Non-parametric Models for Non-negative Functions

Ulysse Marteau-Ferey, Francis Bach, Alessandro Rudi

NeurIPS 2020, December 2020

Sierra project-team – INRIA Paris – Département d'informatique de l'ENS Paris

Prototypical machine learning task

Goal : find $f_\star : \mathcal{X} \to \mathbb{R}$ using n training points $(x_i)_{1 \leqslant i \leqslant n}$.

$$f_\star \in \arg\min_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^{n} \ell(f(x_i)) + \Omega(f) \quad (1)$$

Prototypical machine learning task

Goal : find $\theta_\star \in \mathcal{H}$ using n training points $(x_i)_{1 \leqslant i \leqslant n}$.

$$\theta_\star \in \arg\min_{\theta \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^{n} \ell(\theta^\top \varphi(x_i)) + \Omega(\theta) \quad (1)$$

Linear Models

Features : $\varphi(x) \in \mathcal{H}$ (built features, kernels...)

Parametrization : by a vector $\theta \in \mathcal{H}$, $f_\theta : \mathcal{X} \to \mathbb{R}$

$$f_\theta(x) = \theta^\top \varphi(x), \ \theta \in \mathcal{H}$$

Prototypical machine learning task

Goal : find $\theta_\star \in \mathcal{H}$ using n training points $(x_i)_{1 \leqslant i \leqslant n}$.

$$\theta_\star \in \arg\min_{\theta \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^{n} \ell(\theta^\top \varphi(x_i)) + \Omega(\theta) \tag{1}$$

Linear Models have good properties

- **preserve convexity** of ℓ
- **rich classes of functions** when $\mathcal{H}$ infinite dimensional (kernel methods)
- finite dimensional representation with $\boldsymbol{n}$ **degrees of freedom**:

$$\theta_\star = \sum_{i=1}^{n} \alpha_i \varphi(x_i)$$

Proposed model for non-negative functions

What if we want $f \geqslant 0$?

$$f_\star \in \arg\min_{\substack{f \in \mathcal{F} \\ f \geqslant 0}} \frac{1}{n} \sum_{i=1}^{n} \ell(f(x_i)) + \Omega(f) \tag{2}$$

Linear models do not work anymore !

Proposed model for non-negative functions

What if we want $f \geqslant 0$?

$$f_\star \in \arg\min_{\substack{f \in \mathcal{F} \\ f \geqslant 0}} \frac{1}{n} \sum_{i=1}^{n} \ell(f(x_i)) + \Omega(f) \tag{2}$$

Classical models lack crucial properties

- (2) becomes **non-convex**;
- $f_\star \geqslant 0$ **only on a grid**;
- Poor approximation properties.

Proposed model for non-negative functions

What if we want $f \geqslant 0$?

$$f_\star \in \arg\min_{\substack{f \in \mathcal{F} \\ f \geqslant 0}} \frac{1}{n} \sum_{i=1}^{n} \ell(f(x_i)) + \Omega(f) \tag{2}$$

We propose a model for non-negative functions :

$f_A(x) = \varphi(x)^\top A \varphi(x)$, A symmetric matrix $A \succeq 0$

$$\boxed{A \succeq 0 \implies f_A \geqslant 0}$$

Proposed model for non-negative functions

What if we want $f \geqslant 0$?

$$A_\star \in \arg \min_{\substack{A \in \mathcal{S}(\mathcal{H}) \\ A \succeq 0}} \frac{1}{n} \sum_{i=1}^{n} \ell(\varphi(x_i)^\top A \varphi(x_i)) + \Omega(A) \tag{2}$$

We propose a model for non-negative functions :

$$f_A(x) = \varphi(x)^\top A \varphi(x), \; A \; \text{ symmetric matrix } \; A \succeq 0$$

$$\boxed{A \succeq 0 \implies f_A \geqslant 0}$$

The proposed model keeps the interesting properties

$$A_\star \in \arg \min_{\substack{A \in \mathcal{S}(\mathcal{H}) \\ A \succeq 0}} \frac{1}{n} \sum_{i=1}^{n} \ell(\varphi(x_i)^\top A \varphi(x_i)) + \Omega(A) \tag{2}$$

We prove that it has all the good properties of linear models:

- (2) is convex
- approximation properties **match those of linear models**, **$\mathcal{H}$ infinite dimensional**
- **finite dimensional representation** with **n^2 parameters**:

$$A_\star = \sum_{i,j=1}^{n} a_{ij}\ \varphi(x_i)\varphi(x_j)^\top$$

- **dual representation** using only **n parameters**;
- statistical complexity **matches that of linear models**
- and many more ... **Check the paper !**