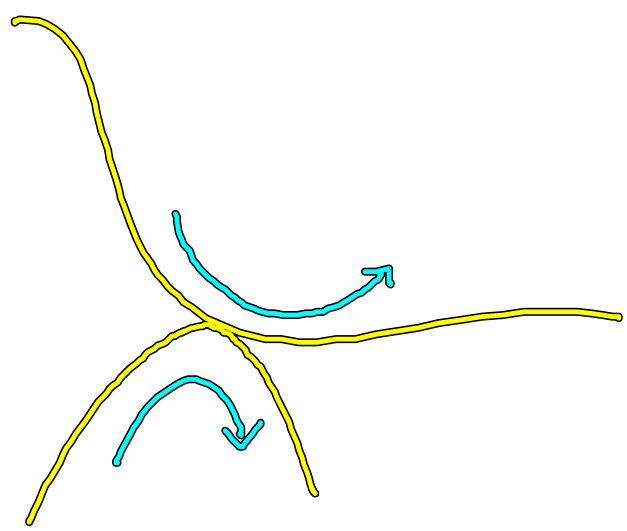
1000 Gewichte

$$\Delta w_{ij} = -\gamma \frac{\partial E}{\partial w_{ij}}$$

$$\gamma << 1$$

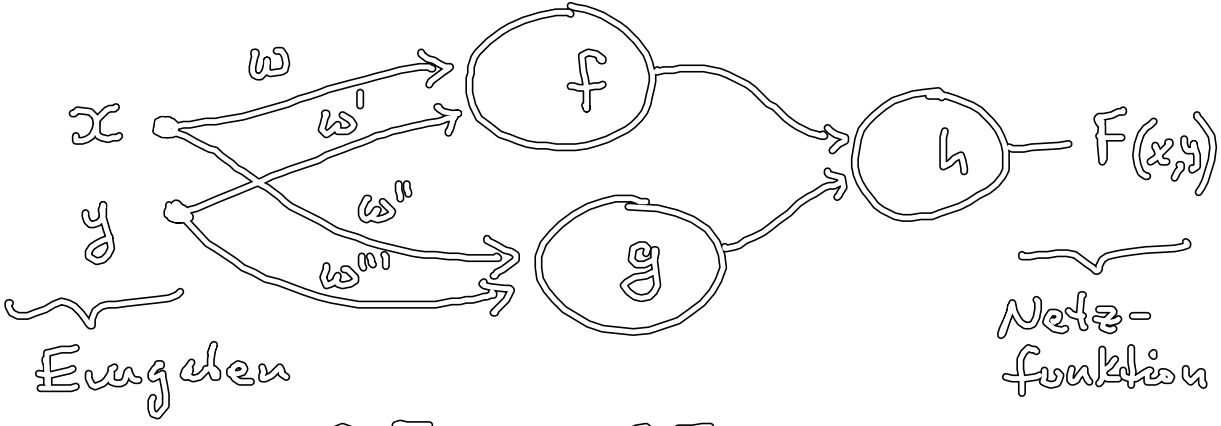


Gradientenabstieg



Gradientenberechnung

$\frac{\partial E}{\partial w_{ij}}$ ← partielle Ableitung



$$\frac{\partial F}{\partial x} \qquad \frac{\partial F}{\partial y}$$

Verschiedene Fälle

Backprop-Algorithmus (Gradienten-abstieg)

I)

