Theorem 9.1 With probability $1-\delta$, the ERM algorithm outputs $\hat{f}_n$ satisfying

$$L(\hat{f}_n) \le L^*(\mathcal{F}_\lambda) + \frac{16(M+C_k\lambda)^2}{\sqrt{n}} + (M^2+C_k^2\lambda^2)\cdot\sqrt{\frac{8\log(1/\delta)}{n}}$$

$$\lambda = \frac{M}{\sqrt{\gamma}}$$

Section 9.2 Penalty regularized least squares

$\mathcal{X}$ $\quad K(x,x')$

P - dist^{n} on $\mathcal{X}\times\mathbb{R}$

$$L(f) = E[(Y-f(X))^2]$$

$$J_\gamma(f) := L(f) + \gamma\|f\|_K^2$$

$$\hat{f}_{n,\gamma} = \arg\min_{f\in\mathcal{H}_K} L_n(f) + \gamma\|f\|_K^2$$

$$A(\gamma) = J_\gamma^*(\mathcal{H}_K) - L^*$$

$L^* = E[(Y-E[Y|X])^2]$ — best for unconstrained estimators

Theorem 9.2 For any $\delta\in(0,1)$ with probability at least $1-\delta$: ERM alg satisfies:

$$L(\hat{f}_{n,\gamma}) - L^* \le A(\gamma) + \frac{16M^2\left(1+\frac{C_K}{\sqrt{\gamma}}\right)^2}{\sqrt{n}} + M^2\left(1+\frac{C_k^2}{\gamma}\right)\sqrt{\frac{8\log 1/\delta}{n}}$$

$\min L_n(\cdot)$ $\qquad \min_{f\in\mathcal{F}_\lambda} L(f) + \lambda\|f\|_K^2$

Recall $\quad E R_n(\mathcal{F}_\lambda(X^n)) \leq \lambda \sqrt{\frac{c_K}{n}}$

$\mathcal{F}_\lambda = \{f \in \mathcal{H}_K : \|f\|_K \leq \lambda\}$ $\quad l_f(z) = (Y - f(x))^2$

Note If $f \equiv 0$ then $J_\gamma(f) = L(f) + \gamma \|f\|_K^2 \leq M^2$

$L(f) = E[(Y-0)^2]$, $\gamma\|f\|_K^2 = 0$

$Y \in [-M, M]$

If $\|f'\|_K > \frac{M}{\sqrt{\gamma}}$

then $J_\gamma(f') = L(f') + \underbrace{\gamma \|f'\|_K^2} > M^2$

$\therefore$ If $\|f'\|_K > \frac{M}{\sqrt{\gamma}}$, f' is worse than 0 classifier

define to be λ.

$\lambda = \frac{M}{\sqrt{\gamma}}$

$$\therefore \min_{f \in \mathcal{H}_K} J_\gamma(f) = \min_{f \in \mathcal{F}_\lambda} J_\gamma(f)$$

$$J_{\gamma,n}(f') = L_n(f') + \gamma \|f'\|_K^2$$

minimize $L_n(f) + \gamma \|f\|_K^2$ instead of

$L(f) + \gamma \|f\|_K^2$

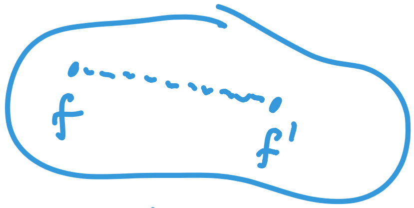
difference is $L_n(f) - L(f)$

Same $\Delta_n(Z^n)$ as before.

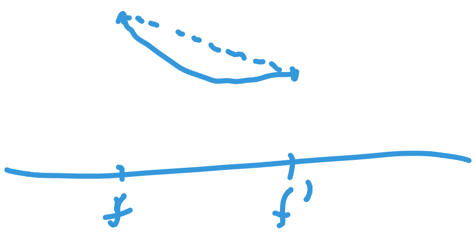
Sections 3.4 + 3.5

m-strongly convex functions + m-smooth convex functions

set is convex:



f convex $\Rightarrow$ $\varphi(\lambda f + (1-\lambda) f') \le \lambda \varphi(f) + (1-\lambda)\varphi(f')$



$\mathcal{F}$ = domain

Definition φ is strongly convex if for some $m > 0$,

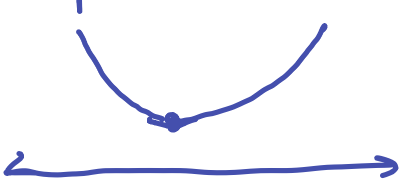
$$\varphi(f') \ge \varphi(f) + \langle g, f' - f \rangle + \frac{m}{2}\|f - f'\|^2$$

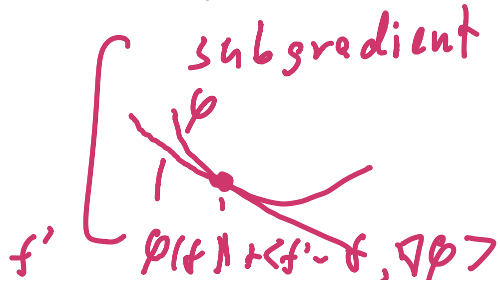
for all $f, f' \in \mathcal{F}$

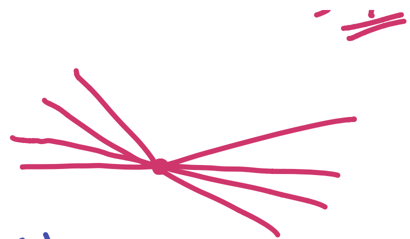
$g = \nabla \varphi(f)$

for any subgradient of φ at x.

