

Aufgaben

- 1 Arbeit an einem CNN für den Pascal VOC 2012 Datensatz fortsetzen
- 2 Implementierung des Ansatz der Deep Taylor Decomposition für DNN (insbesondere CNN)
- 3 Vergleich LRP \leftrightarrow Deep Taylor Decomposition

Aufgabe

- Einführung CNN Part und Wdh. Deep Taylor

Herleitung z^B -Regel

- Eine der Grundannahmen der Deep Taylor Decomposition ist die Anwendung der ReLU-Aktivierungsfunktion.
 \Rightarrow für Input eines Layers gilt $x \in \mathbb{R}^+$
- Für Inputwerte des ersten Layers (z.B. Pixelwerte) gilt dies i.A. nicht
- Diese sind meistens beschränkt und können auch Werte kleiner 0 annehmen
- Formal: $x \in \mathcal{B}$ mit

$$\mathcal{B} = \{x \in \mathbb{R}^d : l_i \leq x_i \leq h_i \quad \forall i \in \{0, \dots, d\}\}$$

Herleitung z^B -Regel

- Gesucht wird Nullstelle für R_j auf dem Intervall

$$[l_i 1_{w_{ij}>0} + h_i 1_{w_{ij}<0}, x_i]^d$$

- Auf diesem Intervall existiert mindestens eine Nullstelle, denn

$$\begin{aligned} & R_j (i 1_{w_{ij}>0} + h_i 1_{w_{ij}<0}) \\ &= \max \left(0, \sum_i l_i 1_{w_{ij}>0} \cdot w_{ij} + h_i 1_{w_{ij}<0} \cdot w_{ij} + b_j \right) \\ &= \max \left(0, \sum_i l_i \cdot w_{ij}^+ + h_i \cdot w_{ij}^- + b_j \right) = 0 \end{aligned}$$

Herleitung z^B -Regel

- Dieses Intervall ist also für die Suche zulässig und wir fügen die Richtung $v_i^{(j)}$, mit

$$v_i^{(j)} = x_i - l_i 1_{w_{ij} > 0} + h_i 1_{w_{ij} < 0},$$

- in die Grundformel

$$\sum_j \frac{v_i^{(j)} w_{ij}}{\sum_i v_i^{(j)} w_{ij}} R_j$$

- ein und erhalten unsere finale z^B -Regel

$$R_i = \sum_j \frac{z_{ij} - l_i w_{ij}^+ - h_i w_{ij}^-}{\sum_i z_{ij} - l_i w_{ij}^+ - h_i w_{ij}^-} R_j$$

Anwendung auf den Pascal-Datensatz

- VGG16 trainiert für Multilabel-Klassifizierung für die Klassen *Mensch* und *Pferd*
- z^{β} -Regel wurde bei Anwendung aller Regeln für das Inputlayer verwendet
- LRP-Composition:
 - LRP-0 auf Dense-Layern
 - LRP- ϵ auf mittleren sechs Conv-Layern
 - LRP- γ auf den letzten sechs Conv-Layern vor Inputlayer

Aufgabe

- Vergleich LRP DTD theoretisch
- Ggf. Ausblick Min-Max?