CLT Homework 2   20200130 Yujun Kim / /

#4.1

① $\inf_{h\in\mathcal{H}} R(h) \le \mathbb{E}_{S\sim D^m}[R(h_S^{ERM})]$

As $h_S^{ERM} \in \mathcal{H}$, $\inf_{h\in\mathcal{H}} R(h) \le R(h_S^{ERM})$ for any $S\sim D^m$.
Take expectation over $S\sim D^m$ on both side to obtain above inequality

② $\mathbb{E}_{S\sim D^m}[\hat{R}_S(h_S^{ERM})] \le \inf_{h\in\mathcal{H}} R(h)$

Let $\varepsilon>0$ be given. $\exists h_\varepsilon$ s.t. $R(h_\varepsilon) \le \inf_{h\in\mathcal{H}} R(h)+\varepsilon$ (By definition of infimum)
For each $S\sim D^m$, $\hat{R}_S(h_S^{ERM}) \le \hat{R}_S(h_\varepsilon)$
Thus, $\mathbb{E}_{S\sim D^m}[\hat{R}_S(h_S^{ERM})] \le \mathbb{E}_{S\sim D^m}[\hat{R}_S(h_\varepsilon)] = R(h_\varepsilon) \le \inf_{h\in\mathcal{H}} R(h)+\varepsilon$

As $\varepsilon>0$ is arbitrary, $\mathbb{E}_{S\sim D^m}[\hat{R}_S(h_S^{ERM})] \le \inf_{h\in\mathcal{H}} R(h)$.

#4.2

$\Phi(u)=(1+u)^2$. $L_\Phi(x,u) := \eta(x)\Phi(-u) + (1-\eta(x))\Phi(u)$.
$h^*(x) = \eta(x)-\frac{1}{2}$, $h^*_\Phi(x) = \operatorname{argmin}_{u\in[-\infty,\infty]} L_\Phi(x,u)$

① $L_\Phi(x,0) = \Phi(0) = 1$
② $L_\Phi(x,h^*_\Phi(x)) = 4\eta(x)(1-\eta(x))$
Now, to find $h^*_\Phi(x)$,

$$\frac{\partial}{\partial u}L_\Phi(x,u) = -\eta(x)\Phi'(-u) + (1-\eta(x))\Phi'(u)$$
$$= -\eta(x)(1-u) + (1-\eta(x))(1+u)$$
$$= -2\eta(x)+1+u = 0.$$

So that $h^*_\Phi(x) = 2\eta(x)-1$.
Hence, $L_\Phi(x,h^*_\Phi(x)) = \eta(x)\Phi(1-2\eta(x)) + (1-\eta(x))\Phi(2\eta(x)-1)$

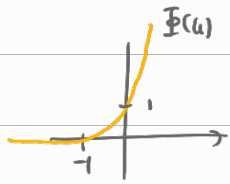
$$= 4\eta(x)(1-\eta(x))^2 + 4(1-\eta(x))\eta(x)^2$$
$$= 4\eta(x)(1-\eta(x))$$

$$L_\Phi(x,0) - L_\Phi(x, h^*_\Phi(x)) = 1 - 4\eta(x)(1-\eta(x))$$
$$= (2\eta(x)-1)^2$$
$$= 2^2 |\eta(x) - \tfrac{1}{2}|^2$$

③ Hence, $|h^*(x)|^s = |\eta(x) - \frac{1}{2}|^s \leq c^s \left[L_\Phi(x,0) - L_\Phi(x, h^*_\Phi(x))\right]$
for $s=2,\ c=\frac{1}{2}$

④ $R(h) - R^* \leq \left[\mathcal{L}_\Phi(h) - \mathcal{L}^*_\Phi\right]^{\frac{1}{2}}$ holds by thm 4.7

#4.3



① $L_\Phi(x,0) = \Phi(0) = \max(0,1)^2 = 1$

② $L_\Phi(x, h^*_\Phi(x)) = 4\eta(x)(1-\eta(x))$.

