

Blatt 3

Luca Krüger, Jonas Otto, Jonas Merkle (Gruppe R)

May 26, 2019

1 Lernregeln

$$E(w, b) = \frac{1}{2} \sum_{\mu=1}^M (T_{\mu} - f(wx_{\mu} + b))^2$$

1.

$$\nabla E(w, b) = \begin{pmatrix} \frac{\partial E}{\partial w} \\ \frac{\partial E}{\partial b} \end{pmatrix} = \begin{pmatrix} -\sum_{\mu=1}^M (T_{\mu} - f(wx_{\mu} + b)) \cdot f'(wx_{\mu} + b) \cdot x_{\mu} \\ -\sum_{\mu=1}^M (T_{\mu} - f(wx_{\mu} + b)) \cdot f'(wx_{\mu} + b) \end{pmatrix}$$

2.

Lernrate η

a) inkrementelle Version:

Woher kommt der Faktor μ ?

$$w(t+1) = w(t) - \mu(T_t - f(wx_t + b)) \cdot \frac{\partial f(wx_t + b)}{\partial w} \cdot x_t$$

$$b(t+1) = b(t) - \mu(T_t - f(wx_t + b)) \cdot \frac{\partial f(wx_t + b)}{\partial w}$$

b) Batch Version:

$$w(t+1) = w(t) + \mu \sum_{\mu=1}^M (T_{\mu} - f(wx_{\mu} + b)) \cdot \frac{\partial f(wx_{\mu} + b)}{\partial w} \cdot x_{\mu}$$

$$b(t+1) = b(t) + \sum_{\mu=1}^M (T_{\mu} - f(wx_{\mu} + b)) \cdot \frac{\partial f(wx_{\mu} + b)}{\partial w}$$

mit $\mu \approx 1/M$

3. a) Python verlinken??
b) Mathematica Grafik mit Startwerten einfügen

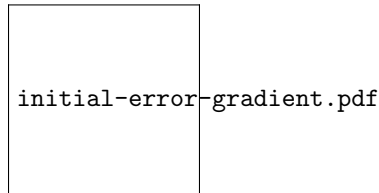


Figure 1: Gradient bei $t = 0$

- c)
d) Der Pfad in Abbildung 3 auf dem Übungsblatt verläuft vom Startpunkt in ein lokales Minimum der Errorfunktion mit hohen Werten für w und b . Problematisch an der Gradienten-Lernregel $E(t)$ ist, dass im allgemeinen gilt $E(t_1) > E(t_2)$ für zwei aufeinanderfolgende Lernschritte t_1, t_2 und somit ein Minimum nur in einer lokal monoton fallenden Umgebung des Startpunktes gefunden werden kann. D.h., der Endpunkt des Pfades hängt bei der Lernregel stark von den Anfangswerten ab und ein gefundenes Minimum ist nicht zwangsweise ein globales Minimum der Errorfunktion.
4. a) Pfad 1 wurde mit der
b)