3/16/23

HW2 Rough Draft

1)

(f is strongly convex)

$(\nabla^2 f(x) \succeq mI)$ $f((1-\alpha)x+\alpha y) \leq (1-\alpha)f(x)+\alpha f(y) - \frac{1}{2}m\alpha(1-\alpha)\|x-y\|^2$,

$(z_\alpha = (1-\alpha)x+\alpha y)$

"$\Rightarrow$" (and only if) $\alpha f(y)-\alpha f(x) \geq f(z_\alpha)-f(x)+\frac{1}{2}m\alpha(1-\alpha)\|x-y\|^2$

$=((\nabla f(x))^T(z_\alpha-x)+O(\|z_\alpha-x\|^2)+\frac{1}{2}m\alpha(1-\alpha)\|x-y\|^2$,

$(z_\alpha=(1-\alpha)x+\alpha y,\ z_\alpha-x=\alpha(y-x))$

$\cancel{\alpha} f(y)-\cancel{\alpha}f(x) \geq \cancel{\alpha}(\nabla f(x))^T(y-x)+\|y-x\|^2 O(\alpha^2) + \frac{1}{2}m\cancel{\alpha}(1-\alpha)\|y-x\|^2$

$f(y)-f(x) \geq (\nabla f(x))^T(y-x)+\|y-x\|^2 O(\alpha) + \frac{1}{2}m(1-\alpha)\|y-x\|^2$

$(\alpha \to 0)$

$f(y)-f(x) \geq (\nabla f(x))^T(y-x)+\frac{1}{2}m\|y-x\|^2$,

HW2 Rough Draft

I cont)

$(y = x + \alpha u, 0 \leq t \leq 1)$

$$f(x+\alpha u) = f(x) + \alpha(\nabla f(x))^T + \frac{1}{2}\alpha^2 u^T \nabla f(x+\alpha t u)u$$

$$\geq f(x) + \alpha(\nabla f(x))^T u + \frac{m}{2}\alpha^2 \|u\|^2,$$

(choose any $u \in R^n$)

$u^T \nabla f(x+\alpha t u) \geq m\|u\|^2$

implies $\nabla^2 f(x) \geq mI$

$\left(\lambda_1, \ldots, \lambda_n \text{ of } \nabla^2 f(x) \text{ satisfies } \min_{1 \leq i \leq n} \lambda_i \geq m\right)$

"$\Leftarrow$" (if) $\nabla^2 f(x) \geq mI$

$(z \in R^n)$

$$(x-z)^T \nabla^2 f(z + t(x-z))(x-z) \geq m\|x-z\|^2$$

$$\nabla f(x) = f(z) + (\nabla f(z))^T(x-z) + \frac{1}{2}(x-z)^T \nabla^2 f(z+t(x-z))(x-z)$$

$$\geq f(z) + (\nabla f(z))^T (x-z) + \frac{m}{2} \|x-z\|^2$$

$$f(y) \geq f(z) + (\nabla f(z))^T (y-z) + \frac{m}{2} \|y-z\|^2$$

$$(z = (1-\alpha)x + \alpha y)$$

$$\begin{aligned}(1-\alpha)f(x) + \alpha f(y) \geq\ & (\alpha + (1-\alpha))f(z) \\ & + (\nabla f(z))^T((1-\alpha)(x-z) + \alpha(y-z)) \\ & + \frac{m}{2}\left((1-\alpha)\|x-z\|^2 + \alpha\|y-z\|^2\right)\end{aligned}$$

$$\begin{aligned}= f(z) &+ (\nabla f(z))^T((1-\alpha)(x-z) + \alpha(y-z)) \\ &+ \frac{m}{2}\left((1-\alpha)\|x-z\|^2 + \alpha\|y-z\|^2\right)\end{aligned}$$

$$\left(z = (1-\alpha)x + \alpha y \Rightarrow \begin{cases} x-z = \alpha(x-y) \\ y-z = (1-\alpha)(y-x) \end{cases} \right)$$

$$\begin{aligned}= f(z) &+ (\nabla f(x))^T\left[(1-\alpha)\alpha(x-y) + \alpha(1-\alpha)(y-x)\right] \\ &+ \frac{m}{2}\left[(1-\alpha)\alpha^2\|x-y\|^2 + \alpha(1-\alpha)^2\|y-x\|^2\right]\end{aligned}$$

$$\cancel{\geq \alpha(1-\alpha)\|x-y\|^2}$$

HW2 Rough Draft

1 cont)

$$(1-\alpha)\|x-z\|^2+\alpha\|y-z\|^2=[(1-\alpha)\alpha^2+\alpha(1-\alpha)^2]\|x-y\|^2$$

$$=\alpha(1-\alpha)\|x-y\|^2$$

(Sub)

$$(1-\alpha)f(x)+\alpha f(y)\geq f((1-\alpha)x+\alpha y)+\frac{m}{2}\alpha(1-\alpha)\|x-y\|^2$$

$\therefore$

2)

Given: $a^{[1]}=x$, $z^{[l]}=W^{[l]}a^{[l-1]}+b^{[l]}$, $a^{[1]}=x, a^{[L]}=g(x;w)$ weights

