Olivier Bousquet: New Approaches to Statistical Learning Theory (2002)

Slides by Risi Kondor

Error bounds

Training set: $\{(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)\}$ $\quad X \in \mathcal{X},\ Y \in \mathcal{Y} \quad (X, Y) \sim P$.

Algorithm learns $g : \mathcal{X} \mapsto \mathcal{Y} \quad g \in \mathcal{G}$.

Loss function $c : \mathcal{Y} \times \mathcal{Y} \mapsto [0, b]$ for some b.

Error bound:

$$\mathbb{P}\Bigg[\underbrace{\mathbb{E}\left[c(g(X), Y)\right]}_{L(g)} \leq \underbrace{\frac{1}{n}\sum_{i=1}^{n} c(g(X_i), Y_i)}_{L_n(g)} + \epsilon \Bigg] \geq 1 - \delta.$$

De£ne $f_g : \mathcal{X} \times \mathcal{Y} \mapsto [0, b]$ such that $f_g(x, y) = c(g(x), y)$.

Loss class: $\mathcal{F} = \{f_g \mid g \in \mathcal{G}\}$.

Uniform bounds probabilistically bound

$$\sup_{g \in \mathcal{G}} \left[L(g) - L_n(g)\right] = \sup_{f \in \mathcal{F}} \left[\mathbb{E}\left[f(X, Y)\right] - \frac{1}{n}\sum_{i=1}^{n} f(X_i, Y_i)\right] = \sup_{f \in \mathcal{F}} \left[Pf - P_n f\right]$$

called *the supremum of the empirical process indexed by the class* $\mathcal{F}$.

Rademacher averages

Rademacher random variable: $\mathbb{P}[\sigma = 1] = \mathbb{P}[\sigma = -1] = 1/2.$

Rademacher average of class $\mathcal{F}$ is

$$\mathbb{E}\left[\sup_{f \in \mathcal{F}} R_n f\right] \quad \text{where} \quad R_n f = \frac{1}{n} \sum_{i=1}^{n} \sigma_i f(X_i, Y_i)$$

The plan

Concentration:

$$\sup_{f\in\mathcal{F}} [Pf - P_n f] - \mathbb{E}\left[\sup_{f\in\mathcal{F}} [Pf - P_n f]\right] \leq b\sqrt{\frac{2\epsilon}{n}}$$

(with probability $1 - e^{-\epsilon}$).

Symmetrization:

$$\mathbb{E}\left[\sup_{f\in\mathcal{F}} [Pf - P_n f]\right] \leq 2\mathbb{E}\left[\sup_{f\in\mathcal{F}} R_n f\right].$$

Relation to empirical Rademacher:

$$\mathbb{E}\left[\sup_{f\in\mathcal{F}} R_n f\right] \leq \inf_{\alpha>0}\left[\frac{1}{1-\alpha}\,\mathbb{E}_\sigma\left[\sup_{f\in\mathcal{F}} R_n f\right] + \left(\frac{1}{3} + \frac{1}{2\alpha}\,\frac{b\epsilon}{n}\right)\right]$$

(with probability $1 - e^{-\epsilon}$).

Hoeffding's Inequality

Let $X_1, X_2, \ldots, X_n$ be independent random variables such that $a \leq X_i \leq b$.
Then with probability $1 - e^{-2t^2/\left(n(b-a)^2\right)}$,

$$\pm \left[\frac{1}{n} \sum_{i=1}^{n} X_i - \mathbb{E}X \right] \leq t.$$

Equivalently, with probability $1 - e^{-\epsilon}$,

$$\pm \left[\frac{1}{n} \sum_{i=1}^{n} X_i - \mathbb{E}X \right] \leq |\, b - a \,| \sqrt{\frac{n\epsilon}{2}}.$$

McDiarmid

For random variables $X_1, X_2, \ldots, X_n$ and some function h let

$$Z = h(Z_1, Z_2, \ldots, Z_n)$$

and for some function $h^\dagger$ let

$$Z_k = h^\dagger(Z_1, \ldots, Z_{k-1}, Z_{k+1}, \ldots, Z_n).$$

Assume that $|\, Z - Z_k \,| \leq b$. Then for all $\epsilon > 0$,

$$\mathbb{P}\left[\, \pm (Z - \mathbb{E}Z) \geq b\sqrt{n\epsilon/2}\, \right] \leq e^{-\epsilon}.$$

