Problem 2.1 b)

Q: Is $f_i(w)$ smooth? If so, find a small L. What about f?

A: $f_i(w) = \log\left(1 + e^{-y_i x_i^T w}\right)$.

- Define $g_i(w) = -y_i x_i^T w = -b_i^T w$, $b_i = y_i x_i$.

$\rightarrow f_i(w) = \log\left(1 + e^{g_i(w)}\right)$.

- A twice differentiable function f is L-smooth iff $\nabla^2 f(x) \preceq LI$.

- We will find $\nabla^2 f_i(w)$ and determine L:

$$\nabla f_i(w) = \frac{1}{1+e^{g_i(w)}} \nabla\left(1 + e^{g_i(w)}\right) = \frac{1}{1+e^{g_i(w)}} e^{g_i(w)} \nabla g_i(w)$$

$$= -b_i \frac{e^{g(w)}}{1+e^{g(w)}}.$$

$$\nabla^2 f_i(w) = -b_i \left(\nabla \frac{e^{g_i(w)}}{1+e^{g_i(w)}}\right)^T = -b_i \left(\frac{\left(\nabla\left(e^{g_i(w)}\right)\right)\left(1+e^{g_i(w)}\right) - e^{g(w)} \nabla\left(1+e^{g_i(w)}\right)}{\left(1+e^{g_i(w)}\right)^2}\right)^T$$

$$= -b_i \left(\frac{e^{g_i(w)}(-b_i)\left(1+e^{g_i(w)}\right) - e^{g_i(w)} e^{g_i(w)}(-b_i)}{\left(1+e^{g_i(w)}\right)^2}\right)^T =$$

$$= b_i b_i^T \frac{e^{g_i(w)} \cancel{+ e^{2g_i(w)}} \cancel{- e^{2g_i(w)}}}{\left(1+e^{g_i(w)}\right)^2}$$

$$= b_i b_i^T \frac{e^{g_i(w)}}{\left(1+e^{g_i(w)}\right)^2}.$$

Problem 2.1 b) cont'd

We have that $\nabla^2 f_i(w) = b_i b_i^T \dfrac{e^{g_i(w)}}{(1+e^{g_i(w)})^2}$,

which is bounded for bounded $b_i = y_i x_i$,

because $\dfrac{e^{g_i(w)}}{(1+e^{g_i(w)})}$ is lower bounded by zero, and upper bounded by $1/4$.

Now we find a small L: $\nabla^2 f_i(w)$ is symmetric,

$\rightarrow \nabla^2 f_i(w) \preceq LI \leftrightarrow x^T(LI - \nabla^2 f_i(w))x \geq 0$, $\forall x, w$

$\leftrightarrow Lx^Tx - x^T b_i b_i^T x \dfrac{e^{g_i(w)}}{(1+e^{g_i(w)})^2} \geq 0$

$\rightarrow Lx^Tx \geq x^T b_i b_i^T x \dfrac{e^{g_i(w)}}{(1+e^{g_i(w)})^2}$ $\quad \forall x, w$

$\rightarrow L \geq \dfrac{x^T b_i b_i^T x}{x^T x} \dfrac{e^{g_i(w)}}{(1+e^{g_i(w)})^2}$, $\forall x, w$, $x \neq 0$

in particular, we need $L \geq \max\limits_{x \neq 0} \dfrac{x^T b_i b_i^T x}{x^T x} \dfrac{e^{g_i(w)}}{(1+e^{g_i(w)})^2}$, $\forall w$

$$= \lambda_{max}(b_i b_i^T) \frac{e^{g_i(w)}}{(1+e^{g_i(w)})^2}, \quad \forall w$$

where $\lambda_{max}(b_i b_i^T) = \|b_i\|^2 = b_i^T b_i \geq 0$

$\rightarrow L \geq b_i^T b_i \dfrac{e^{g_i(w)}}{(1+e^{g_i(w)})^2}$ $\quad \forall w$

where the right hand side is upper bounded by $\frac{1}{4} b_i^T b_i$

$\rightarrow$ Smallest L is $L = \frac{1}{4} b_i^T b_i = \frac{1}{4} y_i^2 x_i^T x_i$.

Answer: $f_i(w)$ is L-smooth with $L = \frac{1}{4} y_i^2 x_i^T x_i$.

"What about f ?"

The Hession of f is $\nabla^2 f(w) = \frac{1}{N} \sum_{i \in [N]} \nabla^2 f_i(w) + 2\lambda$

which is bounded since $[N]$ is finite, and f_i's are bounded.

~~We~~ $\rightarrow$ f is L-smooth, and we can find L as :

$$L x^T x - \frac{1}{N} x^T \sum_i (\nabla^2 f_i(w)) x + 2\lambda x^T x \geq 0 \qquad \forall x, w$$

$$\Leftrightarrow L \geq \frac{1}{N} \frac{x^T (\sum_i \nabla^2 f_i(w)) x}{x^T x} - 2\lambda \qquad \forall x \neq 0, w$$

to ensure for $\forall x \neq 0, w$ we choose L as:

$$L = \left(\sum_{i \in [N]} \max_{x_i \neq 0} \frac{1}{4N} \frac{x_i^T b_i b_i^T x_i}{x_i^T x_i} \right) - 2\lambda$$

$$= \frac{1}{4N} \sum_{i \in [N]} b_i^T b_i - 2\lambda$$

$$= \frac{1}{4N} \sum_{i \in [N]} y_i^2 x_i^T x_i$$

Answer: $f(w)$ is L-smooth with L

$$L = \frac{1}{4N} \sum_{i \in [N]} y_i^2 x_i^T x_i - 2\lambda .$$