ERRATA for paper “Minimax Bounds for Active Learning” - In COLT 2007

Rui M. Castro and Robert D. Nowak

It is virtually impossible to get anything exactly right — Carl de Boor, University of Wisconsin

It was noted by the authors that Theorem 3 in [2] is not entirely correct, and it is not valid in all the generality it was stated. This glitch does not affect the conclusions in [2], but it is important to issue a correction. The following theorem is adapted from [1] (page 85, theorem 2.5).

**Theorem** (Tsybakov, 2004). Let \( \mathcal{F} \subseteq \Xi \) be a class of models. Associated with each model \( f \in \mathcal{F} \) we have a probability measure \( P_f \) defined on a common probability space. Let \( M \geq 2 \) be an integer and let \( d_\Delta(\cdot, \cdot) : \Xi \times \Xi \to \mathbb{R} \) be a semi-distance. Suppose we have \( \{f_0, \ldots, f_M\} \in \mathcal{F} \) such that

i) \( d_\Delta(f_j, f_k) \geq 2a > 0, \quad \forall 0 \leq j, k \leq M \),

ii) \( P_{f_0} \ll P_{f_j}, \quad \forall j = 1, \ldots, M \),

iii) \( \frac{1}{M} \sum_{j=1}^{M} KL(P_{f_j} || P_{f_0}) \leq \gamma \log M \),

where \( 0 < \gamma < 1/8 \). The following bound holds.

\[
\inf_{\bar{f}} \sup_{f \in \mathcal{F}} P_f \left( d_\Delta(\bar{f}, f) \geq a \right) \geq \frac{\sqrt{M}}{1 + \sqrt{M}} \left( 1 - 2\gamma - 2\sqrt{\frac{\gamma}{\log M}} \right) > 0,
\]

where the infimum is taken with respect to the collection of all possible estimators of \( f \) (based on a sample from \( P_{f_j} \)), and \( KL \) denotes the Kullback-Leibler divergence.

Furthermore if we have \( d(\cdot, \cdot) : \Xi \times \Xi \to \mathbb{R} \) such that for all \( f \in \Xi \) and \( i \in \{0, \ldots, M\} \)

\[
d(f, f_i) \geq h(d_\Delta(f, f_i)),
\]

where \( h : \mathbb{R} \to \mathbb{R} \) is a monotonically non-decreasing function, then

\[
\inf_{\bar{f}} \sup_{f \in \mathcal{F}} P_f \left( d(\bar{f}, f) \geq h(a) \right) \geq \frac{\sqrt{M}}{1 + \sqrt{M}} \left( 1 - 2\gamma - 2\sqrt{\frac{\gamma}{\log M}} \right) > 0.
\]

The revised proof of Theorem 1 in [2] proceeds by showing that for the collection of distributions \( P_{X^Y} \) constructed we have

\[
d_\Delta(G^*_i, G^*_j) \geq L\|h\|_1 m^{-\alpha},
\]

and

\[
d(G, G^*_i) \geq R_i(G) - R_i(G^*_i) \geq \frac{4c}{k\alpha^2} d_\Delta(G, G^*_i),
\]

for any set \( G \subseteq [0, 1]^d \). Applying the above stated theorem yields the final result of Theorem 1 in [2]. The detailed proof is available in technical report [3] (see Appendix B - “Proof of Theorem 3”).

**References**