$$x \xrightarrow{\quad w \quad} \underbrace{wx}_{\text{Netzfunktion}}$$

A) Feed-forward

B) Backprop

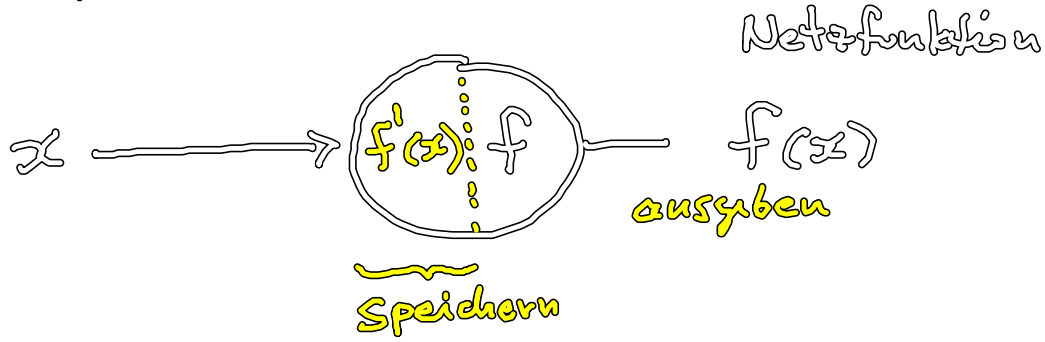
Traversierungs-wert

$$w \xleftarrow{\quad w \quad} 1$$

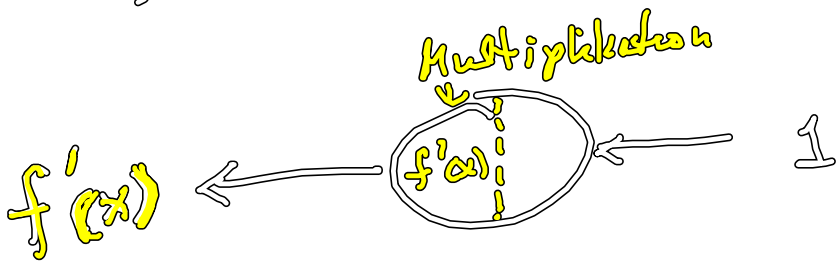
$$w = \frac{d(wx)}{dx}$$

II)

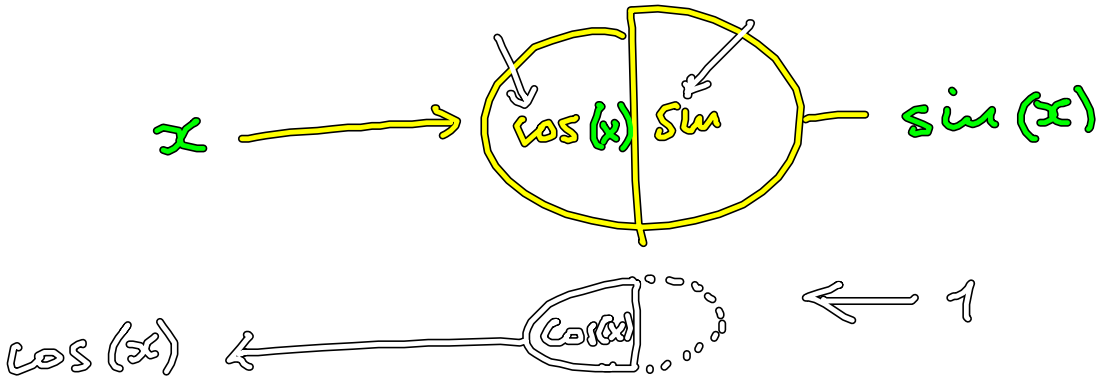
A) Feed-Forward



B) Backprop

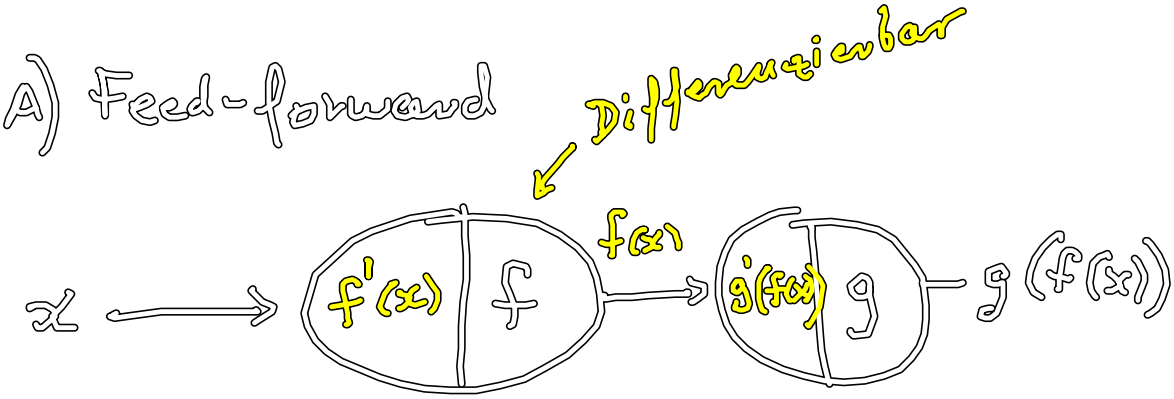


Bsp:

