$$err = \underbrace{opt}_{\text{우시}} + \underline{app} + \underline{gen}$$

$\boxed{\text{UAT}}$

barrun

복잡도 얘기 X

라데마커

Week 8. 10/21

$\hat{f}$ 이 학습으로 얻어진거. $\hat{f} \in F_\delta$ F_δ는 복잡도 δ 이내로 제한된 공간

$$R(\hat{f}) - \inf_{f \in F} R(f) = \text{exess risk}$$

$R(\hat{f})$: population risk

F ← 전체공간

$$= \left(R(\hat{f}) - \inf_{F_\delta} R(f) \right) + \left(\inf_{F_\delta} R(f) - \inf_{F} R(f) \right)$$

approximation err : 우리가 고려가능한 공간과, 전체 공간 사이의 차이

$$= R(\hat{f}) - \hat{R}(\hat{f}) + \underbrace{\hat{R}(\hat{f}) - \inf_{F_\delta} \hat{R}(f)}_{\text{최적값 못찾은 err}} + \inf_{F_\delta} \hat{R}(f) - \inf_{F_\delta} R(f) + \mathcal{E}_{app}$$

$= \mathcal{E}_{opt}$

삼각부등식

$$\leq \underbrace{\left| R(\hat{f}) - \hat{R}(\hat{f}) \right|}_{①} + \mathcal{E}_{opt} + \mathcal{E}_{app} + \underbrace{\left| \inf_{F_\delta} \hat{R}(f) - \inf_{F_\delta} R(f) \right|}_{②}$$

① $\leq \sup_{F_\delta} \left| R(f) - \hat{R}(f) \right|$ 자명

② $\leq \sup_{F_\delta} \left| R(f) - \hat{R}(f) \right|$ 임을 증명하자.

Claim. $\mathcal{F}$: 함수공간 일때

$$\left| \inf_{f\in\mathcal{F}} \hat{R}(f) - \inf_{f\in\mathcal{F}} R(f) \right| \le \sup_{f\in\mathcal{F}} \left| \hat{R}(f) - R(f) \right|$$ 증명하자.

let $\Delta := \sup \left| \hat{R}(f) - R(f) \right| \rightarrow \left| \hat{R}(f) - R(f) \right| \le \Delta$ 자명

$\rightarrow -\Delta \le \hat{R}(f) - R(f) \le \Delta \rightarrow \underbrace{R(f) - \Delta \le \hat{R}(f)}_{\text{① inf 취함}} \le R(f) + \Delta$ ② inf 취함

①

$$\inf(R(f) - \underset{\text{상수}}{\Delta}) \le \inf \hat{R}(f)$$

$\rightarrow \inf R(f) - \Delta \le \inf \hat{R}(f) \longrightarrow \inf R(f) - \inf \hat{R}(f) \le \Delta$ 얻음

② $\inf \hat{R}(f) \le \inf(R(f) + \Delta) = \inf R(f) + \Delta$

$\therefore \inf \hat{R}(f) - \inf R(f) \le \Delta$

따라서 $\left| \inf \hat{R}(f) - \inf R(f) \right| \le \Delta$ 이다

이전수업 이어서 $B>0$, function class $F_{m,\sigma,B}$ 고려하자

$$F_{m,\sigma,B} = \{ f_\theta \in F_{m,\sigma} : C(\theta) \le B \}$$

wher $C(\theta) = \sum_{j=1}^{m} |\beta_j| \, \|w_j\|_2$ with $f_\theta(x) = \sum_{j=1}^{m} \beta_j \sigma(w_j^T x) \in F_{m,\sigma}$

만약 $\|z_i\|_2 \le C$ 가 $i = 1 \sim n$ 이라면

$\sigma = \text{ReLU}$ 로 두었을때 $\text{Rad}(F_{m,\sigma,B}) \le \dfrac{2BC}{\sqrt{m}}$ 상한 존재 한다.