As $u \mapsto L_\Phi(x,u)$ is convex differentiable, $\frac{\partial \Phi}{\partial x}(u) = \max\{0, 1+u\}$,

$$\frac{\partial}{\partial u} L_\Phi(x,u) = -\eta(x)\Phi'(-u) + (1-\eta(x))\Phi'(u)$$
$$= -\eta(x)\max\{0, 1-u\} + (1-\eta(x))\max\{0, 1+u\}$$

i) $u > 1 \Rightarrow \frac{\partial}{\partial u} L_\Phi(x,u) = (1-\eta(x))(1+u) \geq 0$

ii) $u < -1 \Rightarrow \frac{\partial}{\partial u} L_\Phi(x,u) = -\eta(x)(1-u) \leq 0$

iii) $-1 \leq u \leq 1 \Rightarrow \frac{\partial}{\partial u} L_\Phi(x,u) = -\eta(x)(1-u) + (1-\eta(x))(1+u)$
$$= 1 + u - 2\eta(x)$$

Thus, $u = 2\eta(x) - 1 \in [-1, 1]$ has $\frac{\partial}{\partial u} L_\Phi(x,u) = 0$.

i.e. $h^*_\Phi(x) = 2\eta(x) - 1$

It follows $L_\Phi(x, h^*_\Phi(x)) = \eta(x)\Phi(1-2\eta(x)) + (1-\eta(x))\Phi(2\eta(x)-1)$
$$= \eta(x)\max(0, 2-2\eta(x))^2 + (1-\eta(x))\max(0, 2\eta(x))^2$$
$$= 4\eta(x)(1-\eta(x)) \quad (\because 2(1-\eta(x)) \geq 0,\ 2\eta(x) \geq 0).$$

The rest follows #4.2. $L_\Phi(x,0) - L_\Phi(x, h^*_\Phi(x)) = (2\eta(x)-1)^2 = 2^2|\eta(x) - \frac{1}{2}|^2$

③ $|h^*(x)|^2 = |\eta(x) - \frac{1}{2}|^2 = (\frac{1}{2})^2 \left[L_\Phi(x,0) - L_\Phi(x, h^*_\Phi(x))\right]$.

④ $R(h) - R^* \leq \left[\mathcal{L}_\Phi(h) - \mathcal{L}^*_\Phi\right]^{\frac{1}{2}}$ holds by thm 4.7.

#5.2

① Remarks on what we've learned

Theorem 5.9 states $R(h) \le \hat{R}_{S,\rho}(h) + \frac{4}{\rho}\mathfrak{R}_m(H) + \sqrt{\frac{\log\log_2 \frac{2r}{\rho}}{m}} + \sqrt{\frac{\log\frac{2}{\delta}}{m}}$

for any $\rho \in (0, r]$

Theorem 5.8 gives $\mathbb{P}\left[\sup_{h\in H} R(h) - \hat{R}_{S,\rho}(h) > \frac{2}{\rho}\mathfrak{R}_m(H) + \sqrt{\frac{\log\frac{1}{\delta}}{2m}}\right] \le \delta$

$\Rightarrow$ For $\varepsilon > 0, \rho > 0$, $\mathbb{P}\left[\sup_{h\in H} R(h) - \hat{R}_{S,\rho}(h) - \frac{2}{\rho}\mathfrak{R}_m(H) - \varepsilon\right] \le e^{-2m\varepsilon^2}$ -- (*)

② Let $\delta > 0$. Define $\rho_k = \gamma^{-k}$, $\varepsilon_k = \varepsilon + \sqrt{\frac{\log k}{m}}$ for $\varepsilon = \sqrt{\frac{\log 2/\delta}{2m}}$

Denote the event $\left\{\sup_{h\in H} R(h) - \hat{R}_{S,\rho_k}(h) - \frac{2}{\rho_k}\mathfrak{R}_m(H) - \varepsilon_k > 0\right\}$ as A_k

$\mathbb{P}[A_k \text{ happens for some } k \ge 1] \le \delta$.

