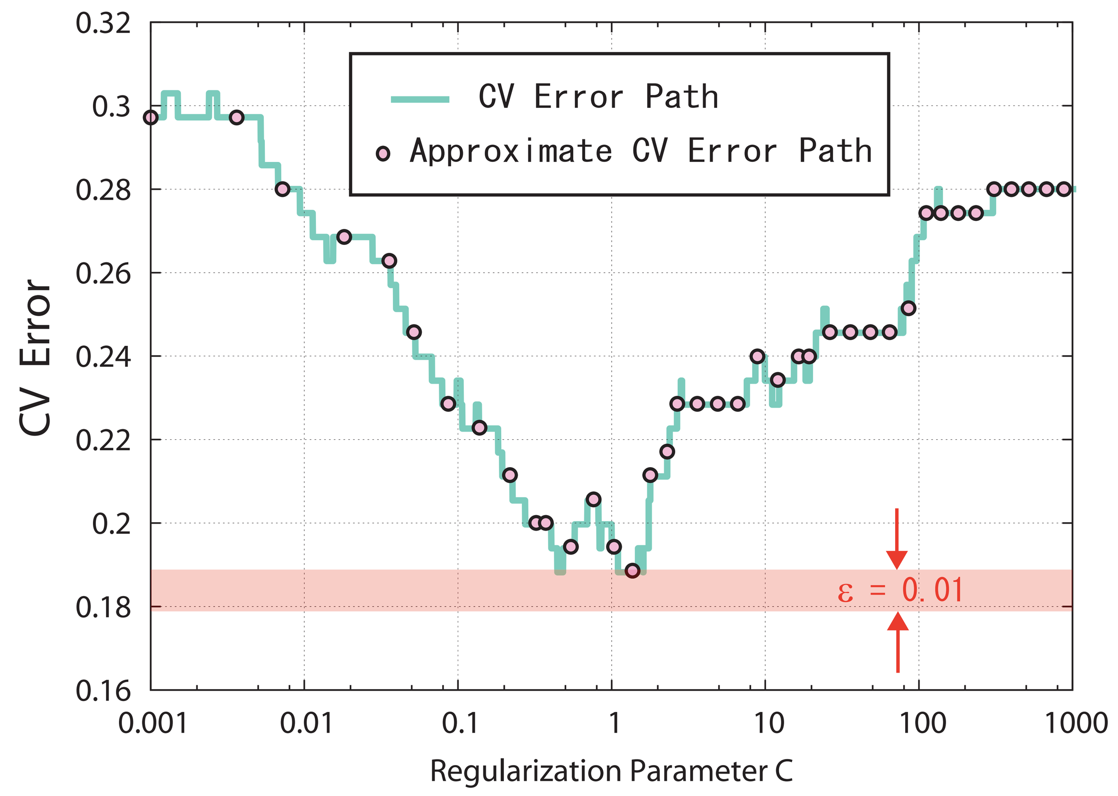
69 Regularization Path of Cross-Validation Error Lower Bounds

Atsushi Shibagaki, Yoshiki Suzuki, Masayuki Karasuyama, Ichiro Takeuchi (Nagoya Institute of Technology, Japan)

