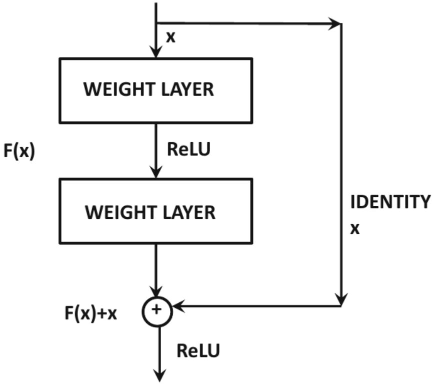
Tomáš Oberhuber

Faculty of Nuclear Sciences and Physical Engineering
Czech Technical University in Prague

Reziduální sítě



- viděli jsme síť ResNet využívá tzv. *skip connections*
- zkratka obcházející jednu vrstvu má význam identity

Reziduální sítě

- aktivace jednotlivých vrstev pak lze popsat jako

$$\vec{x}^{(l+1)} = \vec{x}^{(l)} + F(\vec{x}^{(l)})$$

- to lze přepsat do tvaru

$$\vec{x}^{(l+1)} = \vec{x}^{(l)} + \Delta t \cdot F(\vec{x}^{(l)}),$$

  pro $\Delta t = 1$
- tento tvar odpovídá Eulerově metodě pro řešení ODR

$$\frac{\mathrm{d}\vec{x}}{\mathrm{d}t} = F(\vec{x}, t)$$

E. Weinan, A proposal on machine learning via dynamical systems, Communications in Mathematics and Statistics, vol. 5, no. 1, pp. 1–11, 2017

T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud, “Neural ordinary differential equations,” in Advances in Neural Information Processing Systems, pp. 6571–6583, 2018

Optimalizace s ODR

- místo učení NN se tak můžeme zabývat úlohou hledání vhodných parametrů pro ODR tak, aby např. generovala předepsaná data
- vlastně tak hledáme matematický model pro generování experimentálních dat
- jde o úlohu zvanou *physics from data* - Youtube

Optimalizace s ODR

Mějme ODR tvaru

$$\begin{aligned} \dot{\vec{u}}(t,\vec{p}) &= \vec{f}(\vec{u}(t,\vec{p}),t,\vec{p}) \text{ na } (0,T), & (1)\\ \vec{u}(0,\vec{p}) &= \vec{u}_0(\vec{p}), & (2) \end{aligned}$$

kde

- $\vec{p}$ je vektor parametrů $\vec{p} \in P \equiv \mathbb{R}^m$
- $\vec{u}(t,\vec{p}) : \langle 0,T \rangle \times P \to \mathbb{R}^n$ je neznámá funkce
- $\vec{f}(\vec{u}(t,\vec{p}),t,\vec{p}) : \mathbb{R}^n \times \langle 0,T \rangle \times P \to \mathbb{R}^n$ je daná pravá strana

Optimalizace s ODR

Mějme zobrazení

- $h_1(\vec{u}(t,\vec{p}),t,p) : \mathbb{R}^n \times \langle 0, T\rangle \times P \rightarrow \mathbb{R}$,
- $h_2(\vec{u}(T,\vec{p}),p) : \mathbb{R}^n \times P \rightarrow \mathbb{R}$.

Pak můžeme definovat funkcionál J jako

$$J(\vec{p}) = \int_0^T h_1(\vec{u}(t,\vec{p}),t,p)dt + h_2(\vec{u}(T,\vec{p}),\vec{p})$$

a řešit úlohu

$$\vec{p}^* = \arg\min_{p \in P} J(\vec{p}).$$

Optimalizace s ODR

Example 1

- předpokládejme, že máme zadaná (experimentální) data $\vec{d}(t)$
- volbou

$$
\begin{aligned}
h_1(\vec{u}(t,\vec{p}),t,p) &= \left|\vec{u}(t,\vec{p}) - \vec{d}(t)\right|^2, \\
h_2(\vec{u}(T,\vec{p}),p) &= 0
\end{aligned}
$$

- dostáváme funkcionál

$$
J(\vec{p}) = \int_0^T \left|\vec{u}(t,\vec{p}) - \vec{d}(t)\right|^2 dt
$$

- a hledáme parametry $\vec{p}^*$ tak, že ODR (1) s pravou stranou $\vec{f}(\vec{u}(t,\vec{p}^*),t,\vec{p}^*)$ generuje naše data.