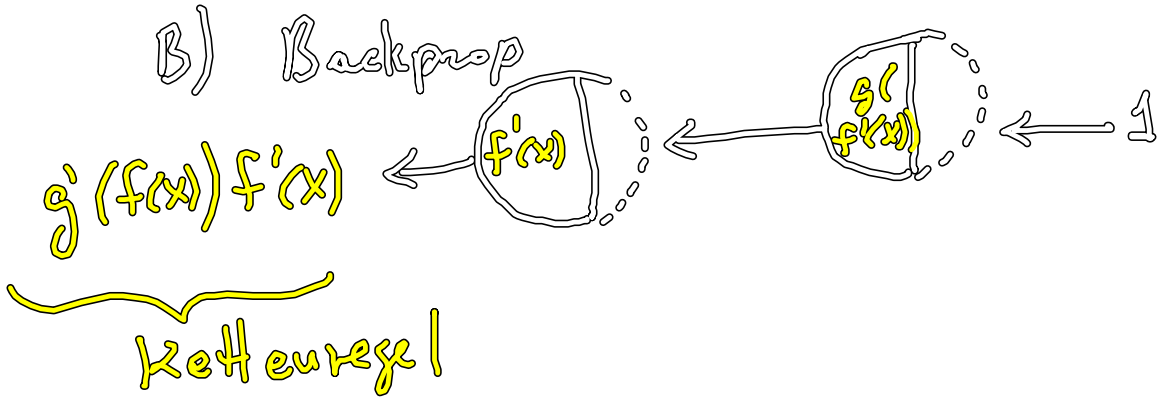


III)

A) Feed-forward

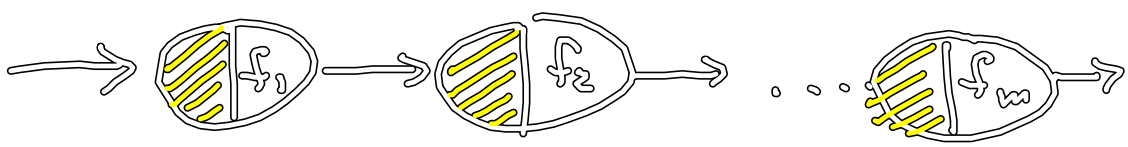


B) Backprop



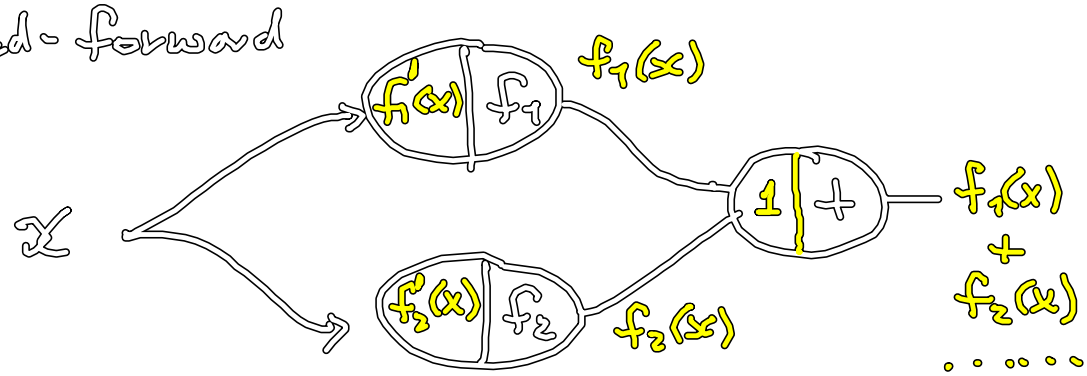
IV)