증명 우선 $\alpha\sigma(x) = \sigma(\alpha x)$ note 하자. $(\forall \alpha > 0)$

따라서 각 $\lambda_j > 0$, $j = 1, \dots, m$ 에 대해 새 파라미터

$\theta = \{ (\beta_j, w_j) \}_{j=1\sim m}$ 파라미터 $\rightarrow \theta' = \{ (\lambda_j \beta_j, \frac{w_j}{\lambda_j}) \}_{j=1\sim m}$ 정의

$\rightarrow \phi(w_j x) = \|w_j\|_2 \, \phi(\bar{w}_j x) \rightarrow$ normalization

라데마커 RV let $\xi_i = \pm 1$ iid 확률 $\frac{1}{2}$ 정의

$$\text{Rad}(F_{m,\sigma,B}) = \mathbb{E}_\xi \left[\sup_{f_\theta \in F_{m,\sigma,B}} \frac{1}{n} \sum_{i=1}^{n} \xi_i f_\theta(z_i) \right]$$ 라데마커 복잡도 정의

$$= \mathbb{E}_\xi \left[\sup_{\theta : C(\theta) \le B} \frac{1}{n} \sum_{i}^{n} \xi_i \sum_{j}^{m} \underline{\beta_j \sigma(w_j^T z_i)} \right]$$

normalize 적용하자

$$= \frac{1}{n} \mathbb{E}_\xi \left[\sup_{\theta : C(\theta) \le B} \sum_{i=1}^{n} \xi_i \sum_{j=1}^{m} \beta_j \|w_j\|_2 \, \sigma(\bar{w}_j^T z_i) \right]$$

순서변경

$$= \frac{1}{n} \mathbb{E}_\xi \left[\sup \sum_{j=1}^{m} \beta_j \|w_j\|_2 \sum_{i=1}^{n} \xi_i \, \sigma(\bar{w}_j^T z_i) \right]$$ 부등호 유도

$$\le \frac{1}{n} \mathbb{E}_\xi \left[\sup \sum_{j=1}^{m} \beta_j \|w_j\|_2 \max_{1 \le k \le m} \left| \sum_{i=1}^{n} \xi_i \, \sigma(\bar{w}_k^T z_i) \right| \right]$$

by $C(\theta) \le B$ 정의가 $\sum_{j=1}^{m} |\beta_j| \, \|w_j\|_2 \le B$

$$\le \frac{B}{n} \mathbb{E}_\xi \left[\sup_{\theta : C(\theta) \le B} \max_{1 \le k \le m} \left| \sum_{i=1}^{n} \xi_i \, \sigma(\bar{w}_k^T z_i) \right| \right]$$

$$= \frac{B}{n} \mathbb{E}_\xi \left[\sup_{\|\bar{w}\|_2 = 1} \left| \sum_{i=1}^{n} \xi_i \, \sigma(\bar{w}^T z_i) \right| \right]$$ $\|\bar{w}_2\| \le 1$ 로 공간 확장 $\underline{\text{sup도 커짐}}$

$$\le \frac{B}{n} \mathbb{E}_\xi \left[\sup_{\|\bar{w}\| \le 1} \left| \sum_{i=1}^{n} \xi_i \, \sigma(\bar{w}^T z_i) \right| \right]$$

(*) 이후증명

$$\le \frac{2B}{n} \mathbb{E}_{\xi}\left[\sup_{\|\bar{w}\|_2 \le 1} \sum_{i=1}^{n} \xi_i \, \sigma(\bar{w}^T z_i) \right] = 2B \, Rad(F_1)$$

where $F_1 = \{ x \mapsto \sigma(w^T x)$ where $\bar{w} \in \mathbb{R}^d, \ \|w\|_2 \le 1 \}$

F_1 에서

set $\tilde{F}_1 = \{ x \to \bar{w}^T x \mid \bar{w} \in \mathbb{R}^d, \ \|\bar{w}\|_2 \le 1 \}$ ReLU 뺀 집합 $\tilde{F}_1$

ReLU 떼면 복잡도 줄어듦 $\therefore$ $Rad(F_1) \underset{(**)}{\le} Rad(\tilde{F}_1) \underset{(***)}{\le} \frac{C}{\sqrt{n}}$

(*), (**), (***) 는 아래에서 증명.

