1) Is f Lipschitz continues? If so, find a small B

$$\| f(w_2) - f(w_1) \| \leq B \| w_2 - w_1 \|$$

$$\| \nabla f(w) \|_2 \leq B$$

$$f(w) = \frac{1}{N} \sum_{i \in [N]} f_i(w) + \lambda \| w \|_2^2$$

where $f_i(w) = \log(1 + \exp\{-y_i w^T x_i\})$

$$\nabla f_i(w) = \frac{\exp\{-y w^T x_i\}}{1 + \exp\{-y w^T x_i\}} = \frac{-y^T x}{\exp\{-y w^T x_i\} + 1}$$

Side note

$$g(x) = \| f(x) \|_2 = \sqrt{\sum_{i=1}^{n} f_i(x)^2}$$

$$g(x)^2 = \| f(x) \|_2^2 = \sum_{i=1}^{n} f(x)^2 \Rightarrow$$

$$\sum_{i=1}^{n} \nabla f(x)^2 = \sum_{i=1}^{n} 2 f(x) \nabla f_i(x) =$$

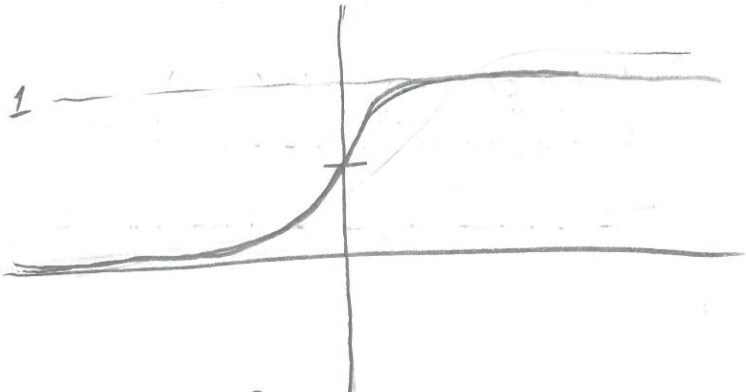
$$\Rightarrow \sum_{i=1}^{n} 2w$$

$$\nabla f(w) = \frac{1}{N} \sum_{i \in [N]} \frac{-y^T x}{\exp\{-y w^T x_i\} + 1} + \lambda \cdot 2 \cdot w$$

$$\frac{|f(x_1) - f(x_2)|}{x_1 - x_2} \leq K$$

Lipschitz continuous

$$\frac{1}{\exp\{-yw^T x_i\}+1} \longrightarrow 1$$



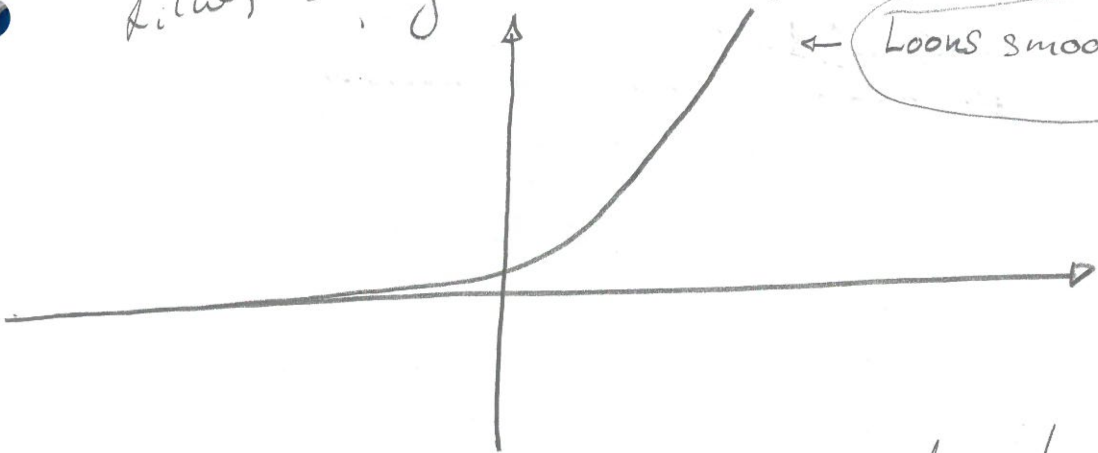
$$\lim_{w \to \infty} f_i(w) = 1$$

Thus $B \geqslant \dfrac{1}{\exp\{-y w_{max}^T x_i\}+1} + 2\lambda w_{max}$

2) Is f_i smooth?

$$f_i(w) = \log\left(1+\exp\{-y_i w^T x_i\}\right)$$

Looks smooth



$\dfrac{df_i(w)}{dw} \Rightarrow$ Sigmoid looking function, thus smooth.