Optimalizace s ODR

Example 2

- mějme zadaný finální stav $\vec{d}(T)$
- volbou

$$\begin{aligned} h_1(\vec{u}(t,\vec{p}),t,p) &= 0, \\ h_2(\vec{u}(T,\vec{p}),p) &= \left|\vec{u}(T,\vec{p}) - \vec{d}(T)\right|^2 \end{aligned}$$

- dostáváme funkcionál

$$J(\vec{p}) = \left|\vec{u}(T,\vec{p}) - \vec{d}(T)\right|^2$$

- a hledáme parametry $\vec{p}^*$ tak, že počáteční podmínka (2) vede k finálnímu stavu $\vec{d}(T)$, tj. můžeme řešit *inverzní úlohu* k úloze (1)-(2)

Optimalizace s ODR

- pro nalezení optimálních parametrů $\vec{p}^*$ opět potřebujeme spočítat

$$\nabla_{\vec{p}} J(\vec{p})$$

- to lze udělat třemi různými způsoby
  1. přímým výpočtem pomocí malých perturbací parametrů $\vec{p}$
     - pro větší počet parametrů to může být výpočetně velmi náročné
  2. algoritmicky pomocí tzv. *automatic differentation* - Julia
     - u ODR je to v současnosti často nejjednodušší přístup
  3. odvozením tzv. *adjoint rovnice* - obdoba backpropagation

Optimalizace s ODR

- pro odvození adjoint rovnice použijeme Lagrangeovy multiplikátory
- definujeme funkcional $J^*(\vec{p})$ jako

$$J^*(\vec{p}) = \int_0^T h_1 + \vec{\lambda}_1 \left(\dot{\vec{u}} - \vec{f} \right) dt + h_2 + \vec{\lambda}_2 (\vec{u}\,|_{t=0} - \vec{u}_0)$$

kde $\vec{\lambda}_{1,2} \in C(\langle 0, T \rangle; \mathbb{R}^n)$ jsou Lagrangeovy multiplikátory

- pokud $\vec{u}$ je řešení ODR (1)-(2) potom

$$\begin{aligned} \dot{\vec{u}} - \vec{f} &= 0 \text{ na } (0, T), \\ \vec{u}\,|_{t=0} - \vec{u}_0 &= 0, \end{aligned}$$

a $J(\vec{p}) = J^*(\vec{p})$ platí pro všechna $\vec{\lambda}_{1,2}$

Optimalizace s ODR

- gradient J^* má tvar

$$
\begin{aligned}
\nabla_{\vec{p}} J^*(\vec{p}) \quad = \quad & \int_0^T \frac{\partial h_1}{\partial \vec{u}} \frac{\partial \vec{u}}{\partial \vec{p}} + \frac{\partial h_1}{\partial \vec{p}} + \vec{\lambda}_1 \left(\frac{\partial \dot{\vec{u}}}{\partial \vec{p}} - \frac{\partial \vec{f}}{\partial \vec{u}} \frac{\partial \vec{u}}{\partial \vec{p}} - \frac{\partial \vec{f}}{\partial \vec{p}} \right) dt + \\
& \frac{\partial h_2}{\partial \vec{u}} \frac{\partial \vec{u}}{\partial \vec{p}} |_{t=T} + \frac{\partial h_2}{\partial \vec{p}} |_{t=T} + \vec{\lambda}_2 \left(\frac{\partial \vec{u}}{\partial \vec{p}} |_{t=0} + \frac{\partial \vec{u}_0}{\partial \vec{p}} \right)
\end{aligned}
$$

- výpočet $\partial \vec{u} / \partial \vec{p}$ a $\partial \dot{\vec{u}} / \partial \vec{p}$ by vyžadoval opětovné řešení původní ODR a postupnými perturbacemi $\vec{p}$
- těchto členů se chceme zbavit