upward curvature:



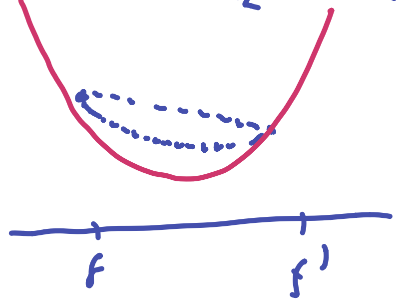




Lemma 3.2

(1) For any f, f' and $0 \le \lambda \le 1$

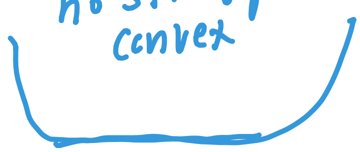
$$\varphi(\lambda f + (1-\lambda) f') \le \lambda \varphi(f) + (1-\lambda)\varphi(f') - \frac{\lambda(1-\lambda)m}{2}\|f-f'\|^2$$



(0) φ is m-strongly convex $\Longleftrightarrow$ $\varphi(f) - \frac{m}{2}\|f\|^2$ is convex.

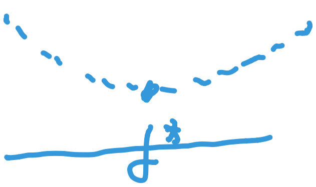
(2) If φ is strongly convex it has a <u>unique</u> minimizer.

no strongly convex



(3) If f^* is the minimizer then for any f

$$\varphi(f) - \varphi(f^*) \ge \frac{m}{2}\|f-f^*\|^2$$



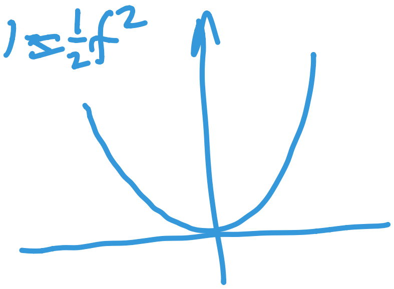
(4) Stability of minimizers under Lipschitz perturbations

$\|f\|^2$

1-strongly convex

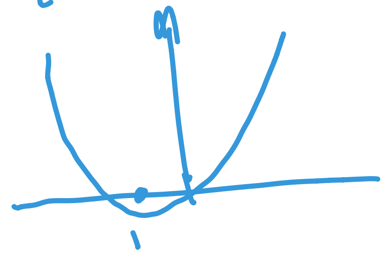
e.g.

$\varphi(f) = \frac{1}{2}f^2$



$f^* = 0$

$\varphi(f) + \alpha f = \frac{1}{2}f^2 + \alpha f$



$-\alpha/2$

$f + \alpha = 0$

$f^*_{new} = -\alpha$

$B(f) = \alpha f$

Suppose B is L-Lipschitz

$(\|B(f) - B(f')\| \leq L\|f - f'\|$

and suppose $\tilde{f}$ is a minimizer of $\varphi + B$

Then $\|f^* - \tilde{f}^*\| \leq \frac{L}{m}$

f^* —— $\tilde{f}^*$

Section 3.5 smooth convex functions

$\mathcal{F}$ - closed subset of $\mathcal{H}$

$\varphi : \mathcal{F} \to \mathbb{R}$ is M-smooth for some $M \geq 0$

if $f \to \nabla\varphi$ is M-Lipschitz:

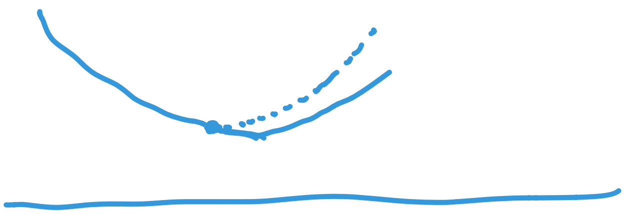
$$\|\nabla\varphi(f) - \nabla\varphi(f')\| \leq M\|f - f'\|$$

[If φ is twice continuously differentiable then φ is M smooth if $\nabla^2\varphi(f) \preceq M \cdot I$ $\mathcal{H} = \mathbb{R}^d$]

(i.e. $\alpha^T[\nabla^2\varphi(f) - MI]\alpha \geq 0$ for all α.)

1d case
$\nabla\varphi = \varphi'$

If φ'' exists, M smoothness is equivalent to $\varphi'' \leq M$



[$\frac{M}{2}\|f\|^2$ - is M-smooth and M strongly convex]

$\varphi(f) = f' A f$ $\quad$ A - symmetric matrix

eigen values $\lambda_1, \dots \lambda_d$

$\underline{M}$ smoothness $\quad \lambda_i \le \underline{M}$ all i

m strongly convex $\quad \lambda_i \ge m$ all i

Gradient descent

$$x^{(k+1)} = x^{(k)} - \alpha \nabla \varphi(x^{(k)})$$