Second Order PAC-Bayesian Bounds for the Weighted Majority Vote

Andrés R. Masegosa, University of Almería

Stephan S. Lorenzen, Christian Igel, Yevgeny Seldin, University of Copenhagen

Introduction

Weighted majority vote is a fundamental technique in machine learning for combining predictions of multiple classifiers.

Our Contributions:

- Second order PAC-Bayesian bound for the weighted majority vote.
- Minimization of the bound does not deteriorate the test error.

Standard first order analysis

Observation: If ρ-weighted majority vote makes an error, then at least a ρ-weighted half of the classifiers make an error. Thus, we have that

$$\underbrace{L(\mathrm{MV}_\rho)}_{\substack{\text{Expected loss of}\\ \rho\text{-weighted majority vote}}} \leq \mathbb{P}\Big(\underbrace{\mathbb{E}_\rho[\mathbb{1}(h(X) \neq Y)]}_{\substack{\rho\text{-weighted mass}\\ \text{of errors}}} \geq 0.5 \Big)$$

First order bound: Applying Markov's inequality ($\mathbb{P}(X \geq \varepsilon) \leq \frac{1}{\varepsilon}\mathbb{E}[X]$)

$$\begin{aligned} &\leq 2\,\mathbb{E}_D[\mathbb{E}_\rho[\mathbb{1}(h(X) \neq Y)]] \\ &= 2\,\underbrace{\mathbb{E}_\rho[L(h)]}_{\substack{\text{Expected loss of } \rho\text{-weighted}\\ \text{randomized classifier}}} \end{aligned}$$

Issues with the first-order bound:

- Ignores correlation of errors among ensemble members.
- Minimization degrades the test error (Lorenzen et al., 2019).

Prior second order analysis

The C-bounds (Lacasse et al., 2007, Germain et al., 2015, Laviolette et al., 2017): Based on Chebyshev-Cantelli inequality

$$\mathbb{P}(X \geq \varepsilon) \leq \frac{\mathbb{E}[X^2] - \mathbb{E}[X]^2}{\mathbb{E}[X^2] - \mathbb{E}[X] + \varepsilon^2}$$

Issues with prior second order bounds:

- $\mathbb{E}[X^2]$ and $\mathbb{E}[X]$ in the denominator make estimation hard.
- Empirically weaker than the first order bound.
- Impossible to optimize except in very restrictive cases.

A novel second order oracle bound

$$\underbrace{L(\mathrm{MV}_\rho)}_{\substack{\text{Expected loss of}\\ \rho\text{-weighted majority vote}}} \leq \mathbb{P}\Big(\underbrace{\mathbb{E}_\rho[\mathbb{1}(h(X) \neq Y)]}_{\substack{\rho\text{-weighted mass}\\ \text{of errors}}} \geq 0.5 \Big)$$

Second-order Markov's inequality $\mathbb{P}(X \geq \varepsilon) \leq \frac{1}{\varepsilon^2}\mathbb{E}[X^2]$:

$$\begin{aligned} &\leq 4\,\mathbb{E}_D[\mathbb{E}_\rho[\mathbb{1}(h(X) \neq Y)]^2] \\ &= 4\,\mathbb{E}_{\rho^2}[\underbrace{\mathbb{E}_D[\mathbb{1}(h(X) \neq Y \wedge h'(X) \neq Y)]}_{\text{Expected \textbf{Tandem Loss}: } L(h,h')}] \\ &= 4\,\mathbb{E}_{\rho^2}[L(h,h')] \end{aligned}$$

Key Points:

- **Tandem loss** counts an error if both h and h' err on a sample.
- Tandem loss is small when the two hypotheses h and h' both have low expected errors and their **errors are anticorrelated**.
- **Second order oracle bound:**

$$L(\mathrm{MV}_\rho) \leq 4\,\mathbb{E}_{\rho^2}[L(h,h')]$$

Second order PAC-Bayesian bound

By applying PAC-Bayes-λ (Thiemann et al., 2017) to $\mathbb{E}_{\rho^2}[L(h,h')]$

For π independent of S, with probability at least $1-\delta$ for all ρ and $\lambda \in (0,2)$

