

Tomáš Oberhuber

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

Fitování a učení modelů pomocí optimalizací

- obecně se budeme zabývat úlohou, kdy se snažíme najít zobrazení

$$\vec{F} : X_1 \rightarrow X_2,$$

za předpokladu, že známe

$$\vec{F}(\vec{x}_i) = \vec{y}_i \text{ pro } i = 1, \dots, n$$

- tuto úlohu lze řešit jako optimalizační úlohu

$$\vec{F} = \arg \min_{F(\vec{\theta}, \cdot)} L(\vec{F}(\vec{\theta}, \vec{x}_1), \dots, \vec{F}(\vec{\theta}, \vec{x}_n), y_1, \dots, y_n)$$

- F není zcela neznámé, je ale dáno až na sadu parametrů $\vec{\theta}$, které je potřeba najít
- F reprezentuje model, který chceme použít řešení dané úlohy

Fitování a učení modelů pomocí optimalizací

- **fitování modelu**

- jde o situaci, kdy víme poměrně dobře, jaký model použít
- většinou pak hledáme menší počet parametrů θ

- **učení modelu**

- jde o situaci, kdy ani nevíme, jaký model máme pro danou úlohu použít
- použijeme některý obecnější model, často s větším počtem parametrů
- doufáme, že obecnější model se správně nastaví na naši úlohu
- to testujeme pomocí dat, která jsme nepoužili pro nastavování parametrů θ
- jde o tzv. *generalizaci modelu*

Fitování a učení modelů pomocí optimalizací

Úlohu

$$\vec{\theta}^* = \arg \min_{\vec{\theta}} L(\vec{F}(\vec{\theta}, \vec{x}_1), \dots, \vec{F}(\vec{\theta}, \vec{x}_n), y_1, \dots, y_n)$$

převédeme na

$$\nabla_{\vec{\theta}} L(\vec{F}(\vec{\theta}, \vec{x}_1), \dots, \vec{F}(\vec{\theta}, \vec{x}_n), y_1, \dots, y_n) = \vec{0}$$

a většinou ji řešíme numericky pomocí **metody největšího spádu** nebo jiné optimalizační metody, tj.

$$\vec{\theta}_{i+1} = \vec{\theta}_i - \nabla_{\vec{\theta}} L(\vec{F}(\vec{\theta}_i, \vec{x}_1), \dots, \vec{F}(\vec{\theta}_i, \vec{x}_n), y_1, \dots, y_n).$$

Fitování a učení modelů pomocí optimalizací

Celý proces se tak skládá z následujících kroků:

- volba vhodného modelu $F(\vec{\theta}, \cdot)$
 - SVM, DNN, CNN, RNN, SPTN, ODE, PDE, ...
- volba vhodné tzv. ztrátové funkce L
 - L_1 , L_2 , cross entropy, KL-divergence
- minimalizace
 - GD, SGD, ADAM, BFGS
- výpočet gradientu $\nabla_{\theta} L$
 - backpropagation, adjoint, automatic differentiation