Induktion .... übung

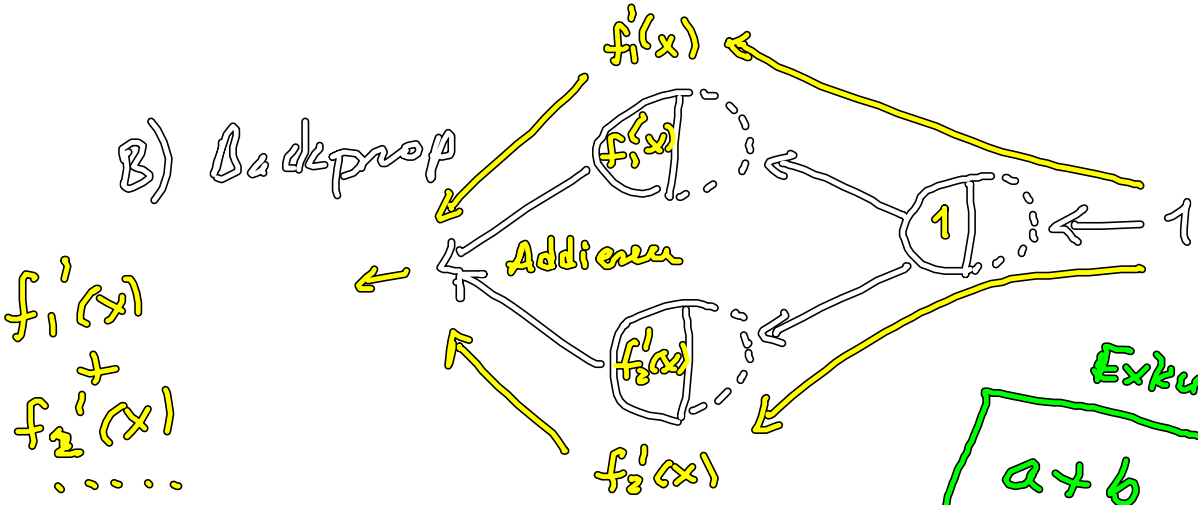


V)

A) Feed-forward



B) Backprop

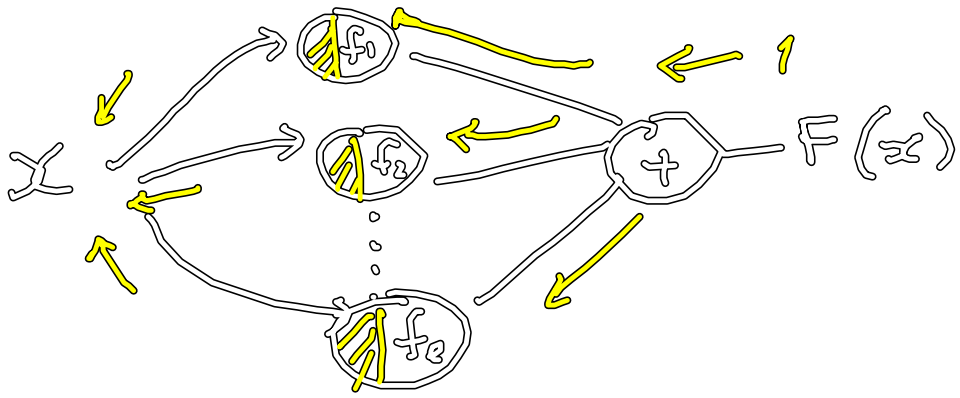


Exkurs

$a+b$

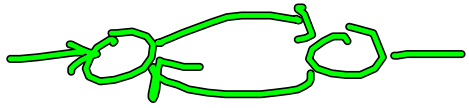
$\frac{\partial}{\partial a}(a+b)=1$

VI) Induktion



Fertig!

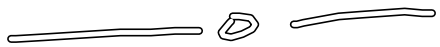
- Funktionskomposition
- Parallele Bearbeitung
- Gewichte
- Differenzierbare Funktionen
- (* Keine Zyklen!)

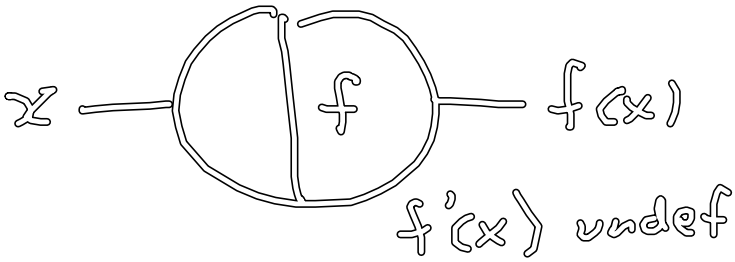


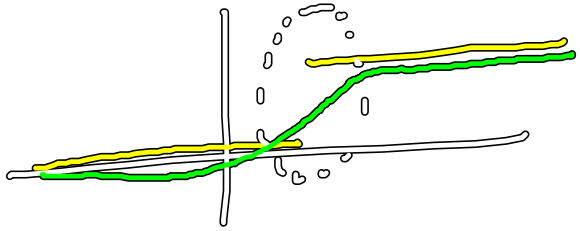
Backprop

→ Feed-forward Stufe:
- Auswertung der Knoten und Weiterpulse
- Speicherung der Ableitung der Knotenfunktion

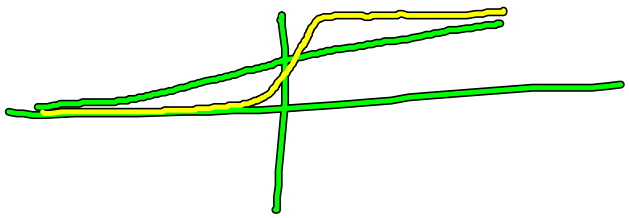
→ Backprop
- Laufe Netz rückwärts
- Sammle linke Seiten multiplikation (an den Knoten) und additiv (wo Kanten sich treffen)