∵) $\mathbb{P}[A_k \text{ happens for some } k]$

$$\le \sum_{k=1}^{\infty} \mathbb{P}[A_k \text{ happens}] \quad \text{(Union Bound)}$$
$$\le \sum_{k=1}^{\infty} e^{-2m\varepsilon_k^2} \quad \text{(By (*))}$$
$$= \sum_{k=1}^{\infty} \frac{1}{k^2} e^{-2m\varepsilon^2}$$
$$= \frac{\pi^2}{6} e^{-2m\varepsilon^2} \le 2e^{-2m\varepsilon^2} = \delta.$$

③ We prove $\mathbb{P}\left[\overbrace{\exists \rho\in(0,1] \text{ s.t. } \sup_{h\in H} R(h) - \hat{R}_{S,\rho}(h) - \frac{2\gamma}{\rho}\mathfrak{R}_m(H) - \sqrt{\frac{\log\log_\gamma \frac{\gamma}{\rho}}{m}} - \sqrt{\frac{\log\frac{2}{\delta}}{2m}} > 0}^{=:\ \text{Event } A}\right] \le \delta$

∵) If $\exists$ such $\rho \in (0,1]$, then $\exists k \ge 1$ with $\rho_k < \rho \le \rho_{k-1} = \gamma\rho_k$

Then $1/\rho_k \le \gamma/\rho$, $\hat{R}_{S,\rho_k}(h) \le \hat{R}_{S,\rho}(h)$, $\sqrt{\log k} = \sqrt{\log\log_\gamma(1/\rho_k)} \le \sqrt{\log\log_\gamma(\gamma/\rho)}$

Thus, $\sup_{h\in H} R(h) - \hat{R}_{S,\rho_k}(h) - \frac{2}{\rho_k}\mathfrak{R}_m(H) - \sqrt{\frac{\log\log_\gamma(1/\rho_k)}{m}} + \sqrt{\frac{\log\frac{2}{\delta}}{2m}}$

$$\ge \sup_{h\in H} R(h) - \hat{R}_{S,\rho}(h) - \frac{2\gamma}{\rho}\mathfrak{R}_m(H) - \sqrt{\frac{\log\log_\gamma(\gamma/\rho)}{m}} - \sqrt{\frac{\log\frac{2}{\delta}}{2m}}$$

> 0.

Hence, A_k happens for some $k \ge 1$

Thus, $\mathbb{P}[A \text{ happens}] \le \mathbb{P}[A_k \text{ happens for some } k] \le \delta$. By considering complement,

$\mathbb{P}\left[\forall \rho\in(0,1],\ \sup_{h\in H} R(h) - \hat{R}_{S,\rho_k}(h) \le \frac{2\gamma}{\rho}\mathfrak{R}_m(H) + \sqrt{\frac{\log\log_\gamma(\gamma/\delta)}{m}} + \sqrt{\frac{\log(2/\delta)}{m}}\right] \ge 1-\delta$

#5.3

① Primal Formulation

The usual soft margin SVM is

$$\min_{w,b,\xi} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m} \xi_i \quad \text{such that} \quad \xi_i \geq 0$$
$$y_i\langle w, x_i\rangle + b \geq 1 - \xi_i.$$

The weighted soft margin SVM is

$$\min_{w,b,\xi} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m} p_i\xi_i \quad \text{such that} \quad y_i(\langle w, x_i\rangle + b) \geq 1 - \xi_i$$
$$\xi_i \geq 0$$

② Dual Formulation

$$\mathcal{L}(w, b, \xi, \alpha, \beta) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{M} p_i\xi_i + \sum_{i=1}^{M}\alpha_i(1 - \xi_i - y_i(\langle w, x_i\rangle + b)) + \sum_{i=1}^{m}\beta_i(-\xi_i)$$

$$\nabla_w \mathcal{L} = w - \sum_{i=1}^{m}\alpha_i y_i x_i = 0$$

$$\nabla_b \mathcal{L} = \sum_{i=1}^{m} -\alpha_i y_i = 0$$

$$\frac{\partial}{\partial \xi_i}\mathcal{L} = Cp_i - \alpha_i - \beta_i = 0.$$

Thus $D(\alpha, \beta) = \inf_{w,b,\xi} \mathcal{L}(w, b, \xi, \alpha, \beta)$

$$= \frac{1}{2}\|\sum_i \alpha_i y_i x_i\|^2 + \sum_i \underbrace{\cancel{(Cp_i - \alpha_i - \beta_i)}}_{0}\xi_i + \sum_i \alpha_i$$
$$- \sum_i \alpha_i y_i \langle \sum_j \alpha_j y_j x_j, x_i\rangle - \underbrace{\cancel{\sum_i \alpha_i y_i}}_{0} b$$

$$= \sum_i \alpha_i - \frac{1}{2}\sum_{i,j}\alpha_i\alpha_j y_i y_j \langle x_i, x_j\rangle$$

such that $\alpha_i \geq 0$, $\beta_i = Cp_i - \alpha_i \geq 0$ so that $Cp_i \geq \alpha_i \geq 0$.

