Learning Safe Prediction for Semi-Supervised Regression
Supplemental Materials

Yu-Feng Li and **Han-Wen Zha** and **Zhi-Hua Zhou**
National Key Laboratory for Novel Software Technology
Collaborative Innovation Center of Novel Software Technology and Industrialization
Nanjing University, Nanjing 210023, China
{liyf,zhouzh}@lamda.nju.edu.cn; zhahw12@gmail.com

Abstract

This file contains proofs of theorems 1-3.

Proofs of Theorem

Theorem 1. $\|\bar{\mathbf{f}} - \mathbf{f}^*\|^2 \leq \|\mathbf{f}_0 - \mathbf{f}^*\|^2$, *if the ground truth* $\mathbf{f}^* \in \Omega = \{\mathbf{f} | \sum_{i=1}^b \alpha_i \mathbf{f}_i, \boldsymbol{\alpha} \in \mathcal{M}\}$.

The proof can be derived following Pythagorean Theorem (theorem 2.4.1 in (Censor and Zenios 1997)).

Theorem 2. $\bar{\mathbf{f}}$ *has already achieved the maximal worst-case performance gain against* $\mathbf{f}_0$*, if the ground truth* $\mathbf{f}^* \in \Omega$.

Proof. The goal is to show $\bar{\mathbf{f}}$ is the optimal solution of the following functional,

$$\bar{\mathbf{f}} = \operatorname*{argmax}_{\mathbf{f} \in \mathbb{R}^u} \min_{\mathbf{f}^* \in \Omega} \Big(\|\mathbf{f}_0 - \mathbf{f}^*\|^2 - \|\mathbf{f} - \mathbf{f}^*\|^2 \Big) \tag{1}$$

Note that Eq.(1) is equivalent to the follows,

$$\max_{\mathbf{f}} \min_{\mathbf{f}^* \in \Omega} \Big(\|\mathbf{f}_0\|^2 - \|\mathbf{f}\|^2 - 2(\mathbf{f} - \mathbf{f}_0)^\top \bar{\mathbf{f}}^* \Big) \tag{2}$$

Eq.(2) is convex to $\mathbf{f}$ and concave to $\mathbf{f}^*$, and thus it is convex. Furthermore, by setting to derivative w.r.t. $\mathbf{f}$ to zero, it can be found that $\mathbf{f}$ has a closed-form solution, i.e., $\mathbf{f} = \mathbf{f}^*$. Substituting such an equality into Eq.(1), Eq.(1) turns out to be following functional w.r.t. $\mathbf{f}^*$ (or equivalently $\mathbf{f}$) only,

$$\bar{\mathbf{f}} = \operatorname*{argmin}_{\mathbf{f} \in \Omega} \Big(\|\mathbf{f}_0 - \mathbf{f}\|^2 \Big) \tag{3}$$

Eq.(3) is exactly the same as the projection problem proposed in the paper. Therefore, $\bar{\mathbf{f}}$ is the optimal solution of Eq.(1) and hence Theorem 2 holds. □

Theorem 3. *The increased loss of the proposal against* $\mathbf{f}_0$*, i.e.,* $\frac{1}{u}\Big(\|\bar{\mathbf{f}} - \mathbf{f}^*\|^2 - \|\mathbf{f}_0 - \mathbf{f}^*\|^2\Big)$*, is at most* $\min\{2\|\boldsymbol{\epsilon}\|_1/u, 2\|\boldsymbol{\epsilon}\|_2/\sqrt{u}\}$.

Proof. Note that $\sum_{i=1}^b \lambda_i^* \mathbf{f}_i \in \Omega$ and thus by employing Theorem 1, one can have,

$$\Big(\|\bar{\mathbf{f}} - \sum_{i=1}^b \lambda_i^* \mathbf{f}_i\|^2 - \|\mathbf{f}_0 - \sum_{i=1}^b \lambda_i^* \mathbf{f}_i\|^2 \Big) \leq 0$$

and it is consequently rewritten as

$$\Big(\big(-\|\mathbf{f}_0\|^2 + \|\bar{\mathbf{f}}\|^2 + 2(\mathbf{f}_0 - \bar{\mathbf{f}})^\top \sum_{i=1}^b \lambda_i^* \mathbf{f}_i \big) \Big) \leq 0$$

Since $\mathbf{f}^* = \sum_{i=1}^b \lambda_i^* \mathbf{f}_i + \boldsymbol{\epsilon}$, we then have,

$$\Big(\|\bar{\mathbf{f}} - \mathbf{f}^*\|^2 - \|\mathbf{f}_0 - \mathbf{f}^*\|^2 \Big) \leq 2(\mathbf{f}_0 - \bar{\mathbf{f}})^\top \boldsymbol{\epsilon}$$

and consequently we have

$$\frac{1}{u}\Big(\|\bar{\mathbf{f}} - \mathbf{f}^*\|^2 - \|\mathbf{f}_0 - \mathbf{f}^*\|^2 \Big) \leq \frac{2}{u}(\mathbf{f}_0 - \bar{\mathbf{f}})^\top \boldsymbol{\epsilon} \tag{4}$$

where the LHS refers to increased loss against $\mathbf{f}_0$. Further note that

$$2(\mathbf{f}_0 - \bar{\mathbf{f}})^\top \boldsymbol{\epsilon} \leq \min\{2\|\boldsymbol{\epsilon}\|_1, 2\sqrt{u}\|\boldsymbol{\epsilon}\|_2\} \tag{5}$$

using the fact that the predictive values in $\mathbf{f}_0$ and $\bar{\mathbf{f}}$ are from $[0, 1]$. With Eqs.(4)-(5), Theorem 3 holds. □

References

Censor, Y., and Zenios, S. A. 1997. *Parallel optimization: Theory, algorithms, and applications*. Oxford University Press.