COMP5212: Machine Learning **Fall 2023**

Homework 1: Due Sunday Oct. 1, 11:59 PM

Instructions: upload a PDF report using LaTeX containing your answers to Canvas (remember to include your name and ID number).

$\partial (1+e^{-z})^{-1} =$

Problem 1. Sigmoid function in logistic regression

Let $g(z) = \frac{1}{1+e^{-z}}$ be the sigmoid activation function

(a) (10 pt) Show that $\frac{\partial g}{\partial z} = g(z)(1-g(z))$

(b) (10 pt) Show that $1 - g(z) = g(-z)$

$\frac{\partial}{\partial z}(1+e^{-z})^{-1} = +1(1+e^{-z})^{-2}$

$\frac{1}{1+e^{-z}} - \frac{1}{(1+e^{-z})^2}$ $\qquad \frac{1}{(1+e^{-z})^2}$

$= \frac{1+e^{-z}-1}{(1+e^{-z})^2} = \frac{e^{-z}}{(1+e^{-z})^2}$

Problem 2. Convexity

(a) (15 pt) Assume that $f : \mathbb{R}^d \to \mathbb{R}$ can be written as $f(\boldsymbol{w}) = g(\langle \boldsymbol{w}, \boldsymbol{x} \rangle + y)$, for some $\boldsymbol{x} \in \mathbb{R}^d$, $y \in \mathbb{R}$, and $g : \mathbb{R} \to \mathbb{R}$. Prove f is convex if g is convex.

(b) (15 pt) For $i = 1, \ldots, r,$, let $f_i : \mathbb{R}^D \to \mathbb{R}$ be a convex function. Prove the $g(x) = \max_{i \in [r]} f_i(x)$ from $\mathbb{R}^d$ to $\mathbb{R}$ is also convex.

Problem 3. Smoothness

A differential function f is said to be L-smooth if its gradietns are Lipschitz continuous, that is

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

let $f : \mathbb{R}^d \to \mathbb{R}$ be a twice differentiable function. If f is L-smooth then prove the following inequality:

- (25 pt) Prove $\langle \nabla^2 f(x)v, v\rangle \leq L\|v\|_2^2, \quad \forall x, v \in \mathbb{R}^d$
- (25 pt) Prove $f(y) \leq f(x) + \langle \nabla f(x), y - x\rangle + \frac{L}{2}\|y - x\|_2^2$

$f: \mathbb{R}^d \to \mathbb{R}$ $\qquad f(y) \leq f(x) + \nabla f(x)^T (y-x) + \frac{L}{2}\|y-x\|_2^2$

$\nabla f \in \mathbb{R}^d \qquad x \in \mathbb{R}^d$

$\nabla^2 f \in \mathbb{R}^{d \times d} \qquad v \in \mathbb{R}^d$

$\underset{\mathbb{R}^{d\times d}\ \mathbb{R}^{d\times 1}}{\nabla^2 f(x)\, v} \in \mathbb{R}^d$

$(\nabla^2 f(x) v)^T v \leq L\|v\|_2^2$

$\nabla^2 f(x)\, \underbrace{v^T v} \leq L\, \underbrace{v^T v}$

$\nabla^2 f(x) v^T v \leq L\, v^T v$

$\nabla^2 f(x)^T 1 \leq L 1$

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

$\nabla^2 f(x) \preceq L\, I$

$$g(z) = \frac{1}{1+e^{-z}}$$

$$\frac{\partial g}{\partial z} = \frac{1}{1+e^{-z}}\left(1-\frac{1}{1+e^{-z}}\right)$$

$$= \left(\frac{1}{1+e^{-z}}\right)\left(\frac{1}{1+e^{z}}\right)$$

$$\frac{1}{1} - \frac{1}{1+e^{-z}} = \frac{1+e^{-z}-1}{1+e^{-z}} = \frac{e^{-z}}{1+e^{-z}} = \frac{1}{e^{z}+1} = \frac{1}{1+e^{z}}$$

1(a)

$$g'(z) = \frac{\partial}{\partial z}(1+e^{-z})^{-1}$$

$$= -1\,(1+e^{-z})^{-2}\,(1+e^{-z})'$$

$$= \frac{1}{(1+e^{-z})^2}\,e^{-z}$$

$$= \frac{1}{1+e^{-z}}\,\frac{e^{-z}}{(1+e^{-z})}$$

$$= \frac{1}{(1+e^{-z})}\,\frac{1}{(1+e^{z})}$$

#

1(b)

$$g(z) = \frac{1}{1+e^{-z}}$$

$$1-g(z) = \frac{1+e^{-z}}{1+e^{-z}} - \frac{1}{1+e^{-z}} = \frac{e^{-z}}{1+e^{-z}} = \frac{1}{1+e^{z}}$$

$$g(-z) = \frac{1}{1+e^{-(-z)}} = \frac{1}{1+e^{z}}$$

#

2(a)

$f: \mathbb{R}^d \to \mathbb{R}$,

$\mathbb{R}^d$

$f(w) = g(w^T x + y)$

$\mathbb{R}^d$. $\quad$ $\mathbb{R}$

proof f is convex if g is convex.

