

Section 3.

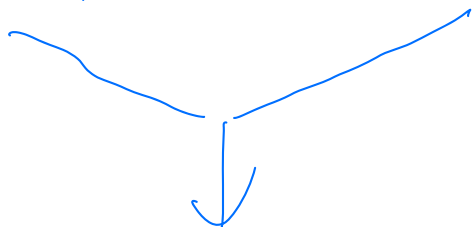
lemma 3.1

$$P(|Z_i| \leq C_2 \frac{\|Z\|_{L_2}}{\sqrt{\varepsilon}}) \geq 1 - 2\exp(-C_1 \varepsilon N)$$

+

lemma 3.7

$$P(|Z_i| \geq K \|Z\|_{L_2}) \geq 1 - 2\exp(-C N \varepsilon)$$



Corollary 3.8

$$P(K \|Z\|_{L_2} \leq |Z_i| \leq C_2 \frac{\|Z\|_{L_2}}{\sqrt{\varepsilon}}) \geq 1 - 2\exp(-C_1 \varepsilon N)$$

uniform estimate
on a class



Theorem 3.11

with $r > r_Q(H, N, \zeta_1, \zeta_2)$

$$P\left(\left(\frac{k_0}{2}\right) \|h\|_{L_2} \leq |h(x_j)| \leq C_5 \left(k_0 + \frac{1}{\sqrt{\varepsilon}}\right) \|h\|_{L_2}\right) \geq 1 - 2\exp(-C_0 \varepsilon^2 N)$$

Theorem 3.11 can hold uniformly with $r > 2r_Q(H, N, \zeta_1, \zeta_2)$

Corollary 3.13

$r = 2r_Q(F - F, N, \zeta_1, \zeta_2)$, $\|f_1 - f_2\|_{L_2} \geq r$

$$P(|(f_1 - f_2)(x_j)| \geq \left(\frac{k_0}{2}\right) \|f_1 - f_2\|_{L_2}) \geq 1 - 2\exp(-C_0 \varepsilon^2 N)$$

or $\inf \frac{1}{N} \sum_{i=1}^N \left(\frac{f - f^*}{\|f - f^*\|_{L_2}} \right)^2(x_i) \geq \frac{\varepsilon k_0^2}{16}$

$\{f \in F : \|f - f^*\|_{L_2} \geq 2r_Q\}$

using the fact that $Q_{f-f^*}(x) \geq l'(z)(f-f^*)^2(x)$, where $Q_{f-f^*} = \int_0^{f+(f-f^*)} (l'(w) - l'(x)) dx$
 $\Rightarrow P_N Q_{f-f^*} \geq \frac{1}{N} \sum_{i=1}^N l'(z)(f-f^*)^2(x)$

Next
to bound the Quadratic term.

Theo 4.1 & Theo 4.2 & Theo 4.6

l is strongly
convex

l is general
 $W \perp\!\!\!\perp X$

l is strongly convex in a neighborhood of 0
 $W \perp\!\!\!\perp X$

★ Theo. 4.6 is more general case, with $\|f - f^*\|_{L_2}^* \geq 2r_Q$

$$P(P_N Q_{f-f^*} \geq c, \varepsilon k_0^2 P(0, t) \|f - f^*\|_{L_2}^2) \geq 1 - 2 \exp(-c_0 \varepsilon^2 N)$$

$$t = c_2(k_0 + \frac{1}{\sqrt{\varepsilon}}) (\|f - f^*\|_{L_2} + \|\xi\|_{L_2})$$

when F is bounded in L_2 : $t = C(\varepsilon, k_0) (\|\xi\|_{L_2} + \text{diam}(F, L_2))$
 $\|f - f^*\|_{L_2} \leq \text{diam}(F, L_2)$

In section 5 we show that we can improve t to the order of $\|\xi\|_{L_2}$.

Theo 4.6

$$\|f - f^*\|_{L_2} \geq 2r_Q$$

$$P_N Q_{f-f^*} \geq 0 \|f - f^*\|_{L_2}$$

Lemma 5.2

$$\|f - f^*\|_{L_2} \geq 2r_m \left(\frac{\theta}{r_0}, \frac{\delta}{\varepsilon} \right)$$

$$P_N M_{f-f^*} \leq \frac{\theta}{4} \|f - f^*\|_{L_2}^2$$

Theo 5.4

$$\left\{ t_1 = 0, t_2 = c_0(k_0, \varepsilon) (\|\xi\|_{L_2} + \text{diam}(F, L_2)) \right\}$$

$$\|\hat{f} - f^*\|_{L_2} \leq 2 \max \{ r_Q, r_m \}$$

$$E \mathcal{L}_{\hat{f}} \leq 2(\theta + \beta) \max \{ r_Q^2, r_m^2 \}$$

Now improve the choice of t to be $2C_0(k_0 + \frac{1}{\sqrt{\varepsilon}}) \max\{\|\xi\|_{L_2}, r_Q\}$

$\Rightarrow$ also show that $\hat{f} \in F : \|\hat{f} - f^*\|_{L_2} \leq \max\{\|\xi\|_{L_2}, r_Q\}$

Lemma 5.6

$$P_N Q_{\lambda}(f - f^*) \geq \lambda \|\theta\|_{L_2} \max\{\|\xi\|_{L_2}^2, 4r_Q^2\}$$