This Work in A Nutshell

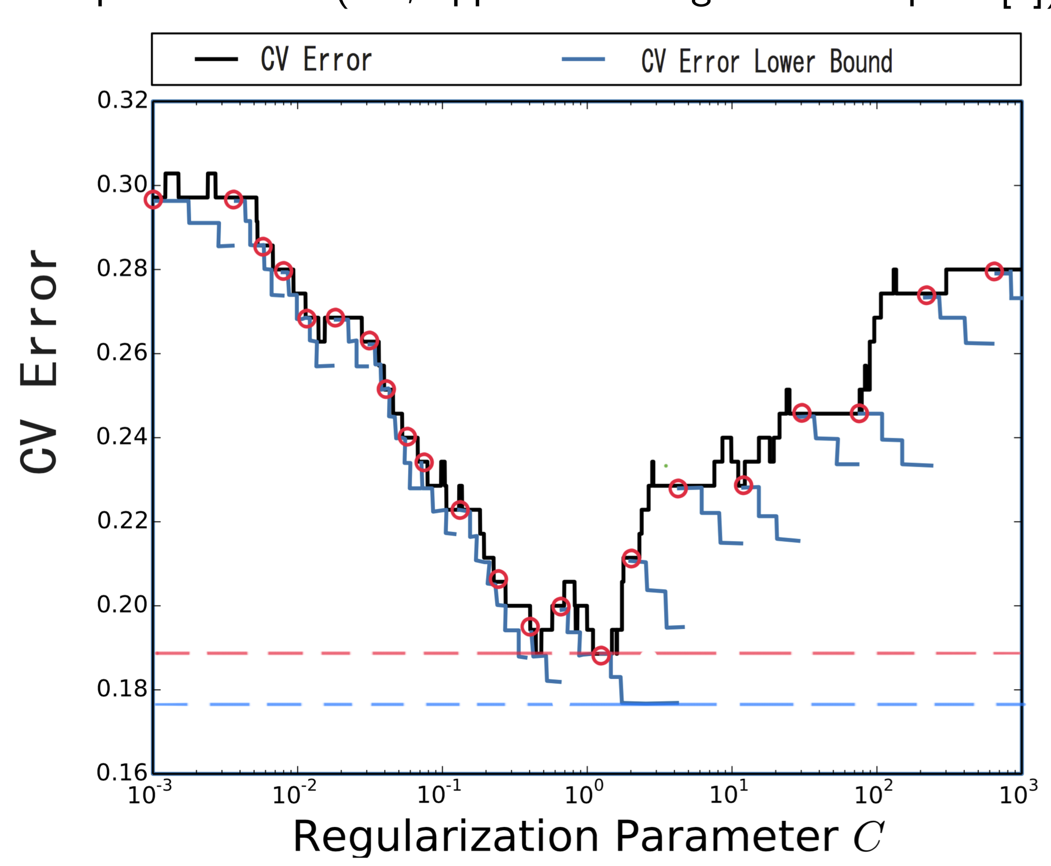
1. Approximation Guarantee in CV

- Our method can find a regularization parameter which has a theoretical approximation guarantee in terms of the cross validation (CV) error.
- Example: L_2-regularized Logistic Regression



2. Path of CV Error Lower Bounds

- We derived a novel CV error lower bound that can be represented as a function of the regularization parameter C (c.f., approximate regularization path [1])