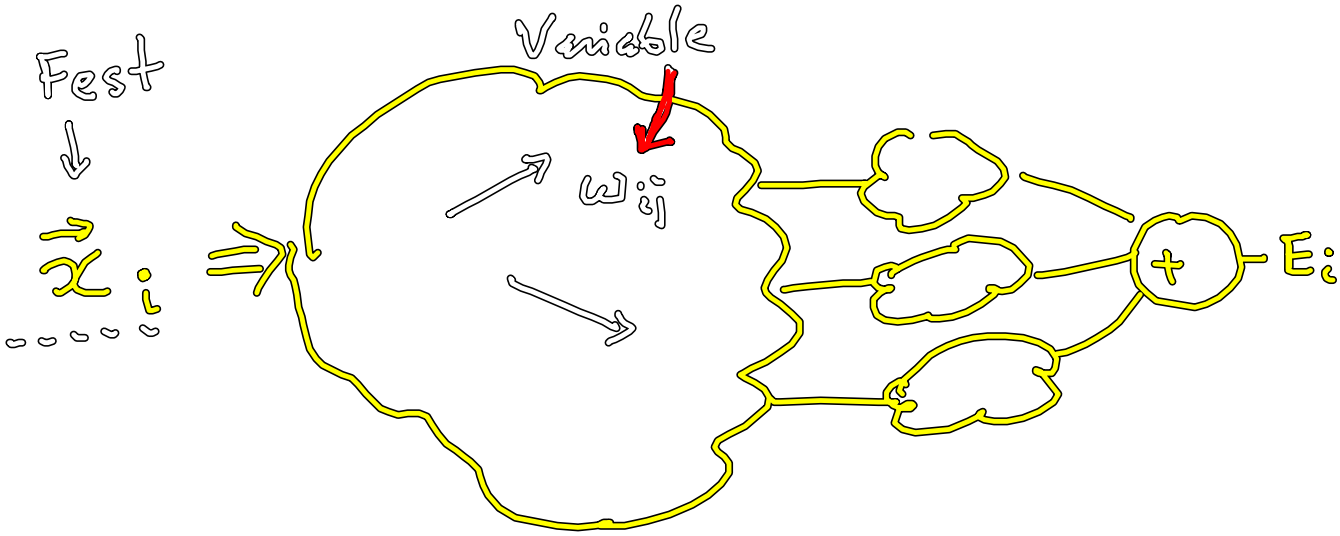


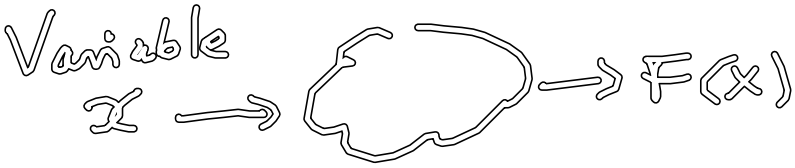
$$s(x) = \frac{1}{1 + e^{-cx}}$$

