COMP5212 Machine Learning
Written Homework 1

LO, Li-yu
20997405
e-mail: lloac@connect.hkust.hk

25/Sep/2023

Problem 1

(a)

△ Get $\frac{\partial g}{\partial z}$,

$$\begin{aligned}
\frac{\partial}{\partial z}(1+e^{-z})^{-1} &= (-1)\cdot(1+e^{-z})^{-2}\cdot(1+e^{-z})' \\
&= (1+e^{-z})^{-2}\cdot e^{-z} \\
&= \frac{1}{(1+e^{-z})}\cdot\frac{e^{-z}}{(1+e^{-z})} \\
&= \frac{1}{(1+e^{-z})}\cdot\frac{1}{(1+e^{z})}
\end{aligned}$$

△ Then, get $g(z)(1-g(z))$, where $g(z)=\frac{1}{1+e^{-z}}$

$$\begin{aligned}
g(z)(1-g(z)) &= \frac{1}{(1+e^{-z})}\cdot(1-\frac{1}{(1+e^{-z})}) \\
&= \frac{1}{(1+e^{-z})}\cdot\frac{(1+e^{-z})}{(1+e^{-z})}-\frac{1}{(1+e^{-z})} \\
&= \frac{1}{(1+e^{-z})}\cdot\frac{e^{-z}}{(1+e^{-z})} \\
&= \frac{1}{(1+e^{-z})}\cdot\frac{1}{(1+e^{z})}
\end{aligned}$$

△ Therefore, from above, it can be seen that

$$\frac{\partial g}{\partial z} = g(z)(1-g(z))$$

(b)

△ From above, $g(z) = \frac{1}{(1+e^{-z})}$.
△ Get $1 - g(z)$,

$$\begin{aligned} 1 - g(z) &= \frac{(1+e^{-z})}{1+e^{-z}} - \frac{1}{1+e^{-z}} \\ &= \frac{e^{-z}}{1+e^{-z}} \\ &= \frac{1}{(1+e^{z})} \end{aligned}$$

△ Then get $g(-z)$,

$$\begin{aligned} g(-z) &= \frac{1}{1+e^{-(-z)}} \\ &= \frac{1}{(1+e^{z})} \end{aligned}$$

△ Hence, from above, it can be seen that

$$1 - g(z) = g(-z)$$

Problem 2

(a)

△ Recall a function is deemed convex, if and only if

$$f(\theta z_1 + (1-\theta)z_2) \leq \theta * f(z_1) + (1-\theta) * f(z_2), \theta \in [0,1]$$

Hence, below will try to prove $f(w) = g(w^T x + y)$ is a convex function, where

$$g((\theta w_1 + (1-\theta)w_2)^T x + y) \leq \theta g(w_1^T x + y) + (1-\theta) g(w_2^T x + y)$$

△ As

$$\begin{aligned} (\theta w_1 + (1-\theta)w_2)^T x + y &= \theta w_1^T x + (1-\theta) w_2^T x + y \\ &= \theta w_1^T x + \theta y + (1-\theta) w_2^T x + y - \theta y \\ &= \theta(w_1^T x + y) + (1-\theta)(w_2^T x + y) \end{aligned}$$

Therefore,

$$\begin{aligned} & g((\theta w_1 + (1-\theta)w_2)^T x + y) \leq \theta g(w_1^T x + y) + (1-\theta) g(w_2^T x + y) \\ \Rightarrow & g(\theta(w_1^T x + y) + (1-\theta)(w_2^T x + y)) \leq \theta g(w_1^T x + y) + (1-\theta) g(w_2^T x + y) \end{aligned}$$

and as g is convex, we hence have

$$g(\theta z_1 + (1-\theta)z_2) \leq \theta g(z_1) + (1-\theta)g(z_2)$$

△ Hence, from above:
we let $z_1 = w_1^T x + y$ and $z_2 = w_2^T x + y$. Hence, $g(\langle w, x\rangle + y)$ is convex, and hence $f(w)$ is convex.

(b)

△ Again, $g(x)$ is convex if and only if,

$$g(\theta x_1 + (1-\theta)x_2) \leq \theta g(x_1) + (1-\theta)g(x_2)$$

△ And,

$$\begin{aligned}
g(\theta x_1 + (1-\theta)x_2) &= max\{f_1, f_2, ..., f_r\} \\
&= max\{f_1(\theta x_1 + (1-\theta)x_2), f_2(\theta x_1 + (1-\theta)x_2), ..., f_r(\theta x_1 + (1-\theta)x_2), \} \\
&\leq max\{ \\
&\quad\quad \theta f_1(x_1) + (1-\theta) + f_1(x_2), \\
&\quad\quad \theta f_2(x_1) + (1-\theta) + f_2(x_2), \\
&\quad\quad ..., \\
&\quad\quad \theta f_r(x_1) + (1-\theta) + f_r(x_2) \\
&\quad \} \\
&\quad\quad (as \quad f \quad is \quad convex) \\
&\leq \theta max\{f_1(x_1), f_2(x_1), ..., f_r(x_1)\} + (1-\theta)max\{f_1(x_2), f_2(x_2), ..., f_r(x_2)\} \\
&\quad\quad (as \quad max\{\} \quad is \quad convex) \\
&= \theta g(x_1) + (1-\theta)g(x_2)
\end{aligned}$$

△ Therefore, $g(x)$ is convex.