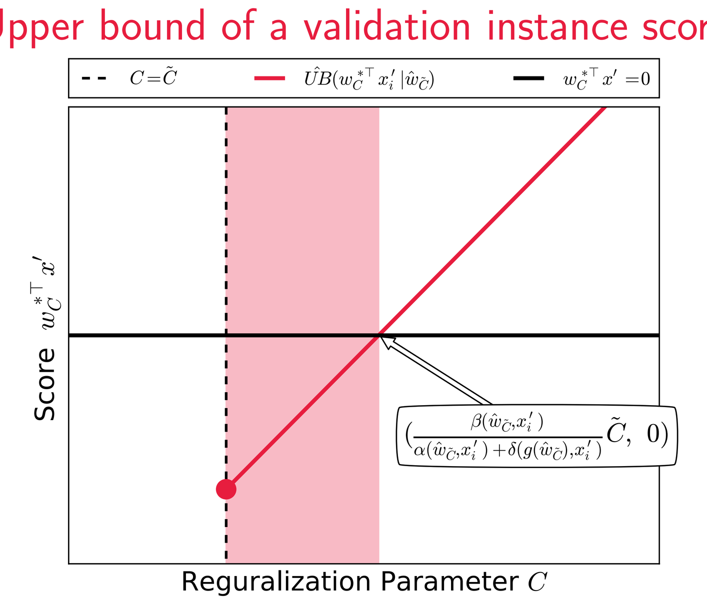


3. Lower and Upper Bounds of A Validation Instance Score

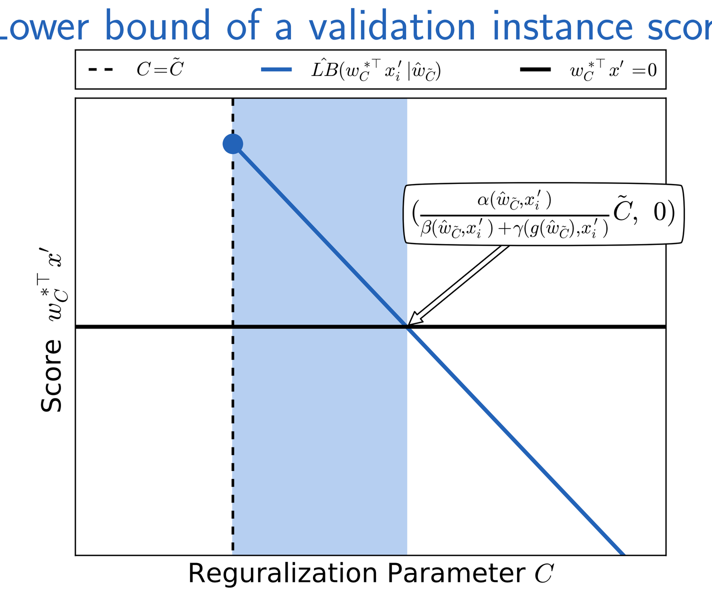
Binary classification for a validation instance x'_i is formulated as:

$$\hat{y}'_i := \begin{cases} +1 & \text{if } f(x'_i) > 0, \\ -1 & \text{if } f(x'_i) < 0, \end{cases} \quad \text{where } f(x'_i) := w_C^{*\top} x'_i.$$

Upper bound of a validation instance score



Lower bound of a validation instance score



Abstract

We propose a novel framework for computing lower bound of the cross-validation (CV) errors as a function of the regularization parameter. The proposed framework can be used for providing a theoretical approximation guarantee on a set of solutions in the sense that how far the CV error of the current best solution could be away from best possible CV error in the entire range of the regularization parameters.

Problem Setup (for hold-out validation set case)

- Convex regularized empirical risk minimization (e.g., SVM)

$$w_C^* := \arg\min_{w \in \mathbb{R}^d} \frac{1}{2}\|w\|_2^2 + C \sum_{i \in [n]} \ell(y_i, w^\top x_i) \qquad (1)$$

- Regularization parameter : C
- Training instances and labels : $\{(x_i, y_i) \in \mathbb{R}^d \times \{-1, 1\}\}_{i \in [n]}$
- Validation instances and labels : $\{(x'_i, y'_i) \in \mathbb{R}^d \times \{-1, 1\}\}_{i \in [n']}$
- Validation error: $E_v(w) := \frac{1}{n'} \sum_{i \in [n']} I(y'_i w^\top x'_i < 0)$
- ε-approximate regularization parameters in $C \in [C_l, C_u]$:

$$\mathcal{C}(\varepsilon) := \left\{ C \in [C_l, C_u] \;\middle|\; E_v(w_C^*) - (\text{the best possible } E_v \text{ in } [C_l, C_u]) \le \varepsilon \right\} \qquad (2)$$

- Goal: finding one of ε-approximate regularization parameter

Algorithm

The algorithm automatically select a sequence of regularization paramters $\{\tilde{C}_t\}_{t=1,2,\dots}$ such that the best among them is shown to be an ε-approximate one.

