T: sequence of k examples

h_T: hypo on T

V: sequence of m−k examples

I: sequence of indices $(i_1 \dots i_k)$

Lemma 30.7

$$P\left[L_D(h_T) - L_V(h_T) \geq \sqrt{\frac{2L_V(h_T)\log(1/\delta)}{|V|}} + \frac{4\log(1/\delta)}{|V|}\right] \leq \delta$$

① $I = (i_1 \dots i_k) \in [m]^k$

② $S = (z_1 \dots z_m)$

③ $T = (z_{i_1} \dots z_{i_k})$ V = rest of

Theorem 30.2

int k

$B: Z^k \to H$

$A: Z^m \to H$

$m \geq 2k$

$A(S) = B(z_{i_1} \dots z_{i_k})$

$1-\delta$:

$$L_D(A(S)) \leq L_V(A(S)) + \sqrt{L_V(A(S))\frac{4k\log(m/\delta)}{m}} + \frac{8k\log(m/\delta)}{m}$$

Proof:

$I \in [m]^k$

$h_I = B(z_{i_1} \dots z_{i_k})$

$n = m-k$

$$P\left[\exists I \in [m]^k \text{ s.t. } L_D(h_I) - L_V(h_I) \geq \sqrt{\frac{2L_V(h_I)\log(1/\delta)}{n}} + \frac{4\log(1/\delta)}{n}\right]$$

$$\leq \sum_{I \in [m]^k} P\left[L_D(h_I) - L_V(h_I) \geq\right.$$

$\leq m^k \delta$

$m^k \delta = \delta'$

$$P\left[\forall I .\ L_D(h_I) - L_V(h_I) \leq \sqrt{\frac{2L_V(h_I)\log(1/\delta)}{n}} + \frac{4\log(1/\delta)}{n}\right]$$

$\left.\begin{matrix} m \geq 2k \\ m-k=n \end{matrix}\right\} \Rightarrow m-k=n \geq \frac{m}{2}$

$$L_D(A(S)) \leq L_V(A(S)) + \sqrt{L_V(A(S))\frac{4k\log(m/\delta)}{m}} + \frac{8k\log(m/\delta)}{m}$$

Coro

30.3 $L_V(A(S)) = 0$

$1-\delta$ Prob

$$L_D(A(S)) \leq \frac{8k\log(m/\delta)}{m}$$

Compression Schema

$A: Z^m \to [m]^d$

$B: Z^k \to H$

$\forall h \in H$

$(x_1, h(x_1)) \dots (x_m, h(x_m)) \to A$

$(x_{i_1}, h(x_{i_1}) \dots x_{i_k}, h(x_{i_k})) \to B \to h'$

$L_S(h') = 0$

Compression schema for unrealizable case

$L_S(h') \leq L_S(h)$

ERM(h)

$z_1 \dots z_m$

z

h'

PAC-Bayes

P: Prior distribution on H

Q: posterior distribution on H

$$\ell(Q, z) \stackrel{\text{def}}{=} \mathop{E}_{h \sim Q}[l(h, z)]$$

$$L_D(Q) \stackrel{\text{def}}{=} \mathop{E}_{h \sim Q}[L_D(h)]$$

$$L_S(Q) \stackrel{\text{def}}{=} \mathop{E}_{h \sim Q}[L_S(h)]$$

Theorem 31.7. Prob $1-\delta$

$$L_D(Q) \leq L_S(Q) + \sqrt{\frac{D(Q\|P) + \ln m/\delta}{2(m-1)}}$$

$$D(Q\|P) \stackrel{\text{def}}{=} \mathop{E}_{h \sim Q}[\ln(Q(h)/P(h))]$$

Proof:

$f(S)$

$$\mathop{P}_{S}[f(S) \geq \varepsilon] = \mathop{P}_{S}[e^{f(S)} \geq e^{\varepsilon}] \leq \frac{\mathop{E}_S[e^{f(S)}]}{e^{\varepsilon}}$$

$\Delta(h) = L_D(h) - L_S(h)$

$$f(S) = \sup_Q \left(2(m-1)\mathop{E}_{h \sim Q}(\Delta(h))^2 - D(Q\|P)\right)$$

$$2(m-1)\mathop{E}_{h \sim Q}(\Delta(h))^2 - D(Q\|P) = \mathop{E}_{h \sim Q}[\ln(e^{2(m-1)\Delta(h)^2} P(h)/Q(h))]$$

$$\| \quad \mathbb{E}_{h \sim Q}\left(-\ln\frac{Q(h)}{P(h)} + \ln\left(e^{\Delta(h)^2 \cdot 2(m-1)}\right)\right)$$

$$\leq \ln \mathop{E}_{h \sim Q}[e^{2(m-1)\Delta(h)^2} P(h)/Q(h)]$$

$$= \ln \mathop{E}_{h \sim P}[e^{2(m-1)\Delta(h)^2}]$$

$$\mathop{E}_S[e^{f(S)}] \leq \mathop{E}_S \mathop{E}_{h \sim P}[e^{2(m-1)\Delta(h)^2}]$$

$$\mathop{E}_S[e^{f(S)}] \leq \mathop{E}_{h \sim P}\mathop{E}_S[e^{2(m-1)\Delta(h)^2}]$$

$\forall h$, $\mathop{E}_S[e^{2(m-1)\Delta(h)^2}] \leq m$

$\mathop{P}_S[\Delta(h) \geq \varepsilon] \leq e^{-2m\varepsilon^2}$

$\mathop{E}_S[e^{2(m-1)\Delta(h)^2}] \leq m$

$\mathop{P}_S[f(S) \geq \varepsilon] \leq \frac{m}{e^{\varepsilon}}$

$\varepsilon = \ln(m/\delta)$

With Prob at least $1-\delta$

$\forall Q$.

$$2(m-1)\mathop{E}_{h \sim Q}(\Delta(h))^2 - D(Q\|P) \leq \varepsilon = \ln(m/\delta)$$

$$\mathop{E}_{h \sim Q}(\Delta(h))^2 \leq \frac{\ln(m/\delta) + D(Q\|P)}{2(m-1)}$$

$$\mathop{E}_{h \sim Q}(\Delta(h)) \leq \sqrt{}$$

$n \in \mathbb{N}$, $x \in \mathbb{R}$ s.t. $n \leq x < n+1$

$P(x \in [n, n+1)) = p$

$P\left(x \in \bigcup_n [n, n+1)\right) = \sum_n p \leq 1$

$[0,1] \cdot f(x) = \frac{1}{|b-a|}$

EMX

Let X be some domain set
let $\mathcal{F} \subseteq P^X$ and let $\mathcal{D}$ be some
distribution over X

Given $S \sim \mathcal{D}^m$ find $f \in \mathcal{F}$
that maximizes $\mathop{\mathbb{E}}_{\mathcal{D}}(f) = \int_{x \in f} \mathcal{D}(x)$

An algo A is an (ε, δ)−EMX learner
if for some $m = m(\varepsilon, \delta)$

$$\mathop{\mathbb{P}}_{S \sim \mathcal{D}^m}\left(\mathop{\mathbb{E}}_{\mathcal{D}}(A(S)) \geq \sup_{f \in \mathcal{F}} \mathop{\mathbb{E}}_{\mathcal{D}}(f) - \varepsilon\right) \geq 1-\delta$$

The EMX learnability of $\mathcal{F}^* = \{f \mid f \text{ is finite subset of } [0,1]\}$
wrt. is independent of ZFC