If so, find a small L for f_i?

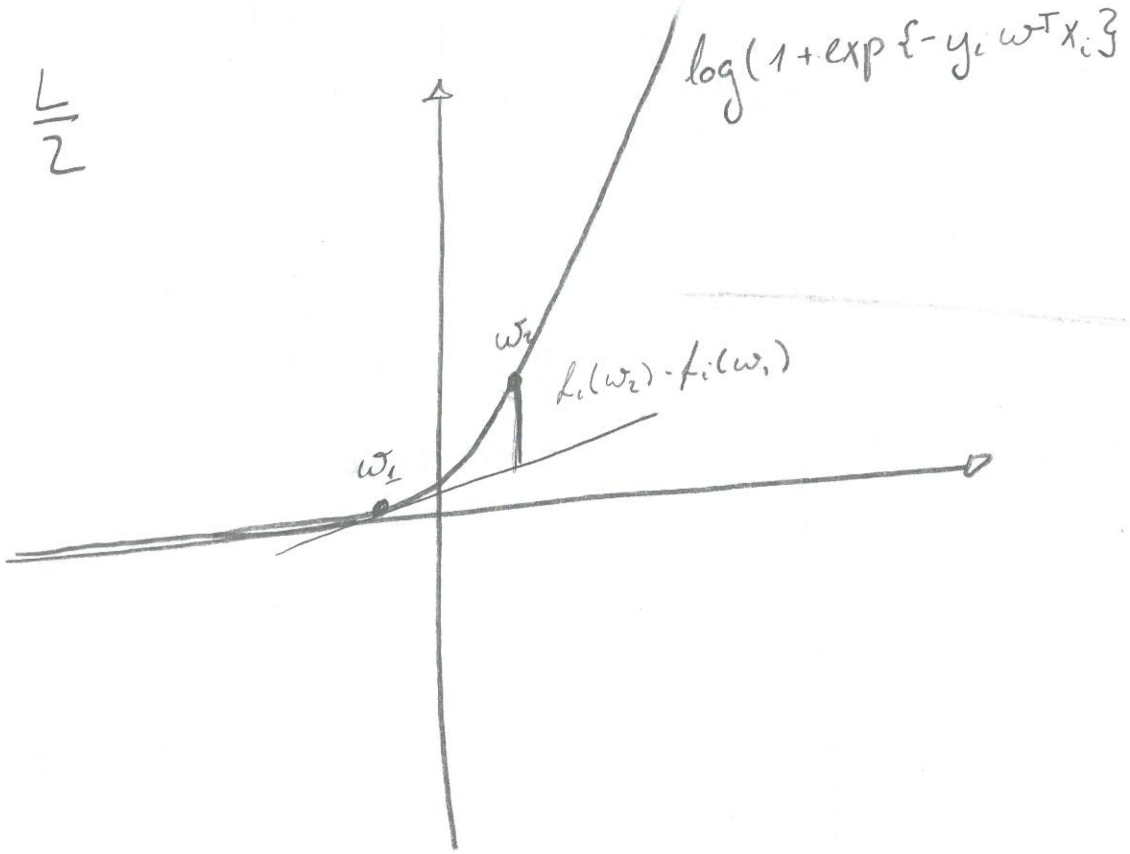
$$f(w_2) \leq f(w_1) + \nabla f(w_1)^T(w_2-w_1) + \frac{L}{2}\|w_2-w_1\|_2^2$$

$$\log\left(1+\exp\{-y_i w_1^T x_i^T\}\right) \leq \log\left(1+\exp\{-y_i w_2 x_i^T\}\right) +$$

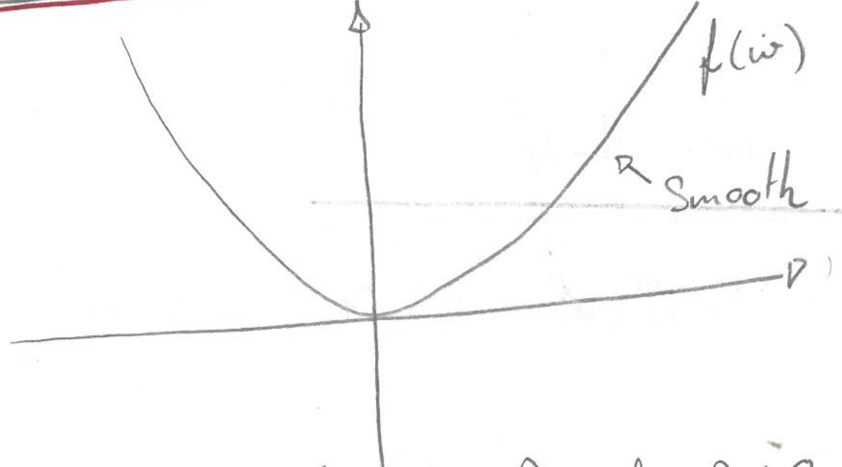
$$\frac{1}{1+\exp\{-y_i w_1 x_i\}}(w_2-w_1) + \frac{L}{2}\|w_2-w_1\|_2^2$$

