3.14 Algorithm stability

Motivation:

Using uniform convergence: [for $\forall h \in H$]

$$P[L(\hat{h}) - \hat{L}(\hat{h}) \geq \varepsilon] \leq P[\sup_{h \in H} L(h) - \hat{L}(h) \geq \varepsilon]$$

Actually, LHS holds for any estimator.

Question: what if an algorithm A do not use all of H?

Let's define the regularized empirical risk minimizer as:

$$\hat{h} = \arg\min_{h \in \mathcal{H}} \hat{L}(h) + \frac{\lambda}{2}\|h\|_{\mathcal{H}}^2$$

where $\|h\|_{\mathcal{H}}^2$ is $\|w\|_2^2$ of the assosiated weight vector.

and constraint is $\|h\|_{\mathcal{H}} \leq B$. [B depends on data].

Define the training set S and the perturbed version S^i as

(i) $S = (z_1, \cdots, z_n)$: drawn i.i.d. from p^*

(ii) $S^i = (z_1, \cdots, z_i', \cdots, z_n)$: i.i.d. copy of the i-th example.

(iii) z_0 is a new test sample.

[Definition 15] Uniform stability

(i) an algorithm $A: Z^n \to \mathcal{H}$ has uniform stability β w.r.t. a loss function ℓ if for all $S \in Z^n$, $S^i \in Z^n$ and $z_0 \in Z$,

$$|\ell(z_0, A(S)) - \ell(z_0, A(S^i))| \leq \beta$$

Note, this is a strong condition and not riliant on distrub-.

[Example 11] stability of mean estimation

Assume $\forall z \in Z \subseteq \mathbb{R}^d$, $\|z\|_2^2 \leq B$.

Define loss : $\ell(z, h) = \frac{1}{2}\|z - h\|_2^2$

Define algorithm: $A(S) := \arg\min_{h \in \mathbb{R}^d} \hat{L}(h) + \frac{\lambda}{2}\|h\|_2^2$

$$\Rightarrow A(S) = \frac{1}{(1+\lambda)n} \sum_{i=1}^{n} z_i$$

Th... $\beta = \frac{6B^2}{\ldots}$

Then $\beta = \frac{\ }{(1+\lambda)n}$.

To derive this, define $V_1 = A(S) - z_0$, $V_2 = \frac{1}{(1+\lambda)n}[z_i' - z_i]$

$$|\ell(z_0, A(S)) - \ell(z_0, A(S^i))| = \left|\frac{1}{2}\|V_1\|_2^2 - \frac{1}{2}\|V_1+V_2\|_2^2\right|$$

$$\le \|V_2\|_2 \left(\|V_1\|_2 + \frac{1}{2}\|V_2\|_2\right) \quad \text{Minkowski ineq.}$$

$$\le \frac{2B}{(1+\lambda)n}(2B + B) \quad \square$$

[Theorem 16] generalization under uniform stability.

Let A be an algorithm with uniform stability β.

Assume the loss bounded: $\sup_{z,h} |\ell(z,h)| \le M$

Then with prob $\ge 1-\delta$, we have

$$L(A(S)) \le \hat{L}(A(S)) + \beta + (\beta n + M)\sqrt{\frac{2\log(1/\delta)}{n}}$$

Note: we must have that $\beta \sim o(\frac{1}{\sqrt{n}})$.

Pf of theorem 16:

Define $D(S) := L(A(S)) - \hat{L}(A(S))$

① Bound $E[D(S)]$

$$E[D(S)] = E\left[\frac{1}{n}\sum_{i=1}^n [\ell(z_0, A(S)) - \ell(z_i, A(S))]\right]$$

$$= E\left[\frac{1}{n}\sum_{i=1}^n [\ell(z_i', A(S)) - \ell(z_i, A(S))]\right] \quad \text{rename } z_0.$$

$$\le \beta$$

② show $D(S)$ satisfies the bounded differences property.

$$|D(S) - D(S^i)| = |L(A(S)) - \hat{L}(A(S)) - L(A(S^i)) + \hat{L}^i(A(S^i))|$$

$$\le |L(A(S)) - L(A(S^i))| + |\hat{L}(A(S)) - \hat{L}(A(S^i))|$$

$$+ |\hat{L}(A(S)) - \hat{L}^i(A(S^i))|$$

$$\le \beta + \beta + \frac{2M}{n}$$

where $L^i(A(S^i)) = \frac{1}{m}\sum_{k=1}^m L(A(S_k^i))$ $\quad S_k^i = \{(z_1, \cdots, z_i^{(k)}, \cdots z_n) : k\}$

③ McDiarmid's inequality:

$-2\varepsilon^2$

$$P\{D(S)-E[D(S)]\geq \varepsilon\}\leq \exp\left\{\frac{}{n\left(2\beta+\frac{2M}{n}\right)^2}\right\}$$

$$=\exp\left\{\frac{-n\varepsilon^2}{2(\beta n+M)^2}\right\}=:\delta \qquad \square$$