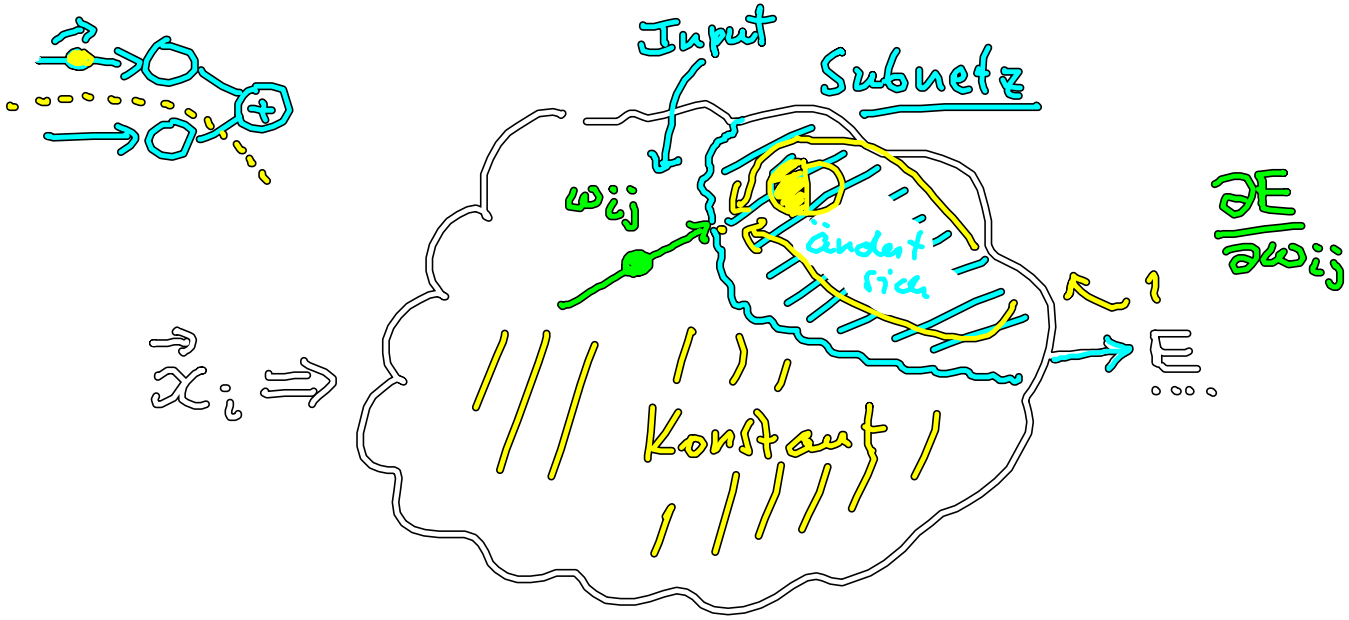


$c = 1$

$c = 100$



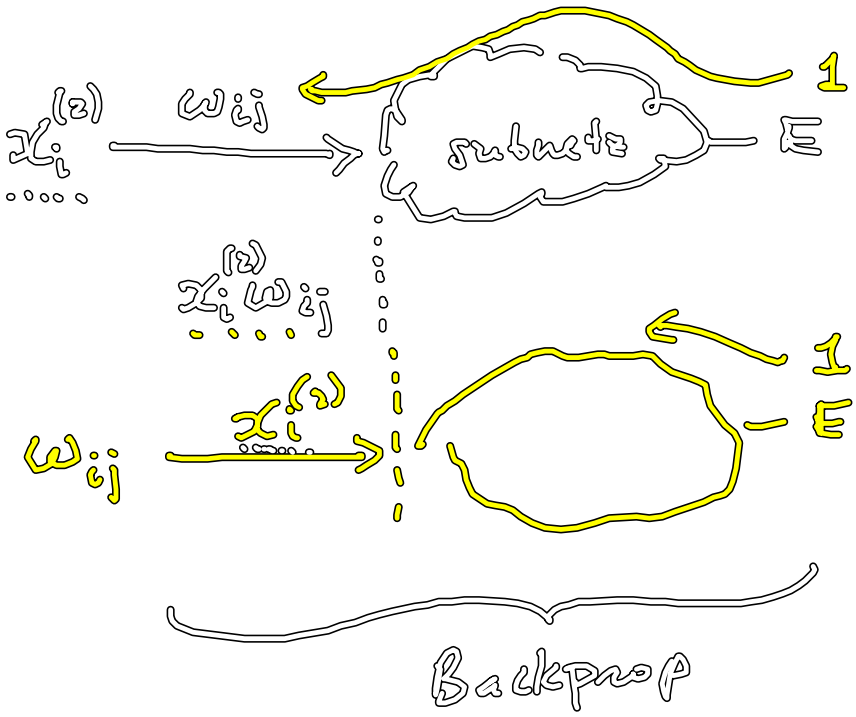




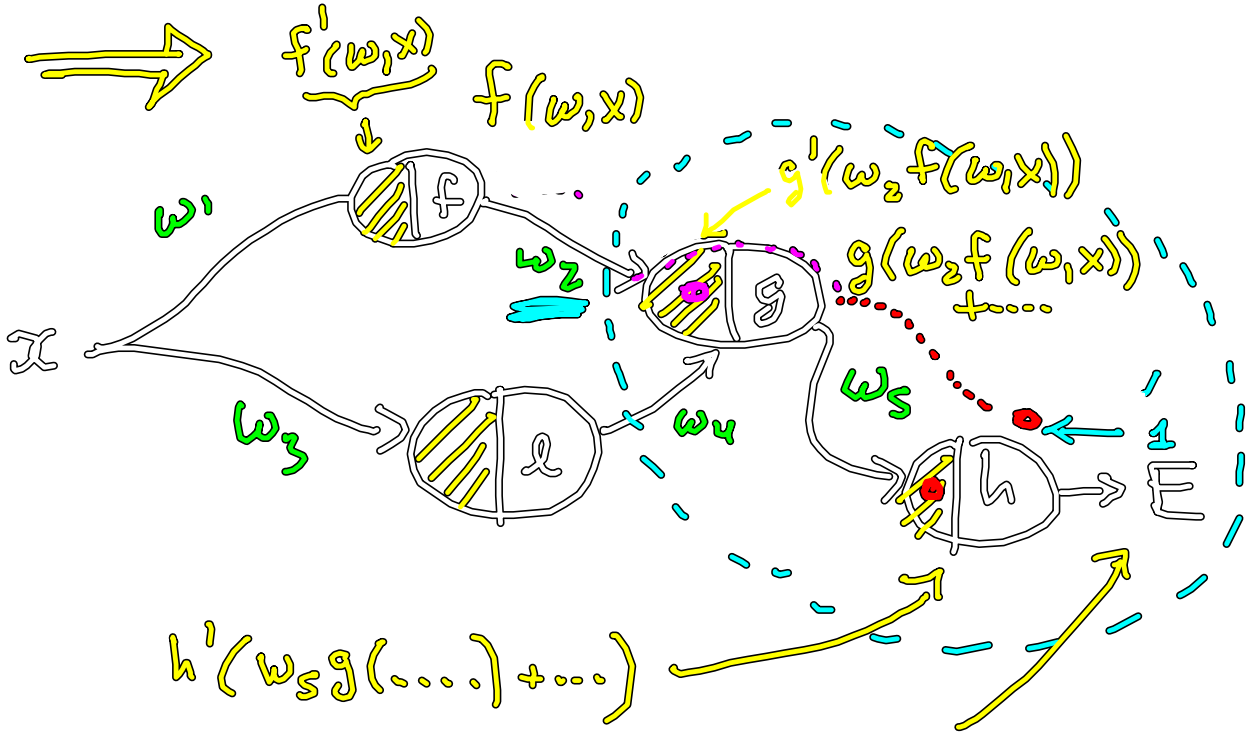
