Learning From Non-iid Data: Fast Rates for the One-vs-All Multiclass Plug-in Classifiers

Vu Dinh[1] Lam Si Tung Ho[2] Nguyen Viet Cuong[3]
Duy Duc Nguyen[4] Binh T. Nguyen[5]

[1]Purdue University [2]University of California, Los Angeles [3]National University of Singapore
[4]University of Wisconsin-Madison [5]University of Science, Vietnam

- Fast and super fast learning rates for plug-in classifier
  - **Multiclass** setting
  - **Non-iid** data
- Non-iid data
  - Exponentially strongly mixing data
  - Converging drifting data
- Generalization of previous result for binary-class and iid case
- Algorithm does not need to know the exponent in the margin assumption
- The rates have nice properties
  - Not depend on the number of classes
  - Retain optimal learning rate for the Hölder class in iid case

Assumptions

1. All label distribution functions $\eta_j(X)$ are *Hölder continuous* with exponent β.
2. Marginal distribution $\mathbf{P}_X$ satisfies *strong density assumption*.
   - Its density has *positive* upper and lower bounds on a compact regular set of $\mathbb{R}^d$.
3. **P satisfies multiclass margin assumption**.

Fast Rates for Exponentially Strongly Mixing Data

Theorem

We can construct a one-vs-all multiclass plug-in classifier $\widehat{f}_n$ *that satisfies: there exist* $C_1, C_2 > 0$ *such that for all large enough* n*,*

$$\mathbf{E}R(\widehat{f}_n) - R(f^*) \leq C_1 n^{-C_2\beta(1+\alpha)/(2\beta+d)}.$$

- α: constant in the margin assumption
- β: exponent in the Hölder continuous assumption
- d: dimension of the input space $\mathbb{R}^d$
- Expected risk of plug-in classifier converges to optimal risk with rate $n^{-C_2\beta(1+\alpha)/(2\beta+d)}$.
  - Fast rate when $C_2\beta(1+\alpha)/(2\beta+d) > 1/2$
  - Super fast rate when $C_2\beta(1+\alpha)/(2\beta+d) > 1$

Fast Rates for Drifting Data

Theorem

We can construct a one-vs-all multiclass plug-in classifier $\widehat{f}_n$ that satisfies: there exists $C > 0$ such that for all large enough n,

$$\mathbf{E}R(\widehat{f}_n) - R(f^*) \leq C\, n^{-\beta(1+\alpha)/(2\beta+d)}.$$

- Expected risk of plug-in classifier converges to optimal risk with rate $n^{-\beta(1+\alpha)/(2\beta+d)}$.
  - Fast rate when $\beta(1+\alpha)/(2\beta+d) > 1/2$
  - Super fast rate when $\beta(1+\alpha)/(2\beta+d) > 1$

Thank you.