Einführung in Neuronale Netze
Backpropagation Learning

Fehlerbestimmung

Die Bestimmung des Netzwerkfehlers ist von besonderer Bedeutung, denn nur so lassen sich Aussagen über die Güte des Netzes machen.

Das Backpropagation-Netz liefert für einen Eingabevektor $\vec{x}$ und eine fest vorgegebene Belegung aller Gewichte im Netz stets dieselbe Ausgabe.

Daher kann man das Ein- und Ausgabeverhalten des Netzes als Funktion o_{Netz} dargestellt werden.

Bezeichne q die Gesamtzahl der Gewichte im Netz und $\vec{w} = (\vec{w}_{U_1}, \ \dots \ , \vec{w}_{U_{H-1}}) \in \mathbb{R}^q$ den Gewichtsvektor des Netzes. Dann gilt für die Netzwerkausgabe: $\vec{o} \ = \ o_{Netz}(\vec{x}, \vec{w})$.

Die Güte der Netzwerkausgabe bezüglich eines gegebenen Trainings- oder Testdatums $(\vec{x}, \vec{o}\,')$ wird durch $\vec{o}\,' - \vec{o}$ beschrieben. Bezeichnet f die zu "lernende" Funktion, so gilt $\vec{o}\,' = f(\vec{x})$ und die Güte des Netzes wird ermittelt durch: $f(\vec{x}) - o_{Netz}(\vec{x}, \vec{w})$.

Allerdings macht es wenig Sinn zwischen positiven und negativen Fehlern zu unterscheiden. Daher bietet es sich an, den Fehler zu quadrieren.

Definition:
*Der **quadratische Fehler F** eines Backpropagtionnetzes ist für gegebenes $\vec{x}$ und $\vec{w}$ definiert durch*

$$F(\vec{x}, \vec{w}) \ = \ | \ f(\vec{x}) - o_{Netz}(\vec{x}, \vec{w}) \ |^2 \ .$$

Der quadratische Fehler ist ein Maß für die Güte eines Neuronalen Netzes.

Die obige Definition liefert lediglich eine Aussage über die Güte hinsichtlich der konkreten Eingabe $\vec{x}$. Aber nur ein Fehlermaß bezüglich aller möglichen Eingabevektoren liefert eine repräsentative Aussage über die Güte eines Neuronalen Netzes.

Dies wird erreicht, indem der quadratische Fehler für möglichst viele Eingabevektoren $\vec{x}_k$, *k = 1,2,...* bestimmt und dann über die Anzahl der Muster gemittelt wird.

Definition:
Der ***mittlere quadratische Fehler (MSE)*** *ist definiert durch*

$$F(\vec{w}) = \lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} F(\vec{x}_k, \vec{w}).$$

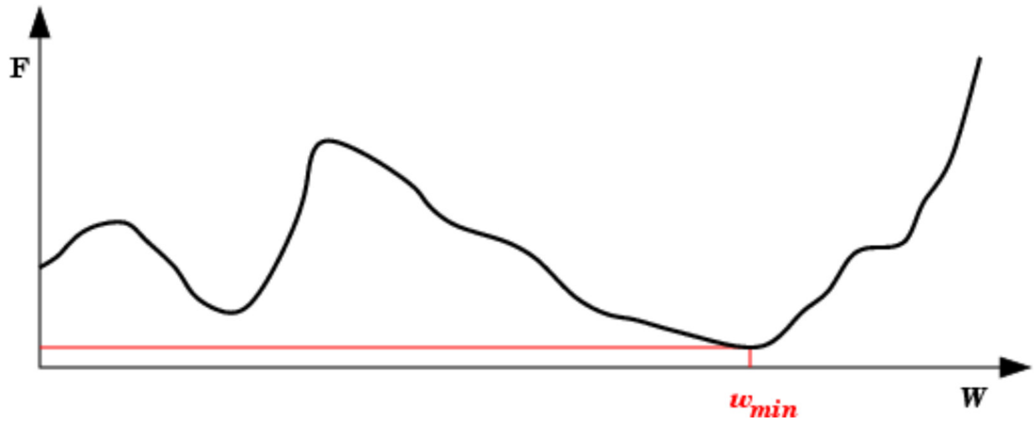
Der MSE ist das am häufigsten verwendete Fehlermaß. Allerdings gibt es auch eine Reihe von Anwendungen, wie beispielsweise die Bildverarbeitung, bei denen andere Fehlermaße sinnvoll sind.

Der mittlere quadratische Fehler hat jedoch eine Reihe von Eigenschaften, die für Backpropagation-Netze von Vorteil sind:

- $F(\vec{w})$ existiert, da die Reihe konvergiert.
- $F(\vec{w})$ ist stetig und differenzierbar in $\vec{w}$.
- $F(\vec{w}) \geq 0$.

Durch variieren von $\vec{w}$ läßt sich $F(\vec{w})$ für den ganzen Gewichtsraum bestimmen. Hat man für alle möglichen Gewichtskombinationen den MSE bestimmt, kann man den Fehler gegen die Gewichte in ein Koordinatensystem eintragen.

Definiert man W als den Vektorraum aller $\vec{w}$ und vereinfacht die Betrachtung, ergibt sich beispielsweise die folgende Fehlerverlaufskurve:



Die Struktur der Fehleroberfläche wird maßgeblich durch die Wahl einer sigmoiden Ausgabefunktion beeinflußt. Sie glättet die Fehleroberfläche und erleichtert so den Gradientenabstieg. Viel entscheidender ist jedoch, daß ein Punkt auf der Fehleroberfläche eine Neigung in die Richtung des nächstliegenden Tals besitzt. Wenn diese Richtung verfolgt wird, wenn also der Gewichtsvektor in Richtung von $-\nabla_{\vec{w}} F(\vec{w})$ verschoben wird, endet der Gradientenabstieg unweigerlich im nächstliegenden Tal.

Ziel der Gewichtsveränderung ist eine Minimierung von F, indem eine Belegung für $\vec{w}$ gefunden wird, für die F ein absolutes Minimum F_{min} erreicht.

Auf den theoretisch erreichbaren Wert wird im Kapitel Lokale Minima eingegangen.

Allerdings müssen an dieser Stelle zwei elementare Sachverhalte beachtet werden:

Die Gewichte $\vec{w}$ auf der W-Achse sind mehrdimensional. Der Fehler entspricht also keiner Kurve, sondern einer Oberfläche. Damit repräsentiert $F(\vec{w})$ somit der Höhe die Fehleroberfläche im Punkt $\vec{w}$.

Viel entscheidender ist jedoch, daß nicht die gesamte Fehleroberfläche bestimmt wird. Der nötige Rechenaufwand wäre inakzeptabel groß. Wie die Aufgabe besser gelöst werden kann, zeigt das folgende Kapitel.

Zurück zum letzten Kapitel Zum nächsten Kapitel