+

$$\theta = C_1 \varepsilon k_0^2 P(0, t),$$

$$t = 2C_0(k_0 + \frac{1}{\sqrt{\varepsilon}}) \max\{\|\xi\|_{L_2}, r_Q\},$$

$$\text{Assumption: } r_Q(\frac{\theta}{16}, \frac{\delta}{2}) \leq \max\{\|\xi\|_{L_2}, 2r_Q\}$$



Theo. 5.5

$$P(\|\hat{f} - f^*\|_{L_2} \leq \max\{\|\xi\|_{L_2}, 2r_Q\}) \geq 1 - \delta - 2\exp(-C_0 N \varepsilon^2)$$

$$\Rightarrow \hat{f} \in F \cap \max\{\|\xi\|_{L_2}, 2r_Q\} D_{f^*}$$



Theo. 5.7

$$(t_1=0, t_2=C_0(\varepsilon, k_0)\|\xi\|_{L_2})$$

Theo. 5.8

l is strongly convex in a interval $[-\delta, \delta]$

$$\text{If } r_m\left(\frac{\theta}{16}, \frac{\delta}{2}\right) \leq \max\{\|\xi\|_{L_2}, 2r_Q\}$$

$$\text{Prob. } 1 - \delta - 2\exp(-C_4 N \varepsilon^2)$$

$$\bullet \|\hat{f} - f^*\|_{L_2} \leq 2 \max\{r_Q, r_m\left(\frac{\theta}{16}, \frac{\delta}{2}\right)\}$$

$$\bullet E \mathcal{L} \hat{f} \leq 2(\theta + \beta) \max\{r_Q^2, r_m^2\left(\frac{\theta}{16}, \frac{\delta}{2}\right)\}$$

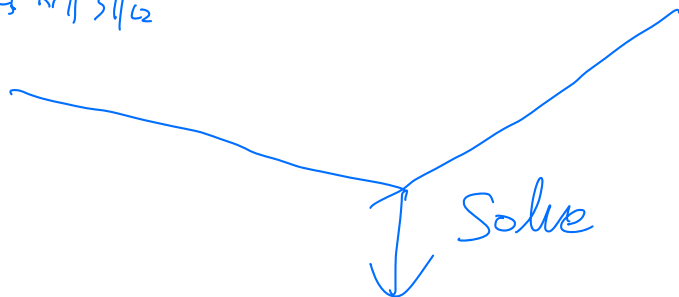
$$\bullet \text{If } W \perp X \text{ can take } t_1 = C_4 k \|\xi\|_{L_2}$$

$$\text{Assume } \|\xi\|_{L_2} \leq C_0 r_Q. \quad r_m\left(\frac{\theta}{16}, \frac{\delta}{2}\right) \leq \delta$$

$$\text{Prob. } 1 - \delta - 2\exp(-C_4 N \varepsilon^2)$$

$$\bullet \|\hat{f} - f^*\|_{L_2} \leq 2 \max\{r_Q, r_m\left(\frac{\theta}{16}, \frac{\delta}{2}\right)\}$$

$$\bullet E \mathcal{L} \hat{f} \leq 2(\theta + \beta) \max\{r_Q^2, r_m^2\left(\frac{\theta}{16}, \frac{\delta}{2}\right)\}$$



Problem 1.2

without using contraction. : without assuming that the class F consists of uniformly bounded or subgaussian functions, target is in L_p for $p > 2$ under a minor smoothness assumption, l need not to be a Lipschitz function.

★ r_Q is intrinsic parameter of class F , has nothing to do with the choice of the loss or with the target.

r_Q remains unchanged, will see in Section 6.2

★ r_m is external complexity para. will vary for different loss func.

Contraction def: class members in F and target Y are uniformly bounded, loss is Lipschitz on the range of the functions $f(x) - Y$

See examples in section 6

$$r_Q(\xi_1, \xi_2) \equiv \inf \{ r > 0 : E \|G\|_{F_r} \leq C_4 \max(\xi_1, \xi_2) r \sqrt{N} \}$$

Let $u(r, \delta) = C(L)$

• For Square Loss (Satisfy Theo. 5.8)

$$r_M\left(\frac{C_2}{4}, \frac{\delta}{2}\right) \leq \frac{\|\xi\|_{L_4}}{\sqrt{N}} + \inf \{ r > 0 : E \|G\|_{F_r} \leq C_3(L) \|\xi\|_{L_4} \beta^{-1} r^2 \sqrt{N} \}$$

• For Log Loss (Satisfy Theo 5.8)

$$r_M\left(\frac{\theta}{16}, \frac{\delta}{2}\right) \leq \frac{1}{\sqrt{N}} + \inf \{ r > 0 : E \|G\|_{F_r} \leq \exp(-C_1 \|\xi\|_{L_2}) r^2 \sqrt{N} \}$$

• For Huber loss (Strongly convex in $(-\delta, r)$ set $\delta = C(L) \max\{\|\xi\|_{L_2}, r_Q\}$)
(Satisfy Theo 5.8)

$$r_M\left(\frac{\theta}{16}, \frac{\delta}{2}\right) \leq C_2 \left(C_0 \frac{\max\{\|\xi\|_{L_2}, r_Q\}}{\sqrt{N}} + \inf \{ r > 0 : \gamma E \|G\|_{F_r} \leq C_3 r^2 \sqrt{N} \} \right)$$

By showing examples. with specific $U(r, \delta) = c_3(L) \left(1 + \sqrt{\frac{\log(1+\delta)}{n}}\right)$

Huber loss obtains the optimal rate and despite the noise being

heavy-tailed.