

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

**Instructions:** upload a PDF report using L<sup>A</sup>T<sub>E</sub>X containing your answers to Canvas (remember to include your name and ID number).

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

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

### Problem 2. Convexity

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

(b) (15 pt) For  $i = 1, \dots, r$ , let  $f_i : \mathbb{R}^D \rightarrow \mathbb{R}$  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 differentiable function  $f$  is said to be  $L$ -smooth if its gradients are Lipschitz continuous, that is

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

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

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

$$f: \mathbb{R}^d \rightarrow \mathbb{R}$$

$$\begin{aligned} \nabla f &\in \mathbb{R}^d & \mathbf{x} &\in \mathbb{R}^d \\ \nabla^2 f &\in \mathbb{R}^{d \times d} & \mathbf{v} &\in \mathbb{R}^d \end{aligned}$$

$$\nabla^2 f(\mathbf{x}) \mathbf{v} \in \mathbb{R}^d$$

$$\begin{aligned} (\nabla^2 f(\mathbf{x}) \mathbf{v})^T \mathbf{v} &\leq L\|\mathbf{v}\|_2^2 \\ \nabla^2 f(\mathbf{x}) \mathbf{v}^T \mathbf{v} &\leq L\mathbf{v}^T \mathbf{v} \end{aligned}$$

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

$$\nabla^2 f(\mathbf{x})^T \mathbf{I} \leq L\mathbf{I}$$

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

$$\nabla^2 f(\mathbf{x}) \preceq L\mathbf{I}$$

$$f(\mathbf{y}) \leq f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}) + \frac{L}{2}\|\mathbf{y} - \mathbf{x}\|_2^2$$

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

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

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

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

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

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

1(a)

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

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

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

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

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

✗

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)

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

$$f(w) = g(\underbrace{w^T x + y}_{\mathbb{R}^d \rightarrow \mathbb{R}})$$

proof  $f$  is convex if  $g$  is convex.

---

$$\text{proof } f(\theta z_1 + (1-\theta)z_2) \leq \theta f(z_1) + (1-\theta)f(z_2)$$

$$w^T x + y$$

$$\begin{aligned} \Delta & [\theta z_1 + (1-\theta)z_2]^T x + y \\ &= \theta z_1^T x + (1-\theta)z_2^T x + y \\ &= \theta(z_1^T x + y) + (1-\theta)(z_2^T x + y) \end{aligned}$$

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

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

proof:

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

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

as  $g$  is convex

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

$$\text{let } w_1^T x + y = z_1$$

$$w_2^T x + y = z_2$$

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

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

hence,

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

hence,

$$f(w) = \text{convex}$$

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

let  $f_i: \mathbb{R}^D \rightarrow \mathbb{R}$  be convex

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

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

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

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

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

$\downarrow$  max func.  $f$  convex

$$\leq \theta \max \{ f_1(x_1), f_2(x_1), \dots, f_r(x_1) \}$$

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

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



3-1

By mean value theorem

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

for  $L$ -smooth

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

$$\Delta \exists \xi, s.t. \nabla^2 f(\xi) = \frac{f(x) - f(y)}{x - y}$$

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

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

since  $\|f(x) - f(y)\| \leq L \|x - y\|$   
 (its max should  $\leq L \|x - y\|$ )  
 $\|\nabla^2 f(\xi) (x - y)\| \leq L \|x - y\|$

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

$$\max(\text{eigen value}) \leq L \therefore \nabla^2 f(\xi) \preceq LI$$

Q

3-2

$$\xi(t) = x + t(y - x)$$

$$g(t) = f(\xi(t))$$

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

$$g(1) = f(y)$$

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

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

1.

Fundamental theorem

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

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

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

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

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

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

$$\leq L \|\xi(t) - x\| \|y - x\|$$

$$\rightarrow = L \|x + t(y - x) - x\| \|y - x\|$$

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

$$\begin{aligned}
& \leq \int_0^1 |g'(t) - g'(0)| dt \\
& = \int_0^1 t L \|y - x\|^2 dt \\
& = L \|y - x\|^2 \left[ \frac{1}{2} t^2 \right]_0^1 \\
& = \frac{1}{2} L \|y - x\|^2
\end{aligned}$$

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

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