Chapter 10 Stability of Learning Algorithms

$(Z, P, \mathcal{F})$ $\quad f: Z \to \mathbb{R}$ $\quad$ Goal $\underbrace{\min_f L(f)}_{L^*} \triangleq E[f(Z)] = P(f)$

A is <u>consistent</u>: $\quad L(\hat{f}_n) \to L^*$ in some sense

$$\hat{f}_n = A(Z^n)$$

(Note $L(\hat{f}_n) \geq L^*$ w/p one

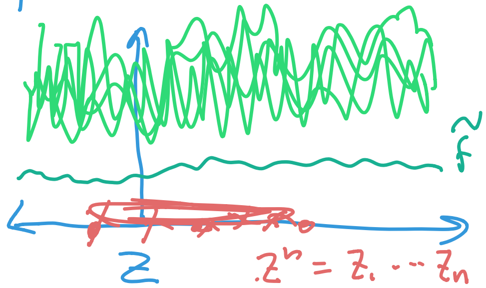
so $E\, L(\hat{f}_n) \xrightarrow{n\to\infty} L^*$ is same as

$$E[\,|L(\hat{f}_n) - L^*|\,] \xrightarrow{n\to\infty} 0$$

ERM $\quad \hat{f}_n = \underset{f \in \mathcal{F}}{\arg\min} \frac{1}{n}\sum_{i=1}^n f(Z_i)$

<u>UCEM</u>: $\quad \sup_P E_P \underbrace{\|P_n - P\|_{\mathcal{F}}}_{\underbrace{\sup_{f\in\mathcal{F}} |P_n(f) - P(f)|}_{\Delta_n(Z^n)}} \to 0$

<u>Note</u> UCEM is <u>not</u> necessary for ERM to be consistent.



$\mathcal{F}$ – so complex that

$\sup_f |P_n(f) - P(f)| \not\to 0$

$\hat{f} \triangleq \underset{f \in \tilde{\mathcal{F}} = \mathcal{F}\cup\{\tilde{f}\}}{\arg\min} \frac{1}{n}\sum_{i=1}^n f(Z_i) = \tilde{f}$ with prob. one.

ERM consistent

10.1 Another abstract framework.

$$(Z, P, \mathcal{F}, \ell) \qquad \ell : \mathcal{F} \times Z \to \mathbb{R}$$

$\ell(f, z)$

$\mathcal{F}$: space of functions on Z

Examples $\quad \ell(f, z) = f(z) \quad$ (reduces to most abstract framework Sect. 6.1)

or $\quad Z = X \times Y \quad \ell(f, (x, y)) = \begin{cases} \phi(-y f(x)) \\ (y - f(x))^2 \end{cases}$

$$\phi(-y \underbrace{\langle f, \psi(x) \rangle}_{\sum_i f_i \psi_i(x)})$$

$$L(f) = \int_Z \ell(f, z)\, dP(z)$$

$$L^*(\mathcal{F}) = \inf_f L(f)$$

$$L_n(f) = \frac{1}{n} \sum_{i=1}^n \ell(f, z_i)$$

$$c_n(A) = \sup_P E_P\left[L(A(z^n)) - \underbrace{L^*}_{L^*_P(\mathcal{F})} \right]$$

Consistency $n \to \infty$ $\to 0$?

$$e_n(A) = \sup_P E_P\left[L_n(A(z^n)) - L_n^* \right]$$

AERM asymptotic $n \to \infty$ $\to 0$?

$$g_n(A) = \sup_P E_P \left| L(A(z^n)) - L_n(A(z^n)) \right|$$

$$\bar{g}_n(A) = \sup_P \left| E_P\left[L(A(z^n)) - L_n(A(z^n)) \right] \right|$$

generalizes $n \to \infty$ $\to 0$

Claim $\quad c_n(A) \le g_n(A) + e_n(A)$

If $f^* = \arg\min L(f)$ so $L^* = L(f)$

then $L_n^* \le L_n(f^*)$ w.p. 1

$$L_n^* = \underset{\text{LERM}}{L_n(\hat{f}_n)}$$

so $EL_n^* \le L^*$

10.2 Suppose $\mathfrak{F}$ is convex subset of H

L, m $f \to \ell(f, z)$ — is an m-strongly convex function and L-Lipschitz continuous

Then the ERM algorithm

$$\hat{f}_n = A(z^n) = \underset{f \in \mathfrak{F}}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} \ell(f, z_i)$$

satisfies $L(\hat{f}_n) - L^* \le \dfrac{2L^2}{\delta m n}$ w/p $\ge 1-\delta$.

Let $z^n = (z_1, \dots, z_n)$ be the training data

$(z_1', \dots z_n')$ is an independent copy of training data.

$$z^n_{(i)} = (z_1, \dots z_{i-1}, z_i', z_{i+1}, \dots z_n) \qquad \text{"}(z_{-i}, z_i')\text{"}$$

$$L_n(f) = \frac{1}{n} \sum_{i=1}^{n} \ell(f, z_i)$$

$$L_n^{(i)}(f) = \frac{1}{n} \ell(f, z_i') + \frac{1}{n} \sum_{j: j \neq i} \ell(f, z_j)$$

$$= \underbrace{L_n(f)}_{\text{m-strongly convex}} + \underbrace{\frac{1}{n}\left[\ell(f, z_i') - \ell(f, z_i)\right]}_{\frac{2L}{n}\text{-Lipschitz continuous}}$$

Lipschitz $|\Phi(f) - \Phi(f')| \leq L\|f - f'\|$

So $\|\hat{f}_n - \hat{f}_n^{(i)}\| \leq \dfrac{2L}{nm}$ b

So $\|\ell(\hat{f}_n, z) - \ell(\hat{f}_n^{(i)}, z)\| \leq L\|\hat{f}_n - \hat{f}_n^{(i)}\| \leq \dfrac{2L^2}{nm}$

$E[L(A(z^n))] - L^* \leq$

(10,11)

$E[L(A(z^n))] - E[L_n(A(z^n))]$

$$= E\left[\frac{1}{n} \sum_{i=1}^{n} \ell(A(z^n), z_i') - \ell(A(z_{(i)}^n), z_i')\right]$$

Has same mean as $\ell(A(z^n), z_i)$

$\leq \dfrac{2L^2}{nm}$

Finish proof using Markov inequality.