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. \( \| \bar{f} - f^* \|^2 \leq \| f_0 - f^* \|^2 \), if the ground truth \( f^* \in \Omega = \{ f : \sum_{i=1}^b \alpha_i f_i, \alpha \in \mathcal{M} \} \).

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

Theorem 2. \( \bar{f} \) has already achieved the maximal worst-case performance gain against \( f_0 \), if the ground truth \( f^* \in \Omega \).

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

\[
\bar{f} = \arg \max_{f \in \mathbb{R}^n} \min_{f^* \in \Omega} \left( \| f_0 - f^* \|^2 - \| f - f^* \|^2 \right) \tag{1}
\]

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

\[
\max_f \min_{f^* \in \Omega} \left( \| f_0 \|^2 - \| f \|^2 - 2(f - f_0)^\top f^* \right) \tag{2}
\]

Eq.(2) is convex to \( f \) and concave to \( f^* \), and thus it is convex. Furthermore, by setting to derivative w.r.t. \( f \) to zero, it can be found that \( f \) has a closed-form solution, i.e., \( f = f^* \).

Substituting such an equality into Eq.(1), Eq.(1) turns out to be following functional w.r.t. \( f^* \) (or equivalently \( f \) only),

\[
\bar{f} = \arg \min_{f \in \Omega} \left( \| f_0 - f \|^2 \right) \tag{3}
\]

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

Theorem 3. The increased loss of the proposal against \( f_0 \), i.e., \( \frac{1}{u} \left( \| \bar{f} - f^* \|^2 - \| f_0 - f^* \|^2 \right) \), is at most \( \min \{ 2\| \epsilon \|_1, 2\| \epsilon \|_2 / \sqrt{u} \} \).

Proof. Note that \( \sum_{i=1}^b \lambda_i^* f_i \in \Omega \) and thus by employing Theorem 1, one can have,

\[
\left( \| \bar{f} - \sum_{i=1}^b \lambda_i^* f_i \|^2 - \| f_0 - \sum_{i=1}^b \lambda_i^* f_i \|^2 \right) \leq 0
\]

and it is consequently rewritten as

\[
\left( -\| f_0 \|^2 + \| \bar{f} \|^2 + 2(f_0 - \bar{f})^\top \sum_{i=1}^b \lambda_i^* f_i \right) \leq 0
\]

Since \( f^* = \sum_{i=1}^b \lambda_i^* f_i + \epsilon \), we then have,

\[
\left( \| \bar{f} - f^* \|^2 - \| f_0 - f^* \|^2 \right) \leq 2(f_0 - \bar{f})^\top \epsilon
\]

and consequently we have

\[
\frac{1}{u} \left( \| f_0 - f^* \|^2 - \| f_0 - f^* \|^2 \right) \leq \frac{2}{u} (f_0 - \bar{f})^\top \epsilon \tag{4}
\]

where the LHS refers to increased loss against \( f_0 \). Further note that

\[
2(f_0 - \bar{f})^\top \epsilon \leq \min \{ 2\| \epsilon \|_1, 2\sqrt{u}\| \epsilon \|_2 \} \tag{5}
\]

using the fact that the predictive values in \( f_0 \) and \( \bar{f} \) are from \([0, 1]\). With Eqs.(4)-(5), Theorem 3 holds.

References