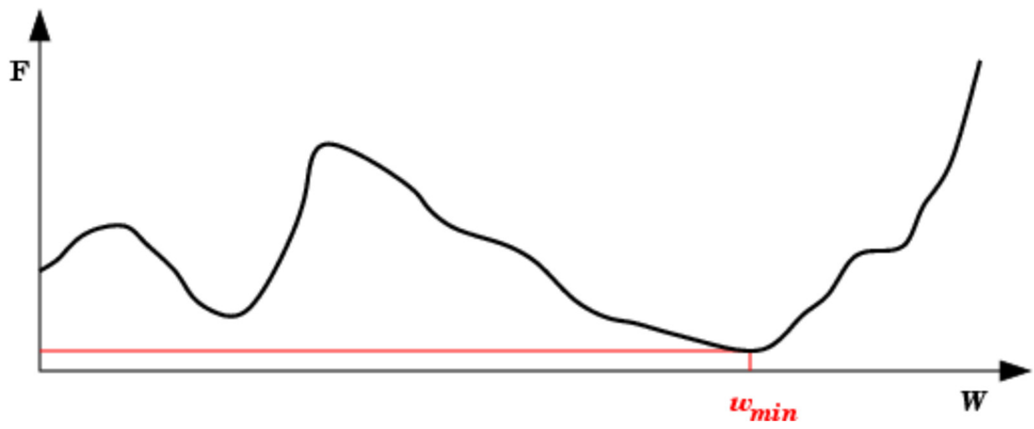
Einführung in Neuronale Netze
Backpropagation Learning

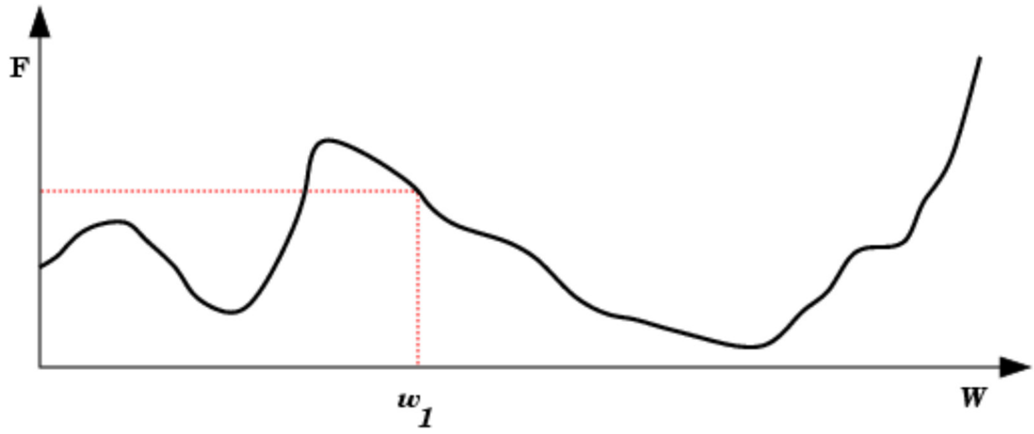
Lernregel

Es stellt sich die Frage, welches Verfahren den Gewichtsvektor $\vec{w}$ derartig verändert, daß $(\vec{w}, F(\vec{w}))$ in ein möglichst globales Minimum bewegt wird, ohne die gesammte Fehleroberfläche zu kennen.

Hierzu soll nochmals der zweidimensionalen Fall betrachtet werden. Es wird ein $\vec{w}$ auf der W-Achse gesucht, für das $F(\vec{w})$ minimal wird.



Das Backpropagation-Verfahren beruht auf einem Gradienabstiegsverfahren. Das Verfahren startet mit einem zufällig gewählten Gewicht w_1. In diesem Punkt wird der Gradient bestimmt und um eine vorgegebene Länge, die Lernrate, herabgestiegen. Die so erhaltene Gewichtsveränderung liefert den Gewichtsvektor w_2, in dem wiederum der Gradient bestimmt wird. Das Verfahren wiederholt sich, bis das lokale Minimum erreicht ist.



Durch Betätigen von vorheriger Schritt und nächster Schritt wird der Gradientenabstieg für eindimensionale Gewichte schrittweise visualisiert. Das Verfahren startet mit der Bestimmung der Tangente in w_1. Diese wird der Lernrate entsprechend weit herabgestiegen und somit die Gewichtsveränderung ermittelt. Zu dem so erhaltenen Gewichtsvektor w_2 wird der zugehörige Fehler bestimmt. Liegt der Fehler oberhalb der zuvor definierten Güteschwelle, wird in w_2 die Tangente bestimmt und das Verfahren wiederholt sich, bis der Punkt w_{min} erreicht wird.

Es muß somit nicht die gesammte Fehleroberfläche bekannt sein, um das Minimum der Fehleroberfläche bestimmen zu können. In diesem Beispiel wird das Minimum durch Kenntnis von lediglich fünf Punkten der Fehleroberfläche bestimmt.

Der Gradient

Q bezeichnet die Anzahl der Gewichte im Netz.
Daher sind die Vektoren $\vec{w}$ wie auch der Gradient q-dimensional.
Damit lautet die allgemeine Form für die Richtung des Anstiegs:

$$\nabla_{\vec{w}} F(\vec{w}) := \left(\frac{\partial F(\vec{w})}{\partial w_1}, \ \dots \ , \frac{\partial F(\vec{w})}{\partial w_q} \right)^T.$$

Backpropagation ist ein Gradientenabstiegs-Verfahren. Durch Verschieben von $\vec{w}$ in Richtung $-\nabla_{\vec{w}}\ F(\vec{w})$ kann also ein Minimum der Fehlerfunktion bestimmt werden. Ob dieses Minimum lokal oder absolut ist und welche Probleme aus diesem Unterschied entstehen, wird im Kapitel Lokale Minima erläutert.

Um den Gradienten zu bestimmen, bietet sich eine Berechnung der einzelnen Komponenten an. Allerdings stellt sich die Frage nach der Existenz dieses Vektors. Die Antwort liefert ein Satz, der von *Hecht-Nielsen* mit Hilfe von statistischen Methoden bewiesen wurde :

Satz:

Sei $F_k\ (\vec{x}_k, \vec{w})\ =\ |\ f(\vec{x}_k) - o_{Netz}(\vec{x}_k, \vec{w})\ |^2$

und $F(\vec{w}) := \lim\limits_{N \to \infty} \frac{1}{N} \sum\limits_{k=1}^{N} F_k\ (\vec{x}_k, \vec{w})$.

Dann ist F differenzierbar in $\vec{w}$ und es gilt :

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

Als abkürzende Schreibweise wird im weiteren verwendet: $\nabla_{\vec{w}}\ F(\vec{w})\ =\ \lim\limits_{N \to \infty} \frac{1}{N} \sum\limits_{k=1}^{N} \nabla_{\vec{w}}\ F_k\ (\vec{w})\,.$

Zurück zum letzten Kapitel Zum nächsten Kapitel