여튼, 샘플수 n을 늘리면 $Rad_{m,\sigma,B} \le \frac{C}{\sqrt{n}}$ 이므로 복잡도가 줄어든다

(**) 증명 : 간단함. $\phi_i : \mathbb{R} \to \mathbb{R}$, $\phi_i(0) = 0$, L-Lip 일때

$$\mathbb{E}_{\xi}\left[\sup_{f \in F} \sum_{i=1}^{n} \xi_i \, \phi_i(f(z_i)) \right] \le L \, \mathbb{E}_{\xi}\left[\sup_{F} \sum_{i=1}^{n} \xi_i f(z_i) \right]$$

$\phi(f) - \phi(0) \le L|f - 0|$ 이니까 당연해보임

(*) 증명 let $F^{\pm} = F \cup \{-F\}$

$$\sup_{f \in F} \left| \sum \xi_i f(z_i) \right| = \sup_{g \in F^{\pm}} \sum \xi_i \, g_i(z_i)$$ 자명, + − 선택가능 하니까

그러면 $$\sup_{g \in F^{\pm}} \sum_i \xi_i \, g(z_i) = \max\left\{ \underbrace{\sup_{f \in F} \sum \xi_i f(z_i)}_{A항}, \underbrace{\sup_{f \in F} \sum -\xi_i f(z_i)}_{B항} \right\}$$

$$= \max(A, B) \le A + B$$

이래서 계수 2배 나옴. 나머진 pdf 참고

(***) 증명

$\mathcal{F} = \{ x \mapsto \langle w, x \rangle \mid \|w\| \le B \}$ 인 집합, $\|z_i\| \le C$ 일때

$$Rad(\mathcal{F}) = \frac{1}{n} \mathbb{E}_{\xi} \left[\sup_{\|w\| \le B} \sum_{i=1}^{n} \xi_i \langle w, z_i \rangle \right] = \frac{1}{n} \mathbb{E}_{\xi} \left[\sup \langle w, \sum_{i=1}^{n} \xi_i z_i \rangle \right]$$

코시 쓰차 $\le \frac{B}{n} \mathbb{E}_{\xi} \left[\| \sum_{i=1}^{n} \xi_i x_i \| \right]$ $\quad \mathbb{E}\|x\| \le \sqrt{\mathbb{E}\|x\|^2}$ 젠센 적용

$$\le \frac{B}{n} \sqrt{\mathbb{E}_{\xi} \| \sum \xi_i x_i \|^2}$$

$$= \frac{B}{n} \sqrt{\mathbb{E}_{\xi} \sum_{i,j} \xi_i \xi_j \langle z_i, z_j \rangle} = \frac{B}{n} \sqrt{\sum^{n} \|z_i\|^2} = \frac{B}{n} \sqrt{nC^2} = \frac{BC}{\sqrt{n}}$$

ξ Rad RV의 독립성 때문 $\qquad$ 같을때만 1

(참고) $\mathbb{E}[\xi_i] = 1 \times \frac{1}{2} - 1 \times \frac{1}{2} = 0$

$\mathbb{E}[\xi_i^2] = 1^2 \frac{1}{2} + (-1)^2 \frac{1}{2} = 1$

$\mathbb{E}[\xi_i \xi_j] = \mathbb{E}[\xi_i]\mathbb{E}[\xi_j] = \begin{cases} 0 & (i \ne j) \\ 1 & (i = j) \end{cases}$

여튼 결론: $Rad(\mathcal{F}_{m,\sigma,B}) \le \frac{2BC}{\sqrt{n}}$

이제 $\|Err\| \le \|app\| + \|gen\| + \|opt\|$ 다 해결함