Computer Vision I

Bjoern Andres, Holger Heidrich, Jannik Presberger

Machine Learning for Computer Vision
TU Dresden

Winter Term 2023/2024

Notation. Let $G = (V, E)$ a digraph.

- For any $v \in V$, let

$$P_v = \{u \in V \mid (u, v) \in E\} \quad \text{the set of } \textbf{parents} \text{ of } v \tag{1}$$
$$C_v = \{w \in V \mid (v, w) \in E\} \quad \text{the set of } \textbf{children} \text{ of } v\ . \tag{2}$$

- For any $u, v \in V$, let $\mathcal{P}(u, v)$ denote the set of all uv-paths. (Any path is a subgraph. For any node u, the uu-path $(\{u\}, \emptyset)$ exists.)

Let G be **acyclic**.

- For any $v \in V$, let

$$A_v = \{u \in V \mid \mathcal{P}(u, v) \neq \emptyset\} \setminus \{v\} \quad \text{the set of } \textbf{ancestors} \text{ of } v \tag{3}$$
$$D_v = \{w \in V \mid \mathcal{P}(v, w) \neq \emptyset\} \setminus \{v\} \quad \text{the set of } \textbf{descendants} \text{ of } v\ . \tag{4}$$

Definition. A tuple $(V, D, D', E, \Theta, \{g_{v\theta} \colon \mathbb{R}^{P_v} \to \mathbb{R}\}_{v \in (D \cup D') \setminus V, \theta \in \Theta})$ is called a **compute graph**, iff the following conditions hold:

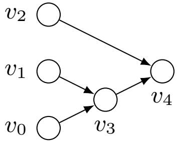
- $G = (V \cup D \cup D', E)$ is an acyclic digraph
- $\forall v \in V : P_v = \emptyset$
- $\forall v \in D' : C_v = \emptyset$
- $\forall v \in D : P_v \neq \emptyset$ and $C_v \neq \emptyset$

Definition. For any compute graph $(V, D, D', E, \Theta, \{g_{v\theta} \colon \mathbb{R}^{P_v} \to \mathbb{R}\}_{v \in (D \cup D') \setminus V, \theta \in \Theta})$, any $v \in V \cup D \cup D'$ and any $\theta \in \Theta$, let $\alpha_{v\theta} \colon \mathbb{R}^V \to \mathbb{R}$ such that for all $\hat{x} \in \mathbb{R}^V$:

$$\alpha_{v\theta}(\hat{x}) = \begin{cases} \hat{x}_v & \text{if } v \in V \\ g_{v\theta}(\alpha_{P_v\theta}(\hat{x})) & \text{otherwise} \end{cases} . \tag{5}$$

We call $\alpha_{v\theta}(\hat{x})$ the **activation** of v for **input** $\hat{x}$ and **parameters** θ. For any $\theta \in \Theta$ let $f_\theta \colon \mathbb{R}^V \to \mathbb{R}^{D'}$ such that $f_\theta = \alpha_{D'\theta}$. We call $f_\theta(\hat{x})$ the **output** of the compute graph for input $\hat{x}$ and parameters θ.

Example. Consider the compute graph below with $V = \{v_0, v_1, v_2\}$, $D = \{v_3\}$ and $D' = \{v_4\}$.



Moreover, consider $\Theta = \{\theta_0, \theta_1\}$ and

- $g_{v_3\theta} \colon \mathbb{R}^{\{v_0,v_1\}} \to \mathbb{R}$ such that $g_{v_3\theta}(x) = x_{v_0} + \theta_0 x_{v_1}$
- $g_{v_4\theta} \colon \mathbb{R}^{\{v_2,v_3\}} \to \mathbb{R}$ such that $g_{v_4\theta}(x) = x_{v_2} + x_{v_3}^{\theta_1}$

This defines the function $f_\theta(x) = x_{v_2} + (x_{v_0} + \theta_0 x_{v_1})^{\theta_1}$.

In the following:

- We assume $\Theta = \mathbb{R}^J$ for some set J.
- We consider compute graphs with $|D'| = 1$, i.e. $f_\theta(\hat{x}) \in \mathbb{R}$ for every $\hat{x} \in \mathbb{R}^V$.

Learning Problem

The l_2-**regularized non-linear logistic regression problem** with respect to labeled data $T = (S, \mathbb{R}^V, x, y)$ and $\sigma \in \mathbb{R}^+$ is to solve

$$\underset{\theta \in \mathbb{R}^J}{\operatorname{argmin}} \quad \frac{1}{|S|} \sum_{s \in S} \left(-y_s f_\theta(x_s) + \log \left(1 + 2^{f_\theta(x)} \right) \right) + \frac{\log e}{2\sigma^2} \|\theta\|^2 \quad . \tag{6}$$

Remark.

