Lecture - 1
Supplementary Notes

Claim: Suppose loss $l(f(x),y)$ is bounded i.e. $a \le l(f(x),y) \le b$ for a function $f \in \mathcal{F}$, then the following is true:

$$P\left[\left|R^m_{emp}[f] - R[f]\right| > \epsilon\right] \le 2e^{-2m\epsilon^2/(b-a)^2}$$

(Note: here we are given a fixed $f \in \mathcal{F}$)

~~Proof~~: In words, for a given $f \in \mathcal{F}$ the probability that the empirical risk deviates from true Risk (by quantity $>\epsilon$) falls exponentially w.r.t. m.

Proof: $R^m_{emp}[f] = \frac{1}{m}\sum_{i=1}^{m} l(f(X_i),Y_i)$; $R[f] = E[l(f(X),Y)]$.

$$P\left[\left|R^m_{emp}[f] - R[f]\right| > \epsilon\right] = P\left(\left[R^m_{emp}[f] - R[f] > \epsilon\right] \cup \left[R[f] - R^m_{emp}[f] > \epsilon\right]\right)$$

$$\le P\left[R^m_{emp}[f] - R[f] > \epsilon\right] + P\left[R[f] - R^m_{emp}[f] > \epsilon\right]$$

(This is called union bound $P\left(\bigcup_{i=1}^{n} A_i\right) \le \sum_{i=1}^{n} P(A_i)$. This is also known as Boole's inequality $\xrightarrow{\text{infact}}$ $P\left(\bigcup_{i=1}^{\infty} A_i\right) \le \sum_{i=1}^{\infty} P(A_i) \longrightarrow$ this is more generic result)

Now, $P\left[R^m_{emp}[f] - R[f] > \epsilon\right] = P\left[\sum_{i=1}^{m} \underbrace{\{l(f(X_i),Y_i) - E[l(f(X),Y)]\}}_{\text{Call it } Z_i} > m\epsilon\right]$ — I

We know $E[l(f(X_i),Y_i)] = E[l(f(X),Y)]$ $\forall i$ (since (X_i,Y_i) are iid)

$\therefore Z_i$ rvs are mean zero. Moreover, $a \le l(f(X_i),Y_i) \le b$

$\Rightarrow a - E[l(f(X),Y)] \le Z_i \le b - E[l(f(X),Y)]$

$\Rightarrow$ length of interval in which Z_i moves is $(b - E[l(f(X),Y)]) - (a - E[l(f(X),Y)]) = b - a$

In other words I is prob. question ~~of~~ of sum of mean zero, iid rvs ~~...~~ all of which move in interval of length $b-a$!

bounds on probability regarding such sum of iid bounded r.v's is efficiently given by Hoeffding bounds which in turn use Chernoff bounding techniques.

Now, I is $P\left[\sum_{i=1}^{m} Z_i > m\varepsilon\right] = P\left[e^{s\sum Z_i} > e^{sm\varepsilon}\right] \quad \forall s > 0$

$$\leq \frac{E\left[e^{s\sum Z_i}\right]}{e^{sm\varepsilon}} \quad \text{(by Markov inequality)}$$

Chernoff bounding scheme

$$= \frac{E\left[\prod_{i=1}^{m} e^{sZ_i}\right]}{e^{sm\varepsilon}}$$

$$= \frac{\prod_{i=1}^{m} E\left[e^{sZ_i}\right]}{e^{sm\varepsilon}} \quad (\because Z_i \text{ are iid})$$

$$\leq \frac{\prod_{i=1}^{m} e^{\frac{s^2(b-a)^2}{8}}}{e^{sm\varepsilon}} \quad \text{(by Hoeffding inequality given below)}$$

Hoeffding Inequality: If X is a r.v. with mean zero & $a \leq X \leq b$, then

$$E\left[e^{sX}\right] \leq e^{s^2(b-a)^2/8} \quad \text{(refer to internet for proof of this)}$$

$$= e^{\frac{s^2 m(b-a)^2}{8} - sm\varepsilon}$$

$$\leq e^{-2m\varepsilon^2/(b-a)^2} \quad \text{(by taking the } s \text{ which minimizes the upper bound)}$$

IIIly we get, $P\left[R[f] - R_{emp}[f] > \varepsilon\right] \leq e^{-2m\varepsilon^2/(b-a)^2}$

$\therefore$ by union bound we get $P\left[|R_{emp}^{m}[f] - R[f]| > \varepsilon\right] \leq 2e^{-2m\varepsilon^2/(b-a)^2}$.

Hence Proved.