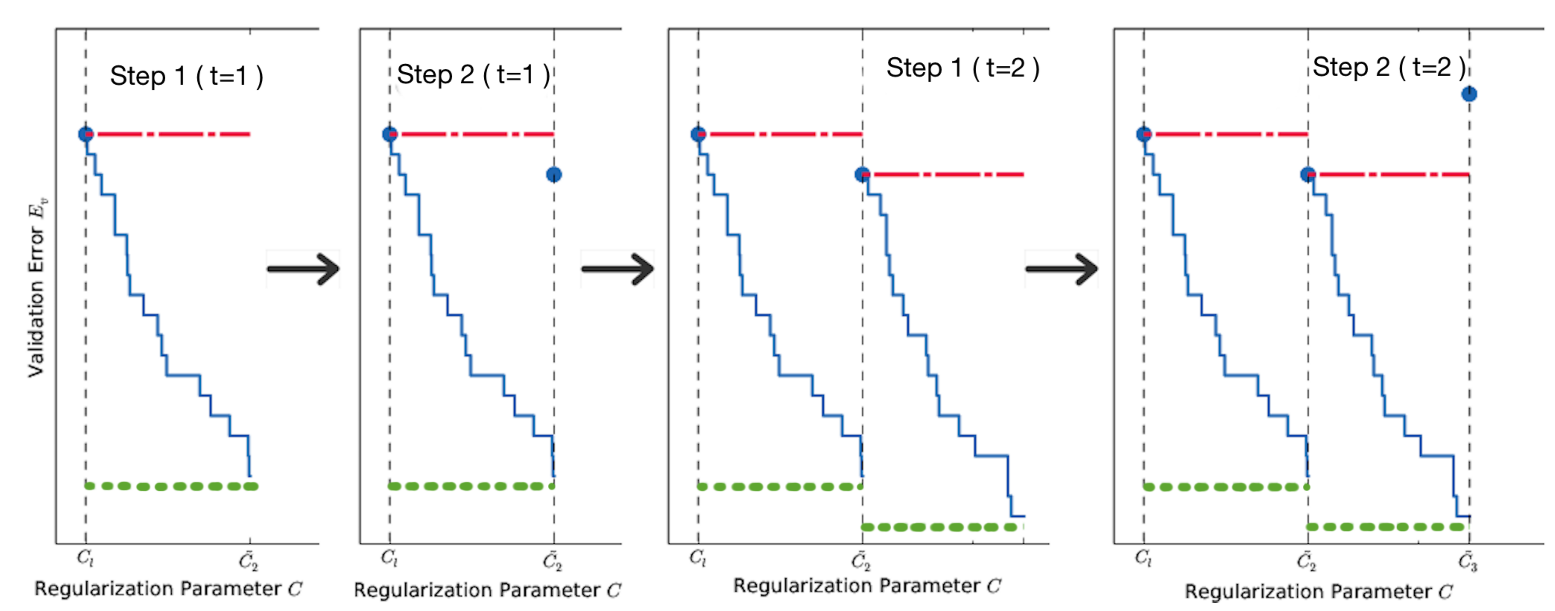
Input: $\varepsilon \in [0, 1]$, $[C_l, C_u]$

Step1: Based on a solution in the previous step, compute VE lower bound $LB(E_v(w_C^*))$ and update the current best VE.

Step2: Find $\tilde{C}_{t+1}$ such that the current best C is ε-approximation in the interval $[C_l, \tilde{C}_{t+1}]$, and compute $w^*_{\tilde{C}_{t+1}}$ at $C = \tilde{C}_{t+1}$.
(VE LB (blue) should not exceed the current best VE - ε (green))