Optimalizace s ODR

- začneme s výrazem $\partial \dot{\vec{u}} / \partial \vec{p}$

$$
\begin{aligned}
\int_0^T \vec{\lambda}_1 \frac{\partial \dot{\vec{u}}}{\partial \vec{p}} dt &= \int_0^T \vec{\lambda}_1 \frac{\partial}{\partial \vec{p}} \frac{\partial \vec{u}}{\partial t} dt \\
&= \int_0^T \vec{\lambda}_1 \frac{\partial}{\partial t} \frac{\partial \vec{u}}{\partial \vec{p}} dt \\
&= -\int_0^T \frac{\partial \vec{\lambda}_1}{\partial t} \frac{\partial \vec{u}}{\partial \vec{p}} dt + \left[\vec{\lambda}_1 \frac{\partial \vec{u}}{\partial \vec{p}} \right]_0^T \\
&= -\int_0^T \frac{\partial \vec{\lambda}_1}{\partial t} \frac{\partial \vec{u}}{\partial \vec{p}} dt + \vec{\lambda}_1 \frac{\partial \vec{u}}{\partial \vec{p}} \Big|_{t=T} - \vec{\lambda}_1 \frac{\partial \vec{u}}{\partial \vec{p}} \Big|_{t=0},
\end{aligned}
$$

Optimalizace s ODR

- a tak dostaneme

$$\nabla_{\vec{p}} J^*(\vec{p}) = \int_0^T \left(\frac{\partial h_1}{\partial \vec{u}} - \dot{\vec{\lambda}}_1 - \vec{\lambda}_1 \frac{\partial \vec{f}}{\partial \vec{u}} \right) \frac{\partial \vec{u}}{\partial \vec{p}} + \frac{\partial h_1}{\partial \vec{p}} - \vec{\lambda}_1 \frac{\partial \vec{f}}{\partial \vec{p}} dt +$$
$$\left(\vec{\lambda}_1 + \frac{\partial h_2}{\partial \vec{u}} \right) \frac{\partial \vec{u}}{\partial \vec{p}} \Big|_{t=T} + \frac{\partial h_2}{\partial \vec{p}} \Big|_{t=T} + \left(\vec{\lambda}_2 - \vec{\lambda}_1 \right) \frac{\partial \vec{u}}{\partial \vec{p}} \Big|_{t=0} + \vec{\lambda}_2 \frac{\partial \vec{u}_0}{\partial \vec{p}}$$

Optimalizace s ODR

- a splněním následujících vztahů

$$\begin{aligned}
\frac{\partial h_1}{\partial \vec{u}} - \dot{\vec{\lambda}}_1 - \vec{\lambda}_1 \frac{\partial \vec{f}}{\partial \vec{u}} &= 0, \quad (3)\\
\left(\vec{\lambda}_1 + \frac{\partial h_2}{\partial \vec{u}} \right)_{t=T} &= 0, \quad (4)\\
\left(\vec{\lambda}_2 - \vec{\lambda}_1 \right) \Big|_{t=0} &= 0. \quad (5)
\end{aligned}$$

- můžeme nulovat výrazy před $\frac{\partial \vec{u}}{\partial \vec{p}}$

Optimalizace s ODR

- (3)-(4) lze splnit řešením adjoint ODR

$$
\begin{aligned}
-\frac{\partial \vec{\lambda}_1}{\partial t} &= \vec{\lambda}_1 \frac{\partial \vec{f}}{\partial \vec{u}} - \frac{\partial h_1}{\partial \vec{u}} \text{ na } (0, T), & (6)\\
\vec{\lambda}_1 \Big|_{t=T} &= -\frac{\partial h_2}{\partial \vec{u}} \Big|_{t=T} & (7)
\end{aligned}
$$

- a následně nastavit

$$
\vec{\lambda}_2|_{t=0} = \vec{\lambda}_1 \Big|_{t=0} .
$$

- uvedená ODR má počáteční podmínku zadanou v bodu $t = T$

Optimalizace s ODR

- provedeme transformaci

$$\vec{\lambda}_1^*(t) = \vec{\lambda}_1(T-t) \text{ for } t \in \langle 0, T \rangle$$

- a vidíme, že

$$\frac{\partial \vec{\lambda}_1^*}{\partial t} = -\frac{\partial \vec{\lambda}_1}{\partial t} \quad \text{a} \quad \vec{\lambda}_1^*(0) = \vec{\lambda}_1(T)$$

- adjoint rovnici (6)-(7) lze tedy transformovat na tvar

$$\frac{\partial \vec{\lambda}_1^*}{\partial t} = \vec{\lambda}_1^* \frac{\partial \vec{f}(\vec{u}(T-t,\vec{p}), T-t, \vec{p})}{\partial \vec{u}} - \frac{\partial h_1(T-t)}{\partial \vec{u}} \text{ na } (0,T), \tag{8}$$

$$\vec{\lambda}_1^* \Big|_{t=0} = -\frac{\partial h_2}{\partial \vec{u}} \tag{9}$$