Hence, dual formulation of weighted soft margin SVM is

$$\max_{\alpha, \beta} D(\alpha, \beta) = \sum_i \alpha_i - \frac{1}{2}\sum_{ij}\alpha_i\alpha_j y_i y_j \langle x_i, x_j\rangle \quad \text{such that} \begin{cases} 0 \leq \alpha_i \leq Cp_i \\ \sum \alpha_i y_i = 0 \end{cases}$$

/ /

#5.7 $S \subseteq \{x \in \mathbb{R}^n : \|x\| \le r\}$. $\mathcal{H} = \{x \mapsto \mathrm{sgn}(w \cdot x) : \min_{x \in S} |w \cdot x| = 1, \|w\| \le \Lambda\}$

(a) Let $\{x_1, \cdots, x_d\}$ be a shattered set, $y = (y_1, \cdots, y_d) \in \{\pm 1\}^d$.

$\exists\, h_w \in \mathcal{H}$ s.t. $h_w(x_i) = \mathrm{sgn}(w \cdot x_i) = y_i \ \forall i$, with $\min_{x \in S} |w \cdot x| = 1$, $\|w\| \le \Lambda$.

Then, $1 = y_i \mathrm{sgn}(w \cdot x_i) \le y_i (w \cdot x_i)$ $(\because |w \cdot x_i| \ge 1)$

$$\Rightarrow d = \sum_{i=1}^{d} 1 = \sum_{i=1}^{d} y_i \,\mathrm{sgn}(w \cdot x_i) \le \sum_{i=1}^{d} y_i (w \cdot x_i) = \Big(\sum_{i=1}^{d} y_i x_i\Big) \cdot w$$
$$\le \Big\|\sum_{i=1}^{d} y_i x_i\Big\| \cdot \|w\|$$
$$\le \Lambda \Big\|\sum_{i=1}^{d} y_i x_i\Big\|$$

(b) Consider $\mathbb{E}_y\Big[\Big\|\sum_{i=1}^{n} y_i x_i\Big\|^2\Big] = \mathbb{E}_y\Big[\sum_i \|x_i\|^2 + \sum_{i \ne j} y_i y_j \langle x_i, x_j \rangle\Big]$

$y_i \sim \mathrm{unif}(\pm 1)$

$$= \sum_{i=1}^{d} \|x_i\|^2 \quad (\because \mathbb{E}\, y_i y_j = \mathbb{E} y_i \,\mathbb{E} y_j = 0 \ \forall i \ne j)$$

By Jensen, $\Big\{\mathbb{E}_y \Big\|\sum_{i=1}^{d} y_i x_i\Big\|\Big\}^2 \le \mathbb{E}_y\Big[\Big\|\sum_{i=1}^{d} y_i x_i\Big\|^2\Big] = \sum_{i=1}^{d} \|x_i\|^2$ --- $(*)$

- $\varphi(t) = t^2$ is convex
- $\varphi\Big(\mathbb{E}_y \Big\|\sum_{i=1}^{d} y_i x_i\Big\|\Big) \le \mathbb{E}_y\Big[\varphi\Big(\Big\|\sum_{i=1}^{d} y_i x_i\Big\|\Big)\Big]$

Take expectation on y for the inequality from (a):

$$d \le \Lambda\, \mathbb{E}_y\Big[\Big\|\sum_{i=1}^{d} y_i x_i\Big\|\Big]$$
$$\le \Lambda \sqrt{\sum_{i=1}^{d} \|x_i\|^2} \quad (\text{by } (*))$$

(c) $\sqrt{\sum_{i=1}^{d} \|x_i\|^2} \le \sqrt{d r^2}$

From (b), $d \le \Lambda \sqrt{d r^2} = \Lambda r \sqrt{d}$

$\Rightarrow \sqrt{d} \le \Lambda r$

$\Rightarrow d \le \Lambda^2 r^2$