Machine Learning 1 – Exercise 4

Machine Learning for Computer Vision
TU Dresden

Learning of composite functions (deep learning)

a) Prove the statement from the lecture that

$$\frac{\partial \alpha_{v\theta'}}{\partial \theta'_j} = \sum_{u \in A_v \setminus V} \frac{\partial g_{u\theta'}}{\partial \theta'_j} \sum_{(V',E') \in \mathcal{P}(u,v)} \prod_{(u',v') \in E'} \frac{\partial g_{v'\theta}}{\partial \alpha_{u'\theta}} \quad . \tag{1}$$

b) Consider a compute graph $(V, D, D', E, \Theta, \{g_{v\theta}\}_{v \in (D \cup D') \setminus V, \theta \in \Theta})$ such that

- $|D'| = 1$
- $D = V^{(1)} \cup V^{(2)}$
- For all $v \in V^{(1)}$: $P_v = V$. For all $v \in V^{(2)}$: $P_v = V^{(1)}$. For the single $v \in D'$: $P_v = V^{(2)}$.

Assume for any $v \in (D \cup D') \setminus V$ and any $u \in P_v$, the derivative $\frac{\partial g_{v\theta}}{\partial \alpha_{u\theta}}\Big|_{\alpha_{P_v\theta}(x)}$ is given. Calculate:

(a) The number of multiplications needed to compute $\Delta_{uv}(x, \theta)$ for $v \in D'$ and all $u \in V$ by means of forward recursion.

(b) The number of multiplications needed to compute $\Delta_{uv}(x, \theta)$ for $v \in D'$ and all $u \in V$ by means of backward recursion.

(c) The speed-up factor from using backward recursion instead of forward recursion.

(d) The speed-up factor under the assumption $|V^{(1)}| = \frac{2}{3}|V|$ and $|V^{(2)}| = \frac{1}{3}|V|$.

c) Let $(V, D, D', E, \Theta, \{g_{v\theta}\}_{v \in (D \cup D') \setminus V, \theta \in \Theta})$ a compute graph such that:

- $D = V^{(1)}$ and $D' = \{v_{\text{out}}\}$
- $E = (V \times V^{(1)}) \cup (V^{(1)} \times \{v_{\text{out}}\})$
- $\Theta = \mathbb{R}^E$
- For all $v \in (D \cup D') \setminus V$, all $\theta \in \Theta$ and all $x \in \mathbb{R}^V$: $g_{v\theta}(\alpha_{P_v\theta}(x)) = \sum_{u \in P_v} \theta_{uv}\, \alpha_{u\theta}(x)$

The function $f_\theta \colon \mathbb{R}^V \to \mathbb{R}^{\{v_{\text{out}}\}}$ defined by this compute graph is such that for all $x \in \mathbb{R}^V$:

$$f_\theta(x) = \sum_{v \in V} \sum_{v' \in V^{(1)}} \theta_{v' v_{\text{out}}}\, \theta_{vv'}\, x_v \quad . \tag{2}$$

Given the objective of the l_2-regularized non-linear logistic regression problem

$$\varphi(\theta) = \sum_{s \in S} \left(-y_s f_\theta(x_s) + \log\left(1 + 2^{f_\theta(x_s)}\right)\right) + \frac{\log e}{2\sigma^2} \|\theta\|_2^2 \quad . \tag{3}$$

i. Calculate the Hessian of φ, i.e. H^φ such that $H^\varphi_{ij} = \frac{\partial^2 \varphi}{\partial \theta_i \partial \theta_j}$, in terms of the gradient $\nabla_\theta f$ and the Hessian H^f of f. Recall that φ is convex in θ if $z^T H^\varphi z \geq 0$ for all $z \in \mathbb{R}^J$.

ii. Calculate the gradient $\nabla_\theta f$ and Hessian H^f for f, as in (2).