(b) Stability in a formal manner.

$$F(w,S)=\sum_{i\in S}\Big[\underbrace{\lambda\|w\|^2}_{\lambda>0}+\ell(\underbrace{w,z_i}_{(x_i,y_i)})\Big] \to (1-y_i w^T x_i)_+ ,\quad (y_i-w^T x_i)^2$$

$\cancel{w^*(S\cup K)}$

$$\mathrm{eig}\left[\frac{\partial^2 \ell}{\partial w^2}\right]\geq 0$$

$$w^*(S)=\arg\min_w F(w,S) \Rightarrow F(w^*(S),S)=\min_w F(w,S)$$

$$\Rightarrow \|w^*(S\cup K)-w^*(S)\|=O\left(\frac{1}{\lambda T}\right)$$

$$F(w,S)=\sum_{i\in S}\Big[\underbrace{\lambda\|w\|^2}_{\lambda>0}+\ell(\underbrace{w,z_i}_{(x_i,y_i)})\Big]$$

assumption (1) ℓ is convex $\Rightarrow \mathrm{eig}\left[\frac{\partial^2 \ell}{\partial w^2}\right]\geq 0 \ \forall$

(2) $\left\|\boxed{\frac{d\ell}{dw}}\right\| = |\ \cdot\ | < B$ $(B>0)$

(3) $0<\Lambda_{min}<\mathrm{eig}\left[\frac{\partial^2 \ell}{\partial w^2}\right]<\Lambda_{max} \ \forall$

$\Lambda_{min}>0$

$\min_w F(w,S) \to w^*(S)$

$\min_w F(w,S\cup K) \to w^*(S\cup K)$

$$F(w^*(S\cup K),S)-F(\overset{\min}{w^*(S)},S)\geq 0$$

$$F(w^*(S\cup K),S)=F(w^*(S),S)+\cancel{\frac{\partial F}{\partial w}\Big|_{w^*(S)}}^{\,T}\ (w^*(S\cup K)-w^*(S))$$

0

$$+\frac{1}{2}\left(w^*(S\cup k)-w^*(S)\right)^T \left.\frac{\partial^2 F}{\partial w^2}\right|_{w\in(w^*(S),\,w^*(S\cup k))} \left(w^*(S\cup k)-w^*(S)\right)$$

$$F(w,S)=\sum_{i\in S}\left[\lambda\|w\|^2+\ell(w,z_i)\right] \geq \Lambda_{min}=0$$

$$\frac{\partial^2 F(w,S)}{\partial w^2}=\sum_{i\in S}\left[2\lambda + \boxed{\frac{\partial^2 \ell}{\partial w^2}(w,z_i)}\right]$$

$$=2\lambda|S|+0$$

$$=2\lambda|S|+\Lambda_{min}|S|$$

$$=2\lambda|S|$$

$$\frac{1}{2}\left(w^*(S\cup k)-w^*(S)\right)\,2\lambda|S|\,\left(w^*(S\cup k)-w^*(S)\right)$$

$$=\lambda|S|\,\|w^*(S\cup k)-w^*(S)\|^2$$

$$F(w^*(S\cup k),S)-F(w^*(S),S)\geq\lambda|S|\,\|w^*(S\cup k)-w^*(S)\|^2$$

$$F(w^*(S\cup k),S)-F(w^*(S),S)\leq ?$$

$$=\boxed{F(w^*(S\cup k),S\cup k)-F(w^*(S),S\cup k)\leq 0}$$

$$+F(w^*(S),k)-F(w^*(S\cup k),k)\leq ?$$

$$\leq F(w^*(S),k)-F(w^*(S\cup k),k)$$

$$f(w_1)-f(w_2)\leq B\,\cancel{\|w_1-w_2\|_1} \to B\sqrt{d}\,\|w_1-w_2\|_2 \qquad \left|\frac{\partial F}{\partial w}\right|<B$$

$$\boxed{F(w,k)=\lambda\|w\|^2+\ell(w,z_k)}$$

$$+\lambda w_m$$

$$|a^T b| < \|a\| \cdot \|b\|$$

$$|a^T b| < a_{max} \left[\sum_{i=1}^{N} |b_i|\right]$$

$$= a_{max} \|b\|_1$$

$$f(w_1) - f(w_2) = \frac{\partial l}{\partial w}\Big|_{w \in (w_1, w_2)}^T (w_1 - w_2)$$

$$\le B \|w_1 - w_2\|_1$$

$$\left|\frac{\partial l}{\partial w}\right| < B$$

$$\Rightarrow \left\|\frac{\partial l}{\partial w}\right\| < B\sqrt{d}$$

$$\sqrt{B^2 + \dots}$$

$$B_d$$

$$F(w^*(S), k) - F(w^*(S \cup k), k)$$

$$\le (2\lambda w_{max} + B\sqrt{d}) \|w^*(S) - w^*(S \cup k)\|$$

$$F(w^*(S \cup K), S) - F(w^*(S), S) \le (2\lambda w_{max} + B\sqrt{d}) \|w^*(S) - w^*(S \cup k)\|$$

$$\partial \quad \ge \lambda \underbrace{\|w^*(S) - w^*(S \cup k)\|^2}_{\Delta w} |S|$$

$$\lambda \|\Delta w\|^2 |S| \le (2\lambda w_{max} + B\sqrt{d}),$$

$$\Rightarrow \|\Delta w\| \le \left[2\lambda w_{max} + B\sqrt{d}\right] / \lambda |S|$$

$$F(w^*(S \cup K), S \cup K) - F(w^*(S), S) \le ?$$

$$= F(w^*(S \cup k), S) - F(w^*(S), S)$$

$$+ \underbrace{F(w^*(S \cup K), k)}_{< F_0}$$

$$\leq \underline{\left\| \frac{\partial F}{\partial w} \right\|} \, \| w^*(S \cup K) - w^*(S) \| + F_0$$

$$= \left(\underbrace{2\lambda |S| \, w_{max}}_{\sum_{i \in S} \frac{\partial}{\partial w} \lambda \|w\|^2} + \underbrace{|S| \, B\sqrt{d}}_{\sum_{i \in S} \frac{\partial}{\partial w} \ell(w, z_i)} \right) \; \text{upper bound} \quad \|\Delta w\| \quad + F_0$$

upperbound

$$= \left(2\lambda |S| \, w_{max} + |S| \, B\sqrt{d} \right) \cdot \frac{B\sqrt{d}}{\lambda |S|}$$

$$= \frac{\left(2\lambda \, w_{max} + B\sqrt{d} \right) B\sqrt{d}}{\lambda} + F_0$$

$$F(w, S) = \sum_{i \in S} \lambda \|w\|^2 + \ell(w, z_i)$$

$\Rightarrow$ $F(w^*(S), S)$ is not ~~grand~~ growing with S.

---

min over $S \cup K$

$$\underline{F(w^*(S \cup K), S \cup K)} - F(w^*(S), S)$$

$$\leq \underline{F(w^*(S), S \cup K)} - F(w^*(S), S)$$

$$= F(w^*(S), K) \leq F_0$$

Generalization

* Chapter 13. Understanding Machine learning
Shai - Ben - David

Stability + Generalization

$$\lambda\|w\|^2|S| \leq \sum_{i \in S} \lambda\|w^*\|^2 + \ell(w^*, z_i) \leq \cancel{|S|\,\ell(0)}$$

$$|S|\,\ell_{max}$$

$$\|w\|^2 \leq \frac{\ell_{max}}{\lambda},$$

$$f(x) = f(0) + f'(0)\,x + \frac{1}{2}x^2 f''(c)$$

$$c \in (0, x)$$