Kurzzusammenfassung neuronale Netzwerke

Ein neuronales Netzwerk ist eine Variante der künstlichen Intelligenz. Man könnte neuronale Netze auch als gerichtete und gewichtete Graphen bezeichnen. Ich werde mir zumindest im Folgenden die Nomenklatur der Graphentheorie zunutze machen und von Knoten und Kanten sprechen. Am meisten werden neuronale Netze die vollständig verbunden sind benutzt. Das heisst, dass alle Knoten aus einer Ebene mit allen aus der nächsten verbunden sind. Diese Eigenschaft ermöglicht mathematisch vereinfachte Darstellung durch Matrizen. In die erste Schicht wird der Input eingegeben. Die Werte der zweiten Schicht werden dann so ausgerechnet, dass alle Knotenwerte, die mit dem Knoten, den wir gerade anschauen, verbunden sind, mit dem Gewicht der jeweiligen Kant multipliziert und dann aufsummiert werden. Häufig wird auch noch eine Aktivierungsfunktion auf dieser Summe angewendet um unhandlich grosse/kleine Ausgaben zu verhindern.

Bevor das Netz trainiert wird werden alle Parameter, die die schlussendliche Ausgabe neben dem Input beeinflussen (das heisst alle Kantengewichte) zufällig festgelegt. Danach bei jedem Trainingsbeispiel für jeden Ausgabeknoten ein Fehler bestimmt. Also wie weit die tatsächliche Ausgabe von der gewünschten Ausgabe entfernt ist. Danach wird der Fehler «backpropagiert» das heisst wir weisen jeder Kante einen Anteil dieses Fehlers proportional zum Gewicht dieser Kante zu. Aus diesen Anteilen weisen wir einem Knoten in der zweitvordersten Schicht einen Fehler zu. Dieser Fehler errechnet sich aus der Summe aller Fehler aller Kanten, welche diesen spezifischen Knoten in Richtung Ausgabeschicht verlassen. Damit das neuronale Netz nun lernen kann müssen die Kantengewichte so angepasst werden, dass der Fehler minimiert wird. Da es meistens viel zu viele Kanten gibt, um mathematisch ein Optimum herzuleiten, wird das Gradient Verfahren verwendet. Für jedes Kantengewicht wird also ein Gradient bestimmt. Der Gradient gibt uns an, wie stark und in welche Richtung wir die Ausgabe beeinflussen, wenn wir dieses Kantengewicht verändern. Dieser Gradient kann durch partielle Ableitung bestimmt werden. Diese partielle Ableitung kann durch cleveres Einsetzen der Kettenregel einigermassen problemlos errechnet werden. Meist stellt die Aktivierungsfunktion das grösste Hindernis dar.