$$\frac{\log\left(\frac{1+\exp\{-y_i w_1 x_i\}}{1+\exp\{-y_i w_2 x_i\}}\right) - \frac{1}{1+\exp\{-y_i w_1 x_i\}}(w_2 - w_1)}{\|w_2 - w_1\|_2^2} \leq$$

$$\leq \frac{L}{2}$$



<u>What about f?</u>



But depends on λ. If $\lambda < 0$, then concave.

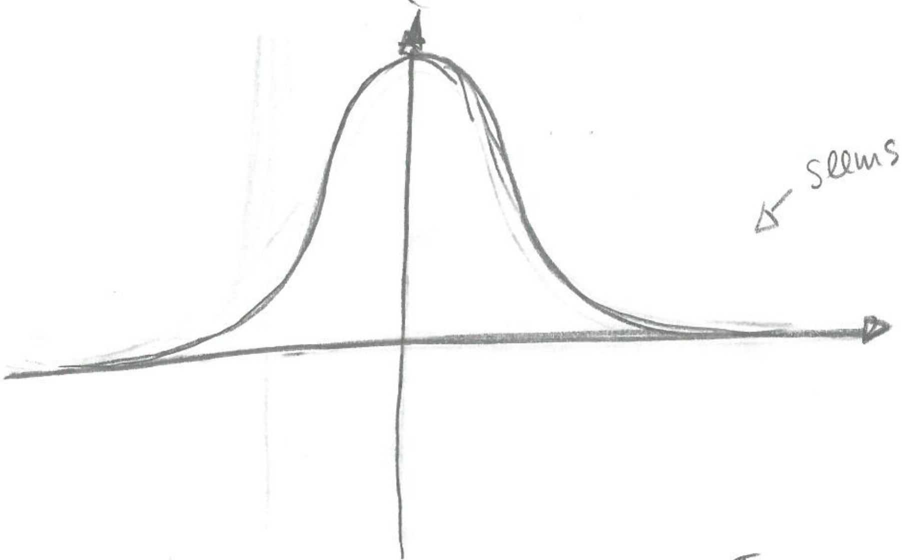
3) Is f strongly convex? If so, find a high μ?

$$f(w_2) \geq f(w_1) + \nabla f(w_1)^T (w_2 - w_1) + \frac{\mu}{2} \|w_2 - w_1\|_2^2$$

$$\nabla^2 f(x) \geq \mu I_d$$

$$\nabla f_i(w) = \frac{1}{\exp\{-y w^T x_i\} + 1}$$

$$\nabla^2 f_i(w) = \frac{\exp\{-y w^T x_i\}}{(1 + \exp\{-y w^T x_i\})^2} \geq \mu I_d$$



seems like it

$$f(w_2) \geq f(w_1) + \nabla f(w_1)^T (w_2 - w_1) + \frac{\mu}{2} \|w_2 - w_1\|_2^2$$

$$f(w) = \frac{1}{N} \sum_{i \in N} \log(1 + \exp\{-y_i w^T x_i\}) + \lambda \|w\|_2^2$$

$$\frac{1}{N} \sum_{i \in N} \log(1 + \exp\{-y_i w_2 x_i^T\}) + \lambda w_2^2 \geq \frac{1}{N} \sum_{i \in N} \log(1 + \exp\{-y_i w_1 x_i^T\}) + \lambda w_1^2$$

$$+ \frac{1}{\exp\{-y_i w_1 x_i^T\} + 1} (w_2 - w_1) + \frac{\mu}{2} \|w_2 - w_1\|_2^2 + 2\lambda w_1$$

$$\frac{1}{N}\sum_{i\in N}\log\left(\frac{1+\exp\{-y_i w_2 x_i^T\}}{1+\exp\{-y_i w_1 x_i^T}\right)+\lambda\left(w_2^2-w_1^2\right)-\frac{1}{\exp\{-y_i w_1 x_i^T\}+1}(w_2-w_1)$$

$$-2\lambda w_1 \geqslant \frac{\mu}{2}\|w_2-w_1\|^2$$

$$\frac{\frac{1}{N}\sum_{i\in N}\log\left(\frac{1+\exp\{-y_i w_2 x_i^T\}}{1+\exp\{-y_i w_1 x_i^T\}}\right)+\lambda\left(\cancel{\ell}w_2^2-w_1^2-2w_1\right)-\frac{1}{\exp\{-y_i w_1 x_i^T\}+1}(w_2-w_1)}{\|w_2-w_1\|^2} \geqslant$$

