

# Blatt 3

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

May 31, 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. a) inkrementelle Version:

$$\begin{aligned} w(t+1) &= w(t) + \eta(T_t - f(wx_t + b)) \cdot f'(wx_t + b) \cdot x_t \\ b(t+1) &= b(t) + \eta(T_t - f(wx_t + b)) \cdot f'(wx_t + b) \end{aligned}$$

b) Batch Version:

$$\begin{aligned} w(t+1) &= w(t) + \eta \sum_{\mu=1}^M (T_{\mu} - f(wx_{\mu} + b)) \cdot f'(wx_{\mu} + b) \cdot x_{\mu} \\ b(t+1) &= b(t) + \eta \sum_{\mu=1}^M (T_{\mu} - f(wx_{\mu} + b)) \cdot f'(wx_{\mu} + b) \end{aligned}$$

3. a) (Siehe Jupyter Notebook)

b) Negativer Gradient (*blau*) im initialen Punkt  $\begin{pmatrix} w(0) \\ b(0) \end{pmatrix} = \begin{pmatrix} -1 \\ 3 \end{pmatrix} :$

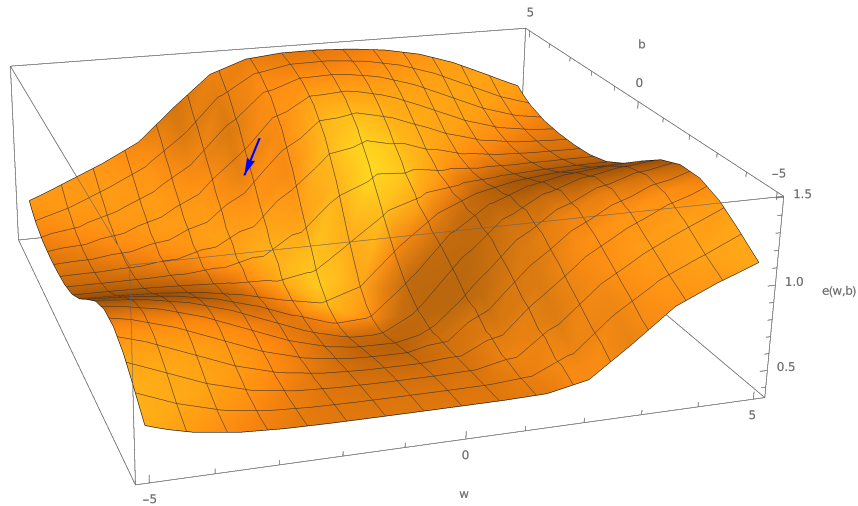


Figure 1: Gradient bei  $t = 0$

- c) (Siehe Jupyter Notebook)
  - 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 Batch-Lernregel und *Pfad 2* mit der inkrementellen Lernregel erzeugt. Der Gradient von *Pfad 1* ist im Gegensatz zu *Pfad 2* eine stetige Funktion, da in jedem Lernschritt die Errorfunktion über den gesamten Trainingsdatensatz berechnet wurde. Für die inkrementelle Lernregel, wird in jedem Lernschritt die Errorfunktion nur über ein Element aus dem Trainingsdatensatz berechnet.
  - b) Die Lernregel über alle Datenpunkte ist insbesondere bei größeren Datensätzen sehr zeitintensiv. Die inkrementelle Lernregel gewichtet einzelne Extremwerte im Datensatz möglicherweise zu stark.
  5. Wenn die Lernrate sehr groß gewählt wird, erreicht der Fehler zwar schnell eine Umgebung des Minimums, springt dann aber über das Minimum und oszilliert um das Minimum herum oder divergiert sogar.
  6. Die Werte sind nicht linear separierbar, das Klassifizierungsproblem kann also von diesem Netz nicht gelöst werden.