Keine Zykeln

A) Feed-forward (ich benutze das Netz)

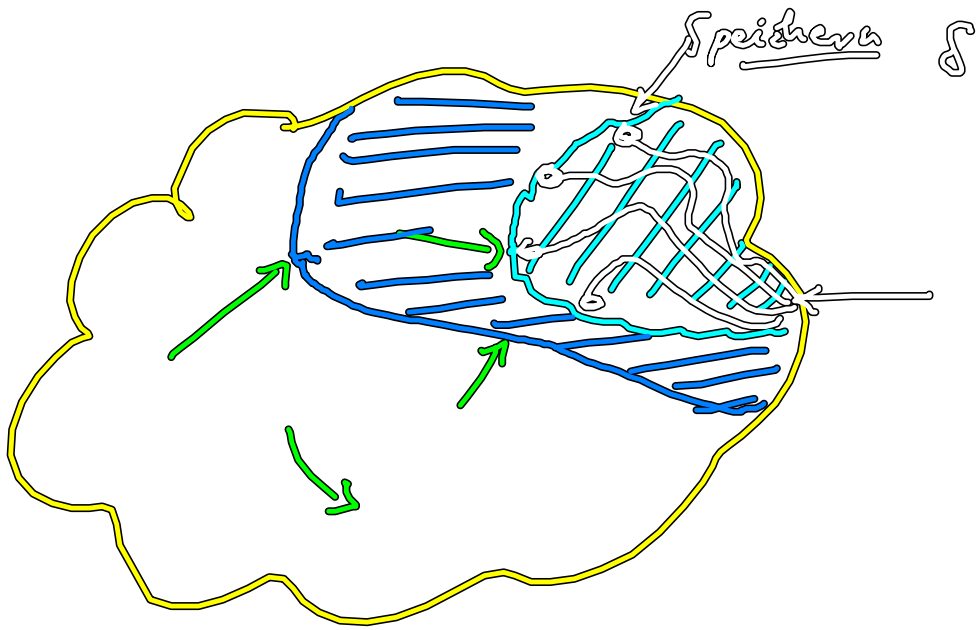
B) Backprop (von E bis w_{ij})



Fertig!