Repeat: Steps 1 and 2 until $\tilde{C}_{t+1} > C_u$

Output: ε-approximate regularization parameter



Can guarantee (the current best VE) - (the best possible VE) $\le \varepsilon$

Regularization path of VELB

1. Compute regularization path of lower and upper bounds of validation instance scores $f(x'_i) = w_C^{*\top} x'_i, i \in [n']$
   - Construct regions of optimal solutions (Sphere) by using $\hat{w}_{\tilde{C}}$

$$\left\| w_C^* - \underbrace{\frac{1}{2}\left(\hat{w}_{\tilde{C}} - \frac{C}{\tilde{C}}(g(\hat{w}_{\tilde{C}}) - \hat{w}_{\tilde{C}})\right)}_{\text{center}} \right\|^2 \le \left(\underbrace{\frac{1}{2}\left\| \hat{w}_{\tilde{C}} + \frac{C}{\tilde{C}}(g(\hat{w}_{\tilde{C}}) - \hat{w}_{\tilde{C}}) \right\|}_{\text{radius}} \right)^2 \qquad (3)$$

   - Solve the following optimization problems :
     Lower bound : $w_C^{*\top} x'_i \ge \hat{LB}(w_C^{*\top} x'_i \mid \hat{w}_{\tilde{C}}) := \min_w w^\top x'_i$ s.t. (3)
     Upper bound : $w_C^{*\top} x'_i \le \hat{UB}(w_C^{*\top} x'_i \mid \hat{w}_{\tilde{C}}) := \max_w w^\top x'_i$ s.t. (3)
     change linearly with a regularized parameter

$$w_C^{*\top} x'_i \ge LB(w_C^{*\top} x'_i | \hat{w}_{\tilde{C}}) = \alpha(\hat{w}_{\tilde{C}}, x'_i) - \frac{C}{\tilde{C}}(\beta(\hat{w}_{\tilde{C}}, x'_i) + \gamma(g(\hat{w}_{\tilde{C}}), x'_i)),$$

$$w_C^{*\top} x'_i \le UB(w_C^{*\top} x'_i | \hat{w}_{\tilde{C}}) = -\beta(\hat{w}_{\tilde{C}}, x'_i) + \frac{C}{\tilde{C}}(\alpha(\hat{w}_{\tilde{C}}, x'_i) + \delta(g(\hat{w}_{\tilde{C}}), x'_i)),$$

$$\alpha(\hat{w}_{\tilde{C}}, x'_i) := \frac{1}{2}(\|\hat{w}_{\tilde{C}}\|\|x'_i\| + \hat{w}_{\tilde{C}}^\top x'_i) \ge 0, \;\; \gamma(g(\hat{w}_{\tilde{C}}), x'_i) := \frac{1}{2}(\|g(\hat{w}_{\tilde{C}})\|\|x'_i\| + g(\hat{w}_{\tilde{C}})^\top x'_i) \ge 0,$$

$$\beta(\hat{w}_{\tilde{C}}, x'_i) := \frac{1}{2}(\|\hat{w}_{\tilde{C}}\|\|x'_i\| - \hat{w}_{\tilde{C}}^\top x'_i) \ge 0, \;\; \delta(g(\hat{w}_{\tilde{C}}), x'_i) := \frac{1}{2}(\|g(\hat{w}_{\tilde{C}})\|\|x'_i\| - g(w_{\tilde{C}})^\top x'_i) \ge 0.$$