$$\geqslant \frac{\mu}{2}$$

HW 2.b. We know $\text{var}[g(w_k, \xi_k)] = E[\|g(w_k, \xi_k)\|_2^2] - \|E[g(w_k, \xi_k)]\|_2^2$

$$\|E[g(w_k, \xi_k)]\|_2 \leq c_0 \|\nabla f(w_k)\|_2$$

$$\text{var}[g(w_k, \xi_k)] \leq M + M_V \cdot \|\nabla f(w_k)\|_2^2$$

$$E[\|g(w_k, \xi_k)\|_2^2] = \text{var}[g(w_k, \xi_k)] + \|E[g(w_k, \xi_k)]\|_2^2$$

$$\leq M + M_V \cdot \|\nabla f(w_k)\|_2^2 + c_0^2 \cdot \|\nabla f(w_k)\|_2^2$$

$$= M + (M_V + c_0^2)\|\nabla f(w_k)\|_2^2$$

HW 2.c. "With square summable but not summable step-size" means

$$\sum_{k=1}^{\infty} \alpha_k = \infty \quad \text{and} \quad \sum_{k=1}^{\infty} \alpha_k^2 < \infty$$

The second condition ensures that $\alpha_k \to 0$

and from Eq. 5 $\quad 0 < \alpha_k \leq \frac{c}{L \cdot M_G} \quad \Rightarrow \quad \alpha_k \cdot L \cdot M_G \leq c$

From Eq. 4, we have

$$E[f(w_{k+1})] - f(w_k) \leq -c \cdot \alpha_k \|\nabla f(w_k)\|_2^2 + \frac{1}{2}\alpha_k^2 \cdot L \cdot E[\|g(w_k, \xi_k)\|_2^2]$$

$$\leq -\left(c - \frac{1}{2}\alpha_k \cdot L \cdot M_G\right)\alpha_k \cdot \|\nabla f(w_k)\|_2^2 + \frac{1}{2}\alpha_k^2 \cdot L \cdot M$$

Taking the expectation

$$E[f(w_{k+1})] - E[f(w_k)] \leq -\left(c - \frac{1}{2}\alpha_k \cdot L \cdot M_G\right) \cdot \alpha_k \cdot E[\|\nabla f(w_k)\|_2^2] + \frac{1}{2}\alpha_k^2 \cdot L M$$

- since we know that $\alpha_k \cdot L \cdot M_G \leq c$

$$\leq -\frac{1}{2} c \cdot \alpha_k E[\|\nabla f(w_k)\|_2^2] + \frac{1}{2} \cdot \alpha_k^2 \cdot L M.$$

Note that $k \in 1, 2, \cdots, K$,

Summing the inequality for all k's

$$E[f(w_{K+1})] - E[f(w_1)] \leq -\frac{1}{2} c \sum_{k=1}^{K} \alpha_k \, E[\|\nabla f(w_k)\|_2^2] + \frac{1}{2} L.M. \sum_{k=1}^{K} \alpha_k^2$$

Re-arranging the terms

$$\sum_{k=1}^{K} \alpha_k . E[\|\nabla f(w_k)\|_2^2] \leq \frac{2}{c}\left(E[f(w_1)] - E[f(w_{K+1})]\right) + \frac{LM}{c} \sum_{k=1}^{K} \alpha_k^2$$

From the proof of Theorem 5, we have

$$f_{inf} - f(w_1) \leq E[f(w_{K+1})] - f(w_1)$$

Therefore

$$\sum_{k=1}^{K} \alpha_k \, E[\|\nabla f(w_k)\|_2^2] \leq \frac{2}{c}\left(E[f(w_1)] - f_{inf}\right) + \frac{LM}{c} \sum_{k=1}^{K} \alpha_k^2$$

We know $\sum_k \alpha_k^2 < \infty \Rightarrow \sum_k \alpha_k \, E[\|\nabla f(w_k)\|_2^2] < \infty$

$$\Rightarrow E\left[\sum \alpha_k . \|\nabla f(w_k)\|_2^2\right] < \infty$$

then

$$E\left[\frac{1}{\sum_k \alpha_k} \sum_k \alpha_k . \|\nabla f(w_k)\|_2^2\right] \xrightarrow{K \to \infty} 0$$

since $\sum_k^{\infty} \alpha_k = \infty$ and $\sum_k \alpha_k \|\nabla f(w_k)\|_2^2 < \infty$.