Vorlesung 15 Nicht-lineare Regression cont'd

Erinnerung: Wir haben ein Modell:

$$f_\Theta : \mathbb{R}^D \to \mathbb{R}^N$$

bzw. ein stat. Modell $P_\Theta(y|x)$.

z.B.: $\begin{cases} \text{das lineare Modell ist } f_\Theta(x) = (x^T\Theta_i^1 + \Theta_0^i)_{i=1}^N, \text{ wobei } \Theta_i = (\Theta_i^1, \Theta_0^i). \\ \text{das nicht-lin. Modell ist } f_\Theta(x) = (\phi(x)^T\Theta_i^1 + \Theta_0^i)_{i=1}^N \end{cases}$

Eine Erweiterung dieses Frameworks bieten <u>neuronale Netze</u>.

<u>Definition 15.1</u>

Sei $L > 0$ und $N_0, N_1, \ldots, N_L > 0$. Für $1 \le i \le L$ sei

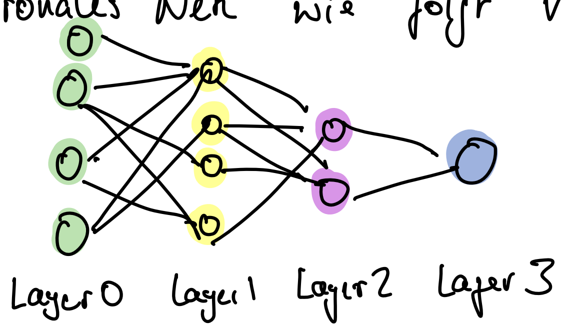
$$g_{\Theta_i} : \mathbb{R}^{N_i} \to \mathbb{R}^{N_{i-1}}$$

ein <u>Modell</u> für nicht-lineare Regression. Das def. Modell

$$f_\Theta = (g_{\Theta_1} \circ \ldots \circ g_{\Theta_L}) : \mathbb{R}^D \to \mathbb{R}^N,$$

wobei $N_L = D$, $N_0 = N$, heißt neuronales Netz von Tiefe L. Die Funktion g_{Θ_i} wird <u>iter Layer</u> genannt.

Warum der Begriff "neuronales Netz"? Die Idee hinter Def. 15.1 ist, dass f_Θ die Struktur eines Gehirns nachbilden soll. Denn z.B. für $L = 3$ und $N = 1$, $D = 4$, können wir ein neuronales Netz wie folgt visualisieren: $(N_1 = 2, N_2 = 4)$



Bemerkung Für Tiefe $L > 1$ können wir oft keine geschlossene Form für ERM, RR, MLE, MAP berechnen.
D.h. für die Parameterschätzung müssen wir Methoden aus der Optimierung verwenden.
Daher werden für die g_{θ_i} oft "einfache" Funktionen.

Eine Wahl, die oft getroffen wird ist

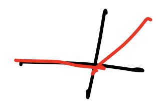
$$g_{\theta_i}(x) = \sigma(f_{\theta_i}(x))$$

mit

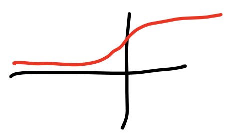
- $f_{\theta_i}(x)$ ist ein nicht-lineares Regressionsmodell (meist sogar linear!)
- σ ist eine nicht-lineare <u>Activation-Function</u>

Beispiele von solchen Activation-Functions:

1) ReLu-Funktion $\sigma(z) = \left(\max(0, z_i)\right)_{i=1}^{k}$, $z \in k$



2) Sigmoid-Funktion $\sigma(z) = \left(\dfrac{1}{1+e^{-z_i}}\right)_{i=1}^{k}$.



3) Softmax $\sigma(z) = \left(\dfrac{\exp(+z_i)}{\sum_{j=1}^{k} \exp(+z_j)}\right)_{i=1}^{k}$

Wir können neuronale Netze für stat. Modelle verwenden:
z.B., wenn die Output-Variable $y \in \{c_1, \dots, c_\ell\}$ kategorisch ist.
Wir erhalten ein stat. Modell für $(y|x)$ durch

$$P_\theta\left((y|x) = c_i\right) = (f_\theta(x))_i$$

mit f_θ ein NN, s.d. $f_\theta(x)_i \geq 0$ und $\sum_i f_\theta(x)_i = 1$.