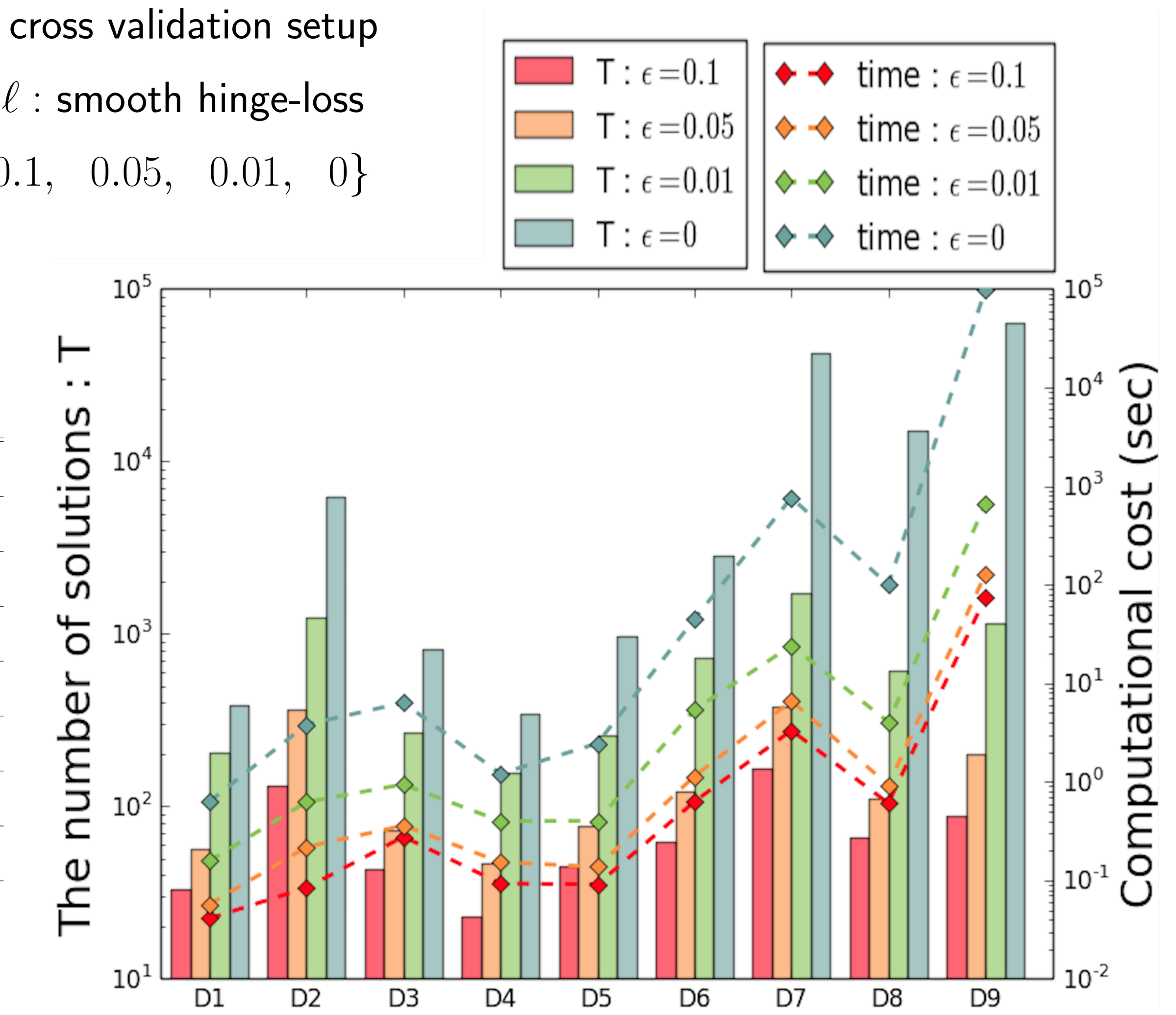
2. Identify the interval of C within which x'_i is guaranteed to be misclassified
   $LB(E_v(w_C^*)) = \#$ (validation instances that are guranteed to be misclassified) $/n'$

Experiments

- The entire interval of regularization parameter : $[10^{-3}, 10^3]$
- Under 10-fold cross validation setup
- Loss function ℓ : smooth hinge-loss
- Input : $\varepsilon = \{0.1, \;\; 0.05, \;\; 0.01, \;\; 0\}$

Benchmark datasets

	n	d
D1	270	13
D2	345	6
D3	351	34
D4	690	14
D5	768	8
D6	1000	24
D7	1284	21
D8	7089	4
D9	32561	123



References

[1] J. Giesen, J. Mueller, S. Laue, and S. Swiercy. Approximating Concavely Parameterized Optimization Problems. *NIPS*, 2012.

[2] K. Ogawa, Y. Suzuki, and I. Takeuchi. Safe screening of non-support vectors in pathwise SVM computation. *ICML*, 2013.