$$L(\mathrm{MV}_\rho) \leq 4\,\mathbb{E}_{\rho^2}[L(h,h')]$$

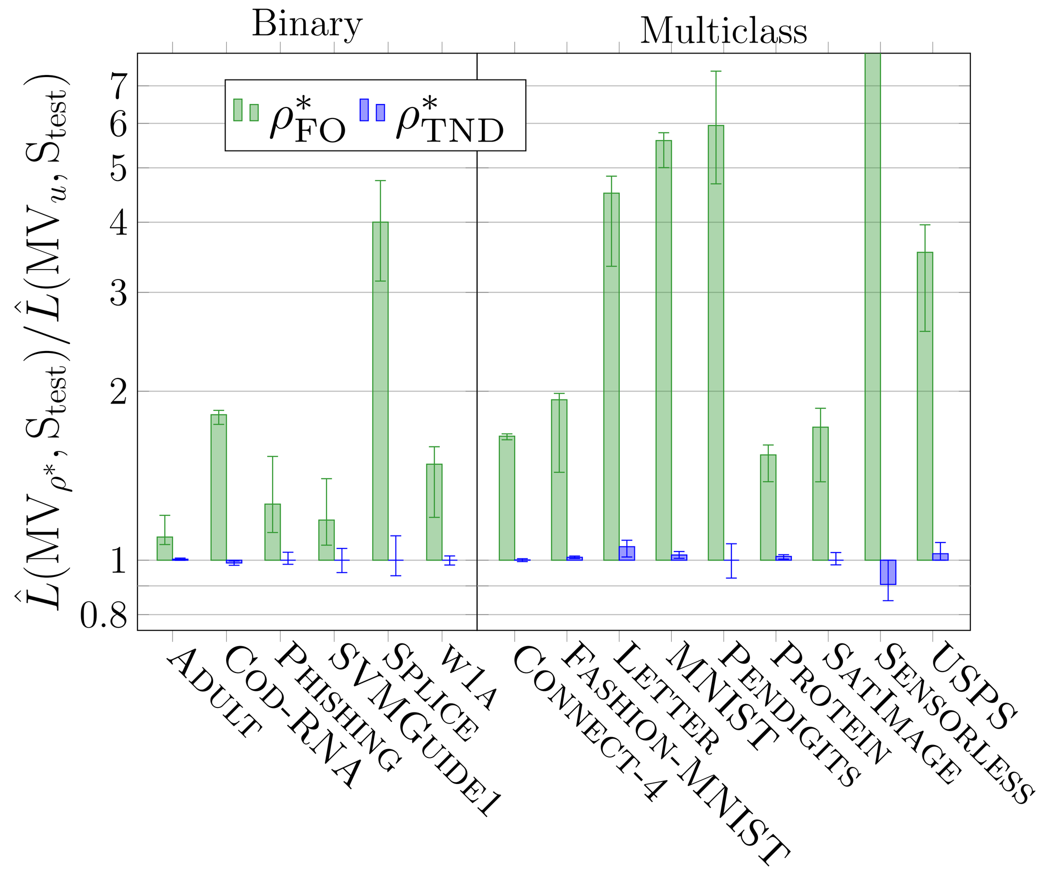
$$\leq 4 \underbrace{\left(\frac{\overbrace{\mathbb{E}_{\rho^2}[\hat{L}(h,h',S)]}^{\text{Empirical \textbf{Tandem loss}}}}{1-\lambda/2} + \frac{2\,\mathrm{KL}(\rho\|\pi) + \ln\frac{2\sqrt{n}}{\delta}}{n\lambda(1-\lambda/2)} \right)}_{\text{PAC-Bayes upper bound on } \mathbb{E}_{\rho^2}[L(h,h')]}$$

Key Points:

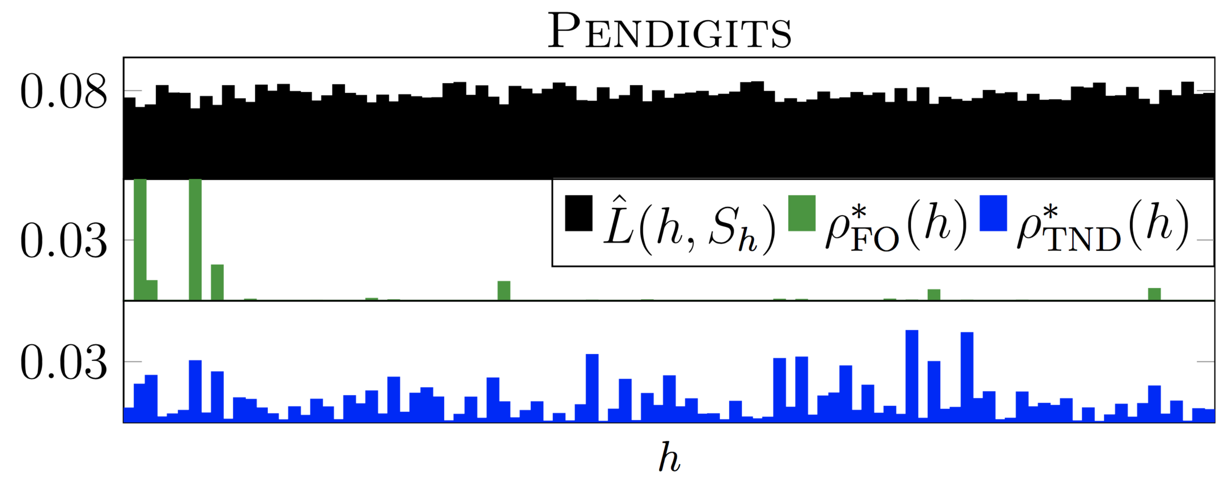
- Takes correlation of errors into account.
- We can easily find the ρ distribution minimizing the bound.
- The minimization of the bound does not degrade the test error

Empirical evaluation

- Test error of optimized majority vote over uniformly weighted baseline for first order [FO] and new second order [TND] bound (the lower the better)



- Optimized weights ρ^* generated by the first order [FO] and the new second order [TND] bound.



Summary

- Minimizing first order bounds deteriorates the test error. Existing second order bounds are looser and can not be optimized.
- Novel second order oracle and PAC-Bayesian bounds for the weighted majority vote based on second order Markov's inequality.
- Minimization of the bound does not deteriorate the test error.