Projekce na kouli v zadané metrice

Pavel Jakš

7. března 2023

Úvod

V oblasti generování adversariálních vzorků, kterými se zabývá tato práce, se v různých algoritmech v hojnosti vyskytuje potřeba tzv. projektovat nějaký vektor do nějakého ϵ-okolí jiného vektoru. To znamená pro vektor $\tilde{x}$ najít vektor $\tilde{x}^*$ v ϵ-okolí vektoru x, takový že nejlépe odpovídá vektoru $\tilde{x}$, tedy je ze všech vektorů v ϵ-okolí x nejblíže vektoru $\tilde{x}$.

Mějme metriku $d : \mathbb{R}^n \times \mathbb{R}^n \to [0, +\infty)$. Označme jako kouli:

$$B^{(d)}(x, \epsilon) = \{\hat{x} \in \mathbb{R}^n | d(x, \hat{x}) \leq \epsilon\}. \tag{1}$$

Poznamenejme, že od standardní definice koule se tato definice liší v tom, že nerovnost definující kouli je neostrá, tedy dle standardní definice otevřené koule jsme jako kouli definovali uzávěr otevřené koule. Formálně bychom tyto myšlenky o projekci mohli zapsat následovně: Projekce je zobrazení $P_{B(x,\epsilon)}$, které pro $\tilde{x}$ má předpis:

$$P_{B^{(d)}(x,\epsilon)}(\tilde{x}) = \operatorname*{argmin}_{\hat{x} \in B^{(d)}(x,\epsilon)} d(\tilde{x}, \hat{x}). \tag{2}$$

Otázka zní: Musíme v projekci mít stejnou metriku definující danou kouli jako máme v minimalizačním problému? Nemusíme, a budeme toho využívat. Proto mějme dvě metriky d_1, d_2 na $\mathbb{R}^n$ a zobecněme definici projekce na následující:

$$P^{(d_2)}_{B^{(d_1)}(x,\epsilon)}(\tilde{x}) = \operatorname*{argmin}_{\hat{x} \in B^{(d_1)}(x,\epsilon)} d_2(\tilde{x}, \hat{x}). \tag{3}$$

V následujících částech textu položme ve vztahu 3 za metriku d_2 metriku indukovanou l_2 normou. Dále poznamenejme, že umocnění l_2 normy rozdílu na druhou nemění řešení optimalizačního problému projekce. Proto vezměme za projekci následující:

$$P_{B^{(d)}(x,\epsilon)}(\tilde{x}) = \operatorname*{argmin}_{\hat{x} \in B^{(d)}(x,\epsilon)} \|\tilde{x} - \hat{x}\|_2^2, \tag{4}$$

kde d je metrika na $\mathbb{R}^n$.

1 Projekce na kouli zadanou Wassersteinovou metrikou

Reference

[1] E. Wong, F. R. Schmidt, J. Z. Kolter, *Wasserstein Adversarial Examples via Projected Sinkhorn Iterations.* Proceedings of the 36th International Conference on Machine Learning, PMLR 97:6808-6817, 2019.

[2] L. Vaserstein, *Markov processes over denumerable products of spaces, describing large systems of automata.* Problemy Peredači Informacii 5, 1969.