Section 3.

lemma 3.1 + **lemma 3.7**

$$P\left(|z_i| \le C_2 \frac{\|z\|_{L_2}}{\sqrt{\varepsilon}}\right) \ge 1 - 2\exp(-C_1 \varepsilon N)$$

$$P\left(|z_i| \ge K\|z\|_{L_2}\right) \ge 1 - 2\exp(-CN\varepsilon)$$

$\downarrow$

Corollary 3.8.

$$P\left(K\|z\|_{L_2} \le |z_i| \le C_2 \frac{\|z\|_{L_2}}{\sqrt{\varepsilon}}\right) \ge 1 - 2\exp(-C_1 \varepsilon N)$$

$\downarrow$ uniform estimate on a class

Theorem 3.11 with $r > r_Q(H, N, \zeta_1, \zeta_2)$

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

$\downarrow$

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

$\downarrow$

Corollary 3.13 $r = 2r_Q(F - F, N, \zeta_1, \zeta_2)$, $\|f_1 - f_2\|_{L_2} \ge r$

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

or $\inf_{\{f \in F : \|f-f^*\|_{L_2} \geq 2r_Q\}} \frac{1}{N}\sum_{i=1}^{N}\left(\frac{f-f^*}{\|f-f^*\|_{L_2}}\right)^2(X_i) \geq \frac{\varepsilon k_0^2}{16}$

$\Big\downarrow$ using the fact that $Q_{f-f^*}(X,Y) \gtrsim l''(\xi)(f-f^*)^2(X)$, where $Q_{f-f^*} = \int_{\xi}^{\xi+(f-f^*)(X)} (l'(w)-l'(x))dx$

$\Rightarrow P_N Q_{f-f^*} \geq \frac{1}{N}\sum_{i=1}^{N} l''(\xi)(f-f^*)^2(X)$

Next to bound the Quadratic term.

Theo 4.1 & Theo 4.2 & Theo 4.6

l is strongly convex | l is general $W \perp X$ | l is strongly convex in a neighborhood of 0 $W \not\perp X$

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

$$P\left(P_N Q_{f-f^*} \geq c_1 \varepsilon k_0^2 \, p(0,t) \|f-f^*\|_{L_2}^2\right) \geq 1 - 2\exp(-c_0\varepsilon^2 N)$$

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

when F is bounded in L_2 : $t = C(\varepsilon, k_0)\left(\|\xi\|_{L_2} + \mathrm{diam}(F, L_2)\right)$

$\|f-f^*\|_{L_2} \leq \mathrm{diam}(F, L_2)$

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

Theo 4.6 + Lemma 5.2 $\longrightarrow$ Theo 5.4 $\{t_1 = 0,\ t_2 = c_0(k_0, \varepsilon)(\|\xi\|_{L_2} + \mathrm{diam}(F, L_2))\}$.

$\|f-f^*\|_{L_2} \geq 2r_Q$ | $\|f-f^*\|_{L_2} \geq 2r_M(\frac{\theta}{16}, \frac{\delta}{2})$ | $\|\hat{f}-f^*\|_{L_2} \leq 2\max\{r_Q, r_M\}$

$P_N Q_{f-f^*} \geq \theta\|f-f^*\|_{L_2}$ | $P_N M_{f-f^*} \leq \frac{\theta}{4}\|f-f^*\|_{L_2}^2$ | $E\mathcal{L}_{\hat{f}} \leq 2(\theta+\beta)\max\{r_Q^2, r_M^2\}$.

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

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

Lemma 5.6 + $\theta = C_1\varepsilon k_0^2 \rho(0,t)$,

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

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

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

Theo. 5.5

$$P(\|\hat{f}-f^*\|_{L_2}\le \max\{\|\xi\|_{L_2}, 2r_Q\}) \ge 1-\delta-2\exp(-C_0N\varepsilon^2)$$

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

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 $[-\gamma, \gamma]$

If $r_M(\frac{\theta}{16}, \frac{\delta}{2}) \le \max\{\|\xi\|_{L_2}, 2 r_Q\}$

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

- $\|\hat{f} - f^*\|_{L_2} \le 2 \max\{r_Q, r_M(\frac{\theta}{16}, \frac{\delta}{2})\}$
- $\mathbb{E}\mathcal{L}_{\hat{f}} \le 2(\theta+\beta) \max\{r_Q^2, r_M^2(\frac{\theta}{16}, \frac{\delta}{2})\}$
- If $W \perp\!\!\!\perp X$ can take $t_1 = C_5 k_1 \|\xi\|_{L_2}$

Assume $\|\xi\|_{L_2} \le C_0 \gamma$. $r_M(\frac{\theta}{16}, \frac{\delta}{2}) \le \gamma$

Prob $1-\delta - 2\exp(-C_4 N \varepsilon^2)$

- $\|\hat{f} - f^*\|_{L_2} \le 2 \max\{r_Q, r_M(\frac{\theta}{16}, \frac{\delta}{2})\}$
- $\mathbb{E}\mathcal{L}_{\hat{f}} \le 2(\theta+\beta) \max\{r_Q^2, r_M^2(\frac{\theta}{16}, \frac{\delta}{2})\}$

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

Solve

Problem 1.2

without using contraction: without assuming that the class F consits 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.

See examples in Section 6

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

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

- For Square Loss (Statisfy Theo. 5.8)

$$r_M\left(\frac{C_2}{4}, \frac{\delta}{2}\right) \le \overset{r_0}{\frac{\|\xi\|_{L_4}}{\sqrt{N}}} + \inf\{ r>0 : E\|G\|_{F_r} \le C_8(L)\|\xi\|_{L_4}\beta^{-1} r^2 \sqrt{N} \}$$

r_M'

- For Log loss (statisfy Theo 5.8)

$$r_M\left(\frac{\theta}{16}, \frac{\delta}{2}\right) \le \frac{1}{\sqrt{N}} + \inf\{ r>0 : E\|G\|_{F_r} \lesssim \exp(-c_1\|\xi\|_{L_2}) r^2 \sqrt{N} \}$$

- For Huber loss (strongly convex in $(-\gamma, \gamma)$ set $\gamma = C_0(L)\max\{\|\xi\|_{L_2}, r_Q\}$)

(Satisfy Theo 5.8)

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

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.