- The optimization problem (6) is analogous to linear logistic regression.
- The optimization problem (6) can be non-convex for non-linear f_θ.
- A local minimum $\hat{\theta} \in \mathbb{R}^J$ can be found by means of a steepest descent algorithm. We describe two techniques, **forward propagation** and **backward propagation**, for computing $\nabla_\theta f_\theta$.

Lemma. Let $j \in J$. For any $v \in V$: $\frac{\partial \alpha_{v\theta}}{\partial \theta_j} = 0$. For any $v \in (D \cup D') \setminus V$:

$$\frac{\partial \alpha_{v\theta}}{\partial \theta_j} = \sum_{u \in (A_v \cup \{v\}) \setminus V} \frac{\partial g_{u\theta}}{\partial \theta_j} \Delta_{uv} \tag{7}$$

with

$$\Delta_{uv} := \sum_{(V',E') \in \mathcal{P}(u,v)} \prod_{(u',v') \in E'} \frac{\partial g_{v'\theta}}{\partial \alpha_{u'\theta}} \; . \tag{8}$$

Remark. For any node u: $\Delta_{uu} = 1$. For any u, v with $\mathcal{P}(u,v) = \emptyset$: $\Delta_{uv} = 0$.

Proof (idea).

$$\begin{aligned}
\frac{\partial \alpha_{v\theta}}{\partial \theta_j} &= \frac{\partial g_{v\theta}}{\partial \theta_j} + \sum_{u \in P_v} \frac{\partial g_{v\theta}}{\partial \alpha_{u\theta}} \frac{\partial \alpha_{u\theta}}{\partial \theta_j} \qquad (9)\\
&= \frac{\partial g_{v\theta}}{\partial \theta_j} + \sum_{u \in P_v} \frac{\partial g_{v\theta}}{\partial \alpha_{u\theta}} \frac{\partial g_{u\theta}}{\partial \theta_j} + \sum_{u \in P_v} \sum_{u' \in P_u} \frac{\partial g_{v\theta}}{\partial \alpha_{u\theta}} \frac{\partial g_{u\theta}}{\partial \alpha_{u'\theta}} \frac{\partial \alpha_{u'\theta}}{\partial \theta_j}\\
&= \text{repeated application (9)}\\
&= \sum_{u \in (A_v \cup \{v\}) \setminus V} \frac{\partial g_{u\theta}}{\partial \theta_j} \sum_{(V',E') \in \mathcal{P}(u,v)} \prod_{(u',v') \in E'} \frac{\partial g_{v'\theta}}{\partial \alpha_{u'\theta}}
\end{aligned}$$

Lemma (backward propagation). For all nodes $u \neq w$ such that $\mathcal{P}(u, w) \neq \emptyset$:

$$\Delta_{uw} = \sum_{v \in C_u} \frac{\partial g_{v\theta}}{\partial \alpha_{u\theta}} \Delta_{vw} \tag{10}$$

Proof.

$$\begin{aligned}
\Delta_{uw} &= \sum_{(V',E') \in \mathcal{P}(u,w)} \prod_{(u',v') \in E'} \frac{\partial g_{v'\theta}}{\partial \alpha_{u'\theta}} \\
&= \sum_{v \in C_u} \sum_{(V'',E'') \in \mathcal{P}(v,w)} \prod_{(u',v') \in E'' \cup \{(u,v)\}} \frac{\partial g_{v'\theta}}{\partial \alpha_{u'\theta}} \\
&= \sum_{v \in C_u} \frac{\partial g_{v\theta}}{\partial \alpha_{u\theta}} \sum_{(V'',E'') \in \mathcal{P}(v,w)} \prod_{(u',v') \in E''} \frac{\partial g_{v'\theta}}{\partial \alpha_{u'\theta}} \\
&= \sum_{v \in C_u} \frac{\partial g_{v\theta}}{\partial \alpha_{u\theta}} \Delta_{vw}
\end{aligned}$$

$\square$

The **backward propagation algorithm** computes Δ_{uw} for one node w and all nodes u. It is defined wrt. an arbitrary partial order $<_C$ of the nodes such that

$$\forall u \in V \cup D \quad \forall v \in C_u \colon \quad v <_C u \ . \tag{11}$$

Input:
Compute graph $(V, D, D', E, \Theta, \{g_{v\theta} \colon \mathbb{R}^{P_v} \to \mathbb{R}\}_{v \in (D \cup D') \setminus V, \theta \in \Theta})$
Node $w \in V \cup D \cup D'$

---

for u ordered by $<_C$ (11)
    if $u = w$
        $\Delta_{uw} := 1$
    else if $\mathcal{P}(u, w) = \emptyset$
        $\Delta_{uw} := 0$
    else
        $\Delta_{uw} := \sum_{v \in C_u} \frac{\partial g_{v\theta}}{\partial \alpha_{u\theta}} \Delta_{vw}$ (10)
