Lecture 16 — November 12, 2024

Prof. Stephen Bates

Scribe: Haichen Hu, Cai Zhou

1 Outline

Agenda:

1. Predictive inference with high-dimensional inputs
2. General Statistical Error Rates
3. Risk control algorithm and guarantees

Last time:

1. Challenge: fairly assess confidence
2. Linear model example
3. Conformal prediction

2 General Statistical Error Model

Setting:

- model: $\hat{f} : \mathcal{X} \to \mathcal{Z}$, e.g. tumor $\hat{f} : \mathcal{X} \to [0,1]^{d \times d}$
- fresh data: $(X_i, Y_i) \in \mathcal{X} \times \mathcal{Y}, i = 1, \ldots, n$, test point $n+1$ i.i.d
- pointwise loss: $L : \mathcal{Y} \times \mathcal{Y}' \to [0,1]$ e.g., $L(y, S) = 1 - \frac{|y \cap S|}{|y|}$
- confidence prediction $T : \mathcal{X} \to \mathcal{Y}'$, e.g. $T(X) = \{(i,j) : \hat{f}(x)_{(i,j)} > -\lambda\}$
- risk $R(T) = \mathbb{E}[L(Y, T(X))]$, e.g., false positive rate or false negative rate
- monotonic loss: $S \subset S', \Rightarrow L(y, S) \geq L(y, S')$
- 1-d nested family $T_\lambda, \lambda \in \mathbb{R} = \Lambda$, if $\lambda_1 < \lambda_2 \Rightarrow T_{\lambda_1}(X) \subset T_{\lambda_2}(X)$
- Notation: confidence bound (UCB) $P(R(T_\lambda) \leq R^+(T_\lambda)) \geq 1 - \delta$, $R^+(T_\lambda)$ is from data and we compute (with Hoeffding or CLT, etc.), while $R(T_\lambda)$ is what we wish to compute the distribution but cannot

Remark: We could view Conformal Prediction as a special type of this model, where the loss function $L(y, C(x)) = I\{y \notin C(x)\}$.

Definition 1. *Let T be a random function $T : \mathcal{X} \to \mathcal{Y}'$, Γ is a (γ, δ)-risk controlling prediction set (RCPS) if*

$$P(R(T) > \gamma) < \delta. \tag{1}$$

3 Property of RCPS

Theorem 2. *Given family of set-valued predictors $T_\lambda(x)$ and aUCB $\hat{R}^+(\lambda)$, choose $\hat{\lambda} = \inf\{\lambda \in \Lambda : \hat{R}^+(T_{\lambda'}) \leq \gamma, \forall \lambda' \geq \lambda\}$. Then with probability at least $1 - \delta$, $R(T_{\hat{\lambda}}) \leq \gamma$.*

Proof. Since the risk is decreasing as λ increases. We define the following λ^*.

$$\lambda^* = \inf\{\lambda : \hat{R}^+(T_{\lambda'}) \leq \gamma, \forall \lambda' \geq \lambda\}.$$

Therefore, $R(T_{\hat{\lambda}}) \geq \gamma \Leftrightarrow \hat{\lambda} \leq \lambda$. By the definition of UCB, we have:

$$P(\hat{\lambda} \leq \lambda^*) \leq P(R^+(T_{\lambda^*}) \geq \gamma) \leq \delta.$$

□

4 Example

- UCB calibration on tumors data
- Multi-label classification
- Hierarchical classification
- Protein structure prediction

Please refer to the slides for more details.

5 Conclusion

Pros of Predictive Inference:

- Finite-Sample risk guarantees
- any predictive model
- any distribution
- incorporate many error notions
- deal with high dimensional outputs