$a^{[l]}=\sigma(z^{[l-1]})$ for $l=2,3,\dots,L$

$C=C(W,b)=\frac{1}{2}\|y-a^{[L]}\|^2$ $\quad x,y\in\mathbb{R}^n$

Training data $=$ 1 point (x,y), $(x\circ y)_i=x_iy_i$

$$\delta^{[l]}=(\delta_j^{[l]})_j=\begin{pmatrix}\delta_1^{[l]}\\ \vdots\\ \delta_{n_L}^{[l]}\end{pmatrix}, \delta_j^{[l]}=\frac{\partial C}{\partial z_j^{[l]}}$$

Prove: $\delta^{[l]}=\sigma'(z^{[l]})\circ(W^{[l+1]})^T\delta^{[l+1]}$

Proof:

$$\delta_j^{[l]} = \frac{\partial C}{\partial z_j^{[l]}} = \sum_{k=1}^{n_{l+1}} \frac{\partial C}{\partial z_k^{[l+1]}} \cdot \frac{\partial z_k^{[l+1]}}{\partial z_j^{[l]}}$$

$$= \sum_{k=1}^{n_{l+1}} \delta_k^{[l+1]} \frac{\partial z_k^{[l+1]}}{\partial z_j^{[l]}}$$

$$z^{[l+1]} = W^{[l+1]}\sigma(z^{[l]}) + b^{[l+1]}$$

$$\frac{\partial z_k^{[l+1]}}{\partial z_j^{[l]}} = \frac{\partial(W^{[l+1]}\sigma(z^{[l]}))_k}{\partial z_j^{[l]}}$$

$$= \left(W^{[l+1]} \frac{\partial \sigma(z^{[l]})}{\partial z_j^{[l]}}\right)_k,$$

$$\sigma(z^{[l]}) = (\sigma(z_1^{[l]}), \ldots, \sigma(z_{n_l}^{[l]})),$$

$$\frac{\partial \sigma(z^{[l]})}{\partial z_j^{[l]}} = (0, 0, \ldots, \sigma'(z_j^{[l]}), \ldots, 0)^T,$$

$\sigma'(z_j^{[l]})$ at kth position,

$$W^{[l+1]} = W_j^{[l+1]},$$

$$W^{[l+1]} \frac{\partial \sigma(z^{[l]})}{\partial z_j^{[l]}} = \sigma'(z_j^{[l]})\, W_j^{[l+1]},$$

HW2 Rough Draft

2 cont)

$$\left(W^{[l+1]} \frac{\partial \sigma(z^{[l]})}{\partial z_j^{[l]}}\right)_k = \sigma'(z_j^{[l]}) \overbrace{W_{jk}^{[l+1]}}^{\text{Element at } W[k,j]}$$

$$\delta_j^{[l]} = \sum_{k=1}^{n_{l+1}} \delta_k^{[l+1]} \sigma'(z_j^{[l]}) W_{jk}^{[l+1]}$$

$$= \sigma'(z_j^{[l]}) \left[(W^{l+1})^T \delta^{[l+1]}\right]_j$$

$$= \sigma'(z_j^{[l]}) \circ \left[(W^{l+1})^T \delta^{[l+1]}\right]_j$$

$$= \sigma'(z_j^{[l]}) \circ (W^{l+1})^T \delta^{[l+1]} \quad \therefore$$

3)

Given $x^{k+1} = x^k + \alpha d^k$, $\alpha > 0$, $d^k \in \mathbb{R}^n$

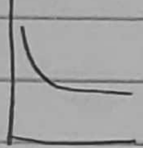
$f(x^{k+1}) \leq f(x^k) - C\|\nabla f(x^k)\|^2$

$C > 0$, $\nabla f = L$-Lipschitz. $\underbrace{x_{n_k}}_{\text{Subsequence}} \to \bar{x}$, $k \to \infty$

Prove: $\nabla f(\bar{x}) = 0$

Proof, $f(x^{k+1}) \leq f(x^k) - C\|\nabla f(x^k)\|^2$

$$\|\nabla f(x^k)\|^2 \leq \frac{f(x^k) - f(x^{k+1})}{C}$$



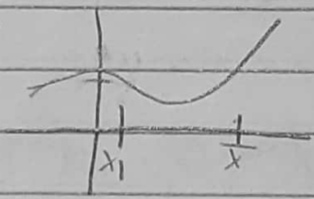
$\{f(x^k)\}$: $f(x^{k+1}) \leq f(x^k)$ (must converge eventually)

monotonic bounded

$f(x^{k+1}) - f(x^k) \to 0$ as $k \to \infty$

(just a guess using class notes

$x_{k_j} \to \bar{x}$



$\{x_k\} \supseteq \{x_{k_j}\}$ $x_{k_j} \to \bar{x}$

$\lim_{j\to\infty} x_{k_j} = \bar{x}$, $\lim_{j\to\infty} f(x_{k_j}) = L$

$$0 \leq \|\nabla f(\vec{x})\|^2 \leq \frac{f(\vec{x}_{k+1}) - f(\vec{x}_k)}{C}$$

$$\lim_{k\to\infty} [f(\vec{x}_{k+1}) - f(\vec{x}_k)] = 0$$

$$\nabla f(\vec{x}) = \lim_{j\to\infty} \nabla f(\vec{x}_{k_j}) + \lim_{k\to\infty} \nabla f(\vec{x}_k) = 0$$

HW 2 Rough Draft

3 cont)

$\{f(\vec{x}_{k_j})\}_{j=1}^{\infty}$

$f(\vec{x}_k) - f(\vec{x}_{k+1}) \to 0$ as $k \to \infty$