→ 10.2 Learnability without uniform convergence
10.3 Learnability and stability
10.4* Stability of SGD

$\ell(f,z)$	$L(f) = E[\ell(f,z)]$	$L_n(f) = E_n[\ell(f,z)]$
$f(z)$		$= \frac{1}{n}\sum_{i=1}^{n} \ell(f,z_i)$
<u>loss</u>	<u>risk</u>	empirical <u>risk</u>

<u>Recall</u> $(Z, \mathcal{F}, P, \ell)$

<u>Thm 10.1</u> Assume $f \to \ell(f,z)$ is m strongly convex
L Lipschitz

$\hat{f}_n$ = ERM output

Then $L(\hat{f}_n) - L^* \leq \frac{2L^2}{\delta m n}$ w/p $\geq 1-\delta$.

(proved that $E[L(\hat{f}_n)] - L^* \leq \frac{2L^2}{mn}$

$$Z_{(i)}^n = (z_1, \cdots z_{i-1}, z_i', z_{i+1}, \cdots z_n)$$

$$|L_n(f) - L_n^{(i)}(f)| \leq \frac{2L^2}{mn}$$ for all f.

$\frac{2L}{n}$ Lipschitz perturbation

$$|\ell(\hat{f}_n, z) - \ell(\hat{f}_n^{(i)}, z)| \le \frac{2L^2}{mn}$$

(10.9) (10.11) $$E[L(\hat{f}_n) - L_n(\hat{f}_n)] = \frac{1}{n}\sum_{i=1}^{n} E[\ell(\hat{f}_n, z_i') - \ell(\hat{f}_n^{(i)}, z_i')]$$

$$E[L_n(\hat{f}_n)] \le L^*$$

Thm 10.2 $\mathcal{F}$ - convex bounded subset of Hilbert space $\mathcal{H}$

$\|f\| \le B$ for $f \in \mathcal{F}$

$f \to \ell(f, z)$ is convex and L-Lipschitz continuous.

$$\hat{f}_{n,\lambda} = A_\lambda(Z^n) := \operatorname*{argmin}_{f \in \mathcal{F}} \frac{1}{n}\sum_{i=1}^{n} \underbrace{\ell(f, z_i) + \frac{\lambda}{2}\|f\|^2}_{\text{new loss - } \lambda \text{ strongly convex}}$$

Then, with $\lambda = \frac{2L}{B\sqrt{n}}$, with prob. $\ge 1-\delta$:

$$L(\hat{f}_{n,\lambda}) \le L^* + \frac{LB}{\sqrt{n}} + \frac{LB}{\delta\sqrt{n}} + \frac{8LB}{\delta n}$$

Proof New ℓ is

$\ell(f, z) + \frac{\lambda}{2}\|f\|^2 \leftarrow L + \lambda B$ Lip. cont.

$\|\lambda f\|$ is $\|\text{grad}\| \le \lambda B$

If ℓ m strongly convex $\frac{2L^2}{mn}$

By proof of Thm. 10.1

$$E[A_\lambda(\hat{f}_{n,\lambda})] \le A_\lambda^*(\mathcal{F}) + \frac{2(L+\lambda B)^2}{\lambda n}$$

$$\le L^* + \frac{\lambda B^2}{2} + \frac{2(L+\lambda B)^2}{\lambda n}$$

$$\left(1+\frac{2}{\sqrt{n}}\right)^2 \le \left(1+\frac{6}{\sqrt{n}}\right)$$

(10.11) $$E[L(\hat{f}_n) - L_n(\hat{f}_n)] = \frac{1}{n}\sum_{i=1}^{n} E[\ell(\hat{f}_n, z_i') - \ell(\hat{f}_n^{(i)}, z_i')]$$

$$\underbrace{\sup_P |\quad\quad\quad\quad|}_{\text{Lemma 10.1} \quad \bar{g}_n(A)} = \sup_P |\quad\quad\quad\quad| = \bar{s}_n(A)$$

so A generalizes on average ($\bar{g}_n(A) \to 0$) iff A is stable on average.

As stated Tuesday

$$c_n(A) \le \bar{g}_n(A) + e_n(A)$$

$e_n(A)$: closeness to minimizing empirical risk

$\therefore$ [stability -or- generalizes on ave] + AERM $\Rightarrow$ consistency

$$c_n(A) = \sup_{\mathcal{P}} E[L(A(Z^n))] - L^*$$

$$\bar{g}_n(A) = \sup_{\mathcal{P}} |E[L(A(Z^n))] - E[L_n(A(Z^n))]|$$

$$E L_n^* < L^*$$

$$e_n = \sup_{P} E[L_n(A(Z^n)) - L_n^*]$$

$$L(A(Z^n)) - L^* = \underbrace{L(A(Z^n)) - L_n(A(Z^n))} + L_n(A(Z^n)) - L_n^* + \cancel{L_n^* - L^*}$$

$$c_n(A) \leq \bar{g}_n(A) + e_n(A)$$

$$\bar{g}_n(A) = s_n(A)$$

10.4 Stability of SGD : stochastic gradient descent

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

GD: $f^{(0)}$ $\qquad f^{(t+1)} = f^{(t)} - \alpha \nabla L_n(f^{(t)})$

$$= f^{(t)} - \alpha \frac{1}{n} \sum_{i=1}^{n} \nabla \ell(f, z_i)$$

requires n evaluations of gradient per time step.

Let I_t be uniformly distributed $\{1, \dots n\}$.

SGD
$$f^{(t+1)} = f^{(t)} - \alpha \nabla \ell(f, z_{I_t})$$

Note
$$\mathbb{E}_{I_t}[\nabla \ell(f^t, z_{I_t})] = \frac{1}{n}\sum_{i=1}^{n} \nabla \ell(f, z_i) = \nabla L_n(f^{(t)})$$

---

on the stability of SGD

suppose for some i that z_i is replaced by z_i'

Let φ be some function,

Let $G_\varphi(f) = f - \alpha \nabla \varphi(f)$ = gradient descent mapping

---

nonexpansive property:

$$G_\varphi(f) - G_\varphi(f') \leq \|f - f'\|$$

contraction property

$$G_\varphi(f) - G_\varphi(f') \leq (1-\eta)\|f - f'\|$$

boundedness

$$|G_\varphi(f)| \leq C$$

Theorem 10.4 Suppose $f \to \ell(f,z)$ is convex, M-smooth, L-Lipschitz

If SGD is run with $\alpha_t \leq \frac{2}{M}$ then

for two data sets differing in one sample:

$$\sup_{z \in Z} E[\ell(f_T, z) - \ell(f_T', z)] \leq \frac{2L^2}{n} \sum_{t=1}^{T} \alpha_t$$

Thm 10.5 Same assumptions + $f \to \ell(f,z)$ is m-strongly convex

$\alpha \leq \frac{2}{m+M}$ then

$$\sup_{z \in Z} E|\ell(f_T, z) - \ell(f_T', z)| \leq \frac{4L^2}{mn}$$

(for all T)

Thm 10.6 $\alpha_t \leq c/t$
