Lecture-2 Supplementary

12-Jan-10

<u>Claim</u> $P\left[\sup_{f\in\mathcal{F}} \left\{R[f] - R^m_{emp}[f]\right\} > \epsilon\right] \longrightarrow 0$ as $m \to \infty$ for all $\epsilon > 0$ (I)

(as a sufficient cond. for)

$$\Downarrow$$

$$\inf_{f\in\mathcal{F}} R^m_{emp}[f] \xrightarrow{P} \inf_{f\in\mathcal{F}} R[f] \qquad \text{(II)}$$

<u>Proof</u> (II) $\Longleftrightarrow$ $P\left[\left|\inf_{f\in\mathcal{F}} R^m_{emp}[f] - \inf_{f\in\mathcal{F}} R[f]\right| > \epsilon\right] \longrightarrow 0$ (as $m\to\infty$ $\forall \epsilon > 0$)

$$P\left[\left|\inf_{f\in\mathcal{F}} R^m_{emp}[f] - \inf_{f\in\mathcal{F}} R[f]\right| > \epsilon\right] = \underbrace{P\left[\inf_{f\in\mathcal{F}} R^m_{emp}[f] - \inf_{f\in\mathcal{F}} R[f] > \epsilon\right] + P\left[\inf_{f\in\mathcal{F}} R[f] - \inf_{f\in\mathcal{F}} R^m_{emp}[f] > \epsilon\right]}_{\text{split into mutually exclusive events}}$$

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ (III) $\qquad\qquad\qquad$ (IV)

$\longrightarrow$ these are mutually exclusive as rightly pointed out by students in the class.

Lets call $\operatorname{argmin}_{f\in\mathcal{F}} R^m_{emp}[f]$ as f^m & $\operatorname{argmin}_{f\in\mathcal{F}} R[f]$ as f^{opt}.

$$\text{(III)} = P\left[R^m_{emp}[f^m] - R[f^{opt}] > \epsilon\right] = P\left[R^m_{emp}[f^m] - R^m_{emp}[f^{opt}] + R^m_{emp}[f^{opt}] - R[f^{opt}] > \epsilon\right]$$

$$\leq P\left[R^m_{emp}[f^{opt}] - R[f^{opt}] > \epsilon\right] \qquad \left(\because R^m_{emp}[f^m] \leq R^m_{emp}[f^{opt}] \text{ by defn. of } f^m\right)$$

$$\leq e^{-2m\epsilon^2/(b-a)^2} \qquad (\because \text{ chernoff bound})$$

$$\downarrow \; 0 \text{ as } m \to \infty$$

By sandwiching (III) goes to zero as $m \to \infty$.

$$\text{(IV)} = P\left[R[f^{opt}] - R^m_{emp}[f^m] > \epsilon\right] = P\left[R[f^{opt}] - R[f^m] + R[f^m] - R^m_{emp}[f^m] > \epsilon\right]$$

$$\leq P\left[R[f^m] - R^m_{emp}[f^m] > \epsilon\right] \qquad \left(\because R[f^{opt}] \leq R(f^m) \text{ by defn. of } f^{opt}\right)$$

note we <u>cannot</u> use chernoff bounds here!! $\longleftarrow$ $\quad \leq P\left[\sup_{f\in\mathcal{F}} R[f] - R^m_{emp}[f] > \epsilon\right]$ $\qquad (\because$ by defn. of sup.$)$

$$\downarrow \; 0 \text{ as } m \to \infty \text{ by (I)}$$

Proof for (modified) Chernoff bound we use after Symmetrization:

Claim: Suppose $a \le X_i \le b$, X_i are i.i.d (all X_i's mean is $E[X]$) for $i = 1$ to $2m$.
Then we have:

$$P\left[\frac{1}{m}\sum_{i=1}^{m} X_i - \frac{1}{m}\sum_{i=m+1}^{2m} X_i > \epsilon\right] \le e^{-m\epsilon^2/(b-a)^2}$$

Proof: $LHS = P\left[\sum_{i=1}^{m}(X_i - E[X]) - \sum_{i=m+1}^{2m}(X_i - E[X]) > m\epsilon\right]$

define $Z_i := X_i - E[X]$ for $i = 1$ to m

$Z_i := E[X] - X_i$ for $i = m+1$ to $2m$

Then Z_i are independent r.v.s, with mean zero and all Z_i (1 to $2m$) move in an interval of length $b-a$!

By apply Hoeffding/chernoff bounding we get

$$P\left[\sum_{i=1}^{2m} Z_i > m\epsilon\right] \le e^{-m\epsilon^2/(b-a)^2}.$$ Hence Proved.

Note that this bound is tighter than bounds used in Smola's book and Lugosi's tutorial. (Why?)