Problem 3

(a)

△ First, by mean value theorem, we have: for a continuous function f on a closed interval $[a, b]$, $\exists c$, such that:

$$f^{'}(c) = \frac{f(b) - f(a)}{b - a}$$

△ Therefore, for function $\nabla f(x)$ there $\exists z$ such that,

$$\nabla^2 f(z) \cdot (x-y) = \nabla f(x) - \nabla f(y)$$

△ Meanwhile, we know that a function is L-smooth when,

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x-y\|$$

△ Therefore, from above,

$$\begin{aligned}\|\nabla^2 f(z) \cdot (x-y)\| &= \|\nabla f(x) - \nabla f(y)\| \\ &\leq L\|x-y\|\end{aligned}$$

and hence,

$$\begin{aligned}&\|\nabla^2 f(z) \cdot (x-y)\| \leq L\|x-y\| \\ \Rightarrow &\|\nabla^2 f(z) \cdot (x-y)\| \leq \|\nabla^2 f(z)\| \cdot \|(x-y)\| \leq L\|x-y\| \\ &\quad (from \quad Cauchy-Schwarz \quad inequality) \\ \Rightarrow &\|\nabla^2 f(z)\| \leq L\end{aligned}$$

assume the case of 2-norm and with the fact that $\nabla^2 f(x)$ is symmetric, the above expression indicates,

$$\nabla^2 f(x) \preceq LI$$

△ Echoing the problem description and from above, we therefore have,

$$\begin{aligned}&\|\nabla^2 f(z) \cdot (x-y)\| \leq L\|x-y\| \\ \Rightarrow &\|(\nabla^2 f(z) \cdot (x-y))^T (x-y)\| \leq L\|x-y\|^2 \\ \Rightarrow &(\nabla^2 f(z) \cdot (x-y))^T (x-y) \leq L\|x-y\|^2 \\ \Rightarrow &\langle \nabla^2 f(z)(x-y), (x-y) \rangle \leq L\|x-y\|^2 \\ \Rightarrow &\langle \nabla^2 f(z) v, v \rangle \leq L\|v\|^2\end{aligned}$$

(b)

△ First, we let

$$\begin{aligned}z(t) &= x + t(y-x) \\ g(t) &= f(z(t)) \\ \therefore g(0) &= f(x) \\ g(1) &= f(y) \\ g^{'}(t) &= \nabla f(z(t))^T (y-x) \\ g^{'}(0) &= \nabla f(x)^T (y-x)\end{aligned}$$

$\triangle$ Also, with fundemental theorem, we have

$$
\begin{aligned}
g(1) - g(0) &= \int_1^0 g^{'}(t)\, dt \\
\therefore g(1) - g(0) - g^{'}(0) &= \int_1^0 (g^{'}(t) - g^{'}(0))\, dt \\
&\leq \int_1^0 \|g^{'}(t) - g^{'}(0)\|\, dt
\end{aligned}
$$

$\triangle$ And for $\|g^{'}(t) - g^{'}(0)\|$,

$$
\begin{aligned}
\|g^{'}(t) - g^{'}(0)\| &= \|\nabla f(z(t)^T(y-x) - \nabla f(x)^T(y-x))\| \\
&= \|(\nabla f(z(t)) - \nabla f(x))^T(y-x)\| \\
&\leq \|(\nabla f(z(t)) - \nabla f(x))\| \cdot \|(y-x)\| \\
&\leq L\|z(t) - x\| \cdot \|y-x\| \\
&= L\|x + ty - tx - x\| \cdot \|y-x\| \\
&= tL\|y-x\|^2
\end{aligned}
$$

$\triangle$ Therefore,

$$
\begin{aligned}
\int_1^0 \|g^{'}(t) - g^{'}(0)\|\, dt &\leq \int_1^0 tL\|y-x\|^2\, dt \\
&= L\|y-x\|^2[\frac{1}{2}t^2]_0^1 \\
&= \frac{1}{2}L\|y-x\|^2
\end{aligned}
$$

$\triangle$ Therefore,

$$
\begin{aligned}
g(1) - g(0) - g^{'}(0) &= f(y) - f(x) - \nabla f(x)^T(y-x) \\
&\leq \frac{1}{2}L\|y-x\|^2
\end{aligned}
$$

and get,

$$
\Rightarrow f(y) \leq f(x) + \nabla f(x)^T(y-x) + \frac{1}{2}L\|y-x\|^2
$$