$$h(w_5\, g(w_2\, f(w, x))) + \dots$$



A) Feed-forward — Einmal

B) Backprop — Einmal pro Gewicht

Übungsaufgabe: BP

$$\boxed{1} \longrightarrow \text{"1"}$$

Klein

BP

$$\begin{array}{ccc} & \text{It.} & \\ w_1 & \longrightarrow & w_1 + \Delta w_1 = w_1 - \gamma \frac{\partial E}{\partial w_1} \\ w_2 & \longrightarrow & \vdots \\ \vdots & & \\ w_{1000} & \longrightarrow & \end{array}$$

<u>on-line</u>

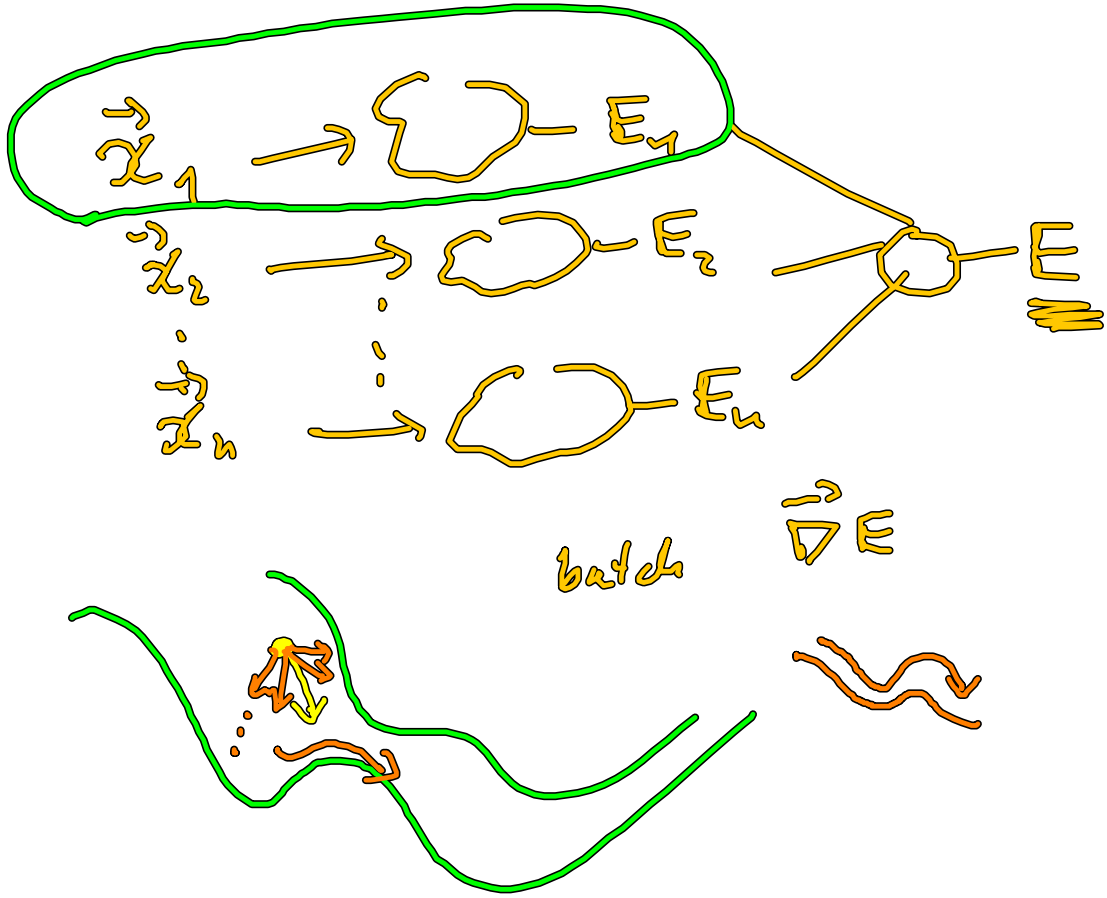
$$\begin{array}{ll} x_1 & w_1 \dots w_{1000} \\ x_2 & w_1 \dots w_{100} \\ \vdots & \\ x_n & \end{array}$$

batch
off-line

$$\begin{array}{l} x_1 \\ x_2 \\ \vdots \\ x_{1000} \end{array}$$

Korrektur $w_1 \; w_2 \dots w_{1000}$

$$w_1 := w_1 - \gamma \frac{\partial E_1}{\partial w_1} - \gamma \frac{\partial E_2}{\partial w_1} - \gamma \frac{\partial E_3}{\partial w_1} \dots$$


$\vec{x}_1$
E_1
$\vec{x}_2$
E_2
$\vec{x}_n$
E_n
E
batch
$\vec{\nabla} E$