Progressive mixture rules are deviation suboptimal

J.-Y. Audibert
Poster ID T14

- $(X_1, Y_1), \ldots, (X_n, Y_n)$ i.i.d. $X \in \mathcal{X}$ $Y \in \mathbb{R}$
- $g_1, \ldots, g_d$ (prediction) functions from $\mathcal{X}$ to $\mathbb{R}$
- Problem: predict as well as the best function g_i, $i = 1, \ldots, d$ for the quadratic risk: $R(g) = \mathbb{E}\big[Y - g(X)\big]^2$
- Known solution: the progressive mixture rule, a.k.a. the exponentially weighted average algorithm, which satisfies

$$\mathbb{E}\, R(\hat{g}) - \min_{i=1,\ldots,d} R(g_i) \leq \text{Cst}\, \frac{\log d}{n} \qquad (A)$$

- Proposed solution: minimize the empirical risk among functions in the star shaped $\cup_{i=1}^{d} [\hat{g}_{\text{ERM}}; g_i]$, where $\hat{g}_{\text{ERM}}$ is the ERM among $\{g_1, \ldots, g_d\}$.

Known solution	Proposed solution
$1/n$ expectation rate (A)	$1/n$ expectation rate
$1/\sqrt{n}$ **deviation rate !!!**	$1/n$ **deviation rate**

New results in blue