---

proof $f(\theta z_1 + (1-\theta) z_2) \le \theta f(z_1) + (1-\theta) f(z_2)$

$w^T x + y$.

$\triangle \quad [\theta z_1 + (1-\theta) z_2]^T x + y$

$= \theta z_1^T x + (1-\theta) z_2^T x + y$

$= \theta(z_1^T x + y) + (1-\theta)(z_2^T x + y)$

$f(w) = g(w^T x + y)$

$g(w^T x + y)$

proof:

$g([\theta w_1 + (1-\theta) w_2]^T x + y) \le \theta g(w_1^T x + y) + (1-\theta)$
$g(w_2^T x + y)$

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

as g is convex

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

let $w_1^T x + y = z_1$

$w_2^T x + y = z_2$

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

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

hence,

$g(\langle w, x\rangle + y)$ is convex

hence,

$f(w) =$ convex

2(b) for $i = 1, \dots, r$

let $f_i : \mathbb{R}^p \to \mathbb{R}$ be convex

prove $g(x) = \max_{i \in [k]} f_i(x)$ from $\mathbb{R}^d$ to $\mathbb{R}$.

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

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

$$= \max \{ f_1(\theta x_1 + (1-\theta)x_2), f_2(\theta x_1 + (1-\theta)x_2) \cdots f_r(\theta x_1 + (1-\theta)x_2 \}$$

$$\leq \max \{ \theta f_1(x_1) + (1-\theta) f_1(x_2), \theta f_2(x_1) + (1-\theta) f_2(x_2) \cdots \theta f_r(x_1) + (1-\theta) f_r(x_2) \}$$

max func. f convex

$$\leq \theta \max \{ f_1(x_1), f_2(x_1) \dots, f_r(x_1) \} + (1-\theta) \max \{ f_1(x_2), f_2(x_2) \cdots, f_r(x_2) \}$$

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

3-1

Δ By mean value theorem

$\exists c,$ s.t. $f'(c) = \dfrac{f(b)-f(a)}{b-a}$

Δ for L-smooth

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

Δ $\exists z,$ s.t. $\nabla^2 f(z) = \dfrac{\nabla f(x) - \nabla f(y)}{x-y}$

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

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

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

(its max should $\leq L \|x-y\|$)

$\| \nabla^2 f(z)(x-y) \| \leq L \| x-y \|$

$\|\nabla^2 f(z)(x-y)\| \leq \| \nabla^2 f(z) \| \, \cancel{\|(x-y)\|} \leq L \, \cancel{\|x-y\|}$

max(eigen value) $\leq L$ $\therefore$ $\nabla^2 f(z) \preceq LI$.

3-2

$z(t) = x + t(y-x)$

$g(t) = f(z(t))$

$\therefore$ $g(0) = f(x)$

$g(1) = f(y)$

$g'(t) = \nabla f(z(t))^T (y-x)$

$g'(0) = \nabla f(x)^T (y-x)$

1. fundamental theorem

$g(1) - g(0) = \int_0^1 g'(t)\, dt$

$\therefore$ $g(1) - g(0) - g'(0) = \int_0^1 [g'(t) - g'(0)]\, dt$

$\leq \int_0^1 | g'(t) - g'(0) |\, dt$

$| g'(t) - g'(0) | = | \nabla f(z(t))^T (y-x) - \nabla f(x)^T (y-x) |$

$= | [\nabla f(z(t)) - \nabla f(x)]^T (y-x) |$

$\leq \| \nabla f(z(t)) - \nabla f(x) \| \, \| y-x \|$

$\leq L \| z(t) - x \| \cdot \| y-x \|$

$= L \| \cancel{x} + ty - tx - \cancel{x} \| \cdot \|y-x\|$

$= tL \| y-x \|^2$

$$\therefore = \int_0^1 |g'(t) - g'(0)|\, dt$$

$$= \int_0^1 t L \|y-x\|^2 dt$$

$$= L\|y-x\|^2 \left[\left(\frac{1}{2} t^2 \right) \right]_0^1$$

$$= \frac{1}{2} L \|y-x\|^2$$

$$\therefore f(y) - f(x) - \nabla f(x)^T (y-x) \le \frac{1}{2} L \|y-x\|^2$$

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