Concentration of $Pf - P_n f$

From the boundedness of the loss, it follows that $\sup_{f \in \mathcal{F}, x, y} | Pf - f(x, y) | \leq b$. De£ning

$$Z = \sup_{f \in \mathcal{F}} [Pf - P_n f]$$

and

$$Z^{\dagger} = \sup_{f \in \mathcal{F}} \left[Pf - \frac{1}{n} \sum_{i \neq k} f(X_i, Y_i) \right]$$

we have $\left| Z - Z^{\dagger} \right| < b/n$. Plugging into McDiarmid gives that with probability $1 - e^{-\epsilon}$

$$\pm \left[\sup_{f \in \mathcal{F}} [Pf - P_n f] - \mathbb{E} \left[\sup_{f \in \mathcal{F}} [Pf - P_n f] \right] \right] \leq b \sqrt{\frac{2\epsilon}{n}}.$$

Symmetrization

$$\begin{aligned}
\mathbb{E}\left[\sup_{f\in\mathcal{F}}[Pf - P_n f]\right] &= \mathbb{E}\left[\sup_{f\in\mathcal{F}}\mathbb{E}\left[P'_n f\right] - P_n f\right] \\
&\leq \mathbb{E}\left[\sup_{f\in\mathcal{F}} P'_n f - P_n f\right] \quad \text{(Jensen)} \\
&= \mathbb{E}\left[\sup_{f\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^{n}\big(f(X'_i, Y'_i) - f(X_i, Y_i)\big)\right] \\
&= \mathbb{E}\left[\sup_{f\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^{n}\sigma_i\big(f(X'_i, Y'_i) - f(X_i, Y_i)\big)\right] \\
&= \mathbb{E}\left[\sup_{f\in\mathcal{F}}\left[\left[\frac{1}{n}\sum_{i=1}^{n}\sigma_i f(X'_i, Y'_i)\right] + \left[\frac{1}{n}\sum_{i=1}^{n}\sigma_i f(X_i, Y_i)\right]\right]\right] \\
&\leq 2\mathbb{E}\left[\sup_{f\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^{n}\sigma_i f(X_i, Y_i)\right] \\
&= 2\mathbb{E}\left[\sup_{f\in\mathcal{F}} R_n f\right]
\end{aligned}$$

Bousquet's one-sided concentration result

Assuming $\sup_{f\in\mathcal{F},x,y}[Pf - f(x,y)] \leq b$, with probability $1 - e^{-\epsilon}$

$$\sup_{f\in\mathcal{F}}[Pf - P_n f] \leq \inf_{\alpha>0}\left[(1+\alpha)\,\mathbb{E}\left[\sup_{f\in\mathcal{F}}[Pf - P_n f]\right] + \sqrt{\frac{2v}{n}} + \left(\frac{1}{3} + \frac{1}{\alpha}\right)\frac{b\epsilon}{n}\right].$$

Bousquet's Localized Concnetration Inequality

Let ϕ be a non-negative, non-decreasing function such that $\phi(r)/\sqrt{r}$ is non-increasing for $r > 0$ and sucht that

$$\mathbb{E}\left[\sup_{f\in\mathcal{F}\,:\,Pf\leq r}[Pf - P_n f]\right] \leq \phi(r)$$

and let r^* be the largest solution of $\phi(r) = r$. Then with probability $1 - e^{-\epsilon}$

$$\sup_{f\in\mathcal{F}} \frac{Pf - P_n f}{\sqrt{Pf}} \leq \inf_{\alpha>0}\left((1+\alpha)\sqrt{r^*}\left(1 + \frac{e}{2}\log\frac{eb}{r^*}\right) + \sqrt{\frac{2b\epsilon}{n}} + \frac{(3+\alpha)\,\epsilon\sqrt{b}}{3\alpha n}\right).$$

Let r and r^* be as above, except that now

$$\mathbb{E}\left[\sup_{f\in\mathcal{F}\,:\,Pf\leq r}|\,R_n f\,|\right] \leq \phi(r).$$

The there exists a universal constant K such that with probability $1 - e^{-\epsilon}$

$$Pf \leq K \inf_{\alpha>0}\left[(1+\alpha)\,P_n f + \left(1 + \frac{1}{4\alpha}\right)\left(31 r^* \log^2\frac{b}{r^*} + 50\frac{b\epsilon}{n}\right)\right].$$