Neuronale Netzwerke

Speziell: "Fully Connected Feed-Forward Network"

Feed-Forward:

- Das Modell besteht aus L Schichten von Neuronen
- In jeder Schicht l haben wir $m(l)$ Neuronen
- Jedes Neuron in der Schicht l besteht aus einem "Bias" $b \in \mathbb{R}$ und einem Vektor von Gewichten $\vec{w} \in \mathbb{R}^{m(l-1)}$
- In anderen Worten, pro Schicht l haben wir eine Matrix von Gewichten $W \in \mathbb{R}^{m(l) \times m(l-1)}$ und einen Vektor von Biases $\vec{b} \in R^{m(l)}$
- Wir haben eine Aktivierungsfunktion $f : \mathbb{R} \to \mathbb{R}$, die in jedem Neuron zur Berechnung der Ausgabe verwendet wird.

Die Ausgabe der l-ten Schicht ist ein Vektor, der auch $\vec{a}^{(l)}$ genannt wird. Diese ist wie folgt definiert (wobei $\vec{x}$ die Eingabe und f die Aktivierungs-Fkt. ist):

$$\vec{a}^{(1)} = \vec{x}$$

$$\vec{a}^{(l)} = f(W^{(l)} \cdot \vec{a}^{(l-1)} + \vec{b}^{(l)})$$

Sinnvolle Aktivierungsfunktionen sind z.B.:

- Sigmoid-Fkt (siehe Klassifikation)
- ReLU-Fkt: $ReLU(t) = max(0, t)$

Fully Connected

Jeder Knoten aus jeder Schicht n ist mit jedem Knoten aus der nächsten Schicht $n+1$ verbunden.

Kostenfunktion

Während $\vec{x}$ die Eingabe darstellt, benennen wir mit $\vec{y}$ die Ausgabe.
Für gegebene Eingabe- und Ausgabewerte, sind die Kosten dann:

$$C_{\vec{x},\vec{y}} = \frac{1}{2}(\vec{a}^{(L)} - \vec{y})^2$$

Für jede Schicht vor der letzten Schicht L werden die Kosten wie folgt berechnet:

$$C^{(l)} = \frac{1}{2n}\sum_{i=1}^{n}(\vec{a}^{(l)} - \vec{y}^{(l)})^2$$