日期: /

Name: ZHONG Qiao Yang Net ID: 24112456g

1. (a) $\frac{d}{dx}\frac{1}{1+e^{-x}} = \boxed{\frac{e^{-x}}{(1+e^{-x})^2}}$,

$\lim_{x\to+\infty}\frac{d}{dx}\text{sigmoid}(x) = \boxed{0,}$

$\lim_{x\to-\infty}\frac{d}{dx}\text{sigmoid}(x) = \boxed{0.}$

(b) $\frac{d}{dx}\frac{e^{2x}-1}{e^{2x}+1} = \frac{d}{dx}\left(1-\frac{2}{e^{2x}+1}\right)$

$= \boxed{\frac{4e^{2x}}{(e^{2x}+1)^2}}$

$\lim_{x\to+\infty}\frac{d}{dx}\tanh(x) = \boxed{0,}$

$\lim_{x\to-\infty}\frac{d}{dx}\tanh(x) = \boxed{0}$

(c) $\text{leaky Relu}(x) = \begin{cases} ax & \text{if } ax > x \\ x & \text{if } ax \le x \end{cases}$, $a \in (0,1)$

$\frac{d}{dx}\text{Leaky Relu}(x) = \boxed{\begin{cases} a & \text{if } ax > x \\ 1 & \text{if } ax < x \end{cases}}$,

$\lim_{x\to+\infty}\frac{d}{dx}\text{LRelu}(x) = \boxed{1,}$

$\lim_{x\to-\infty}\frac{d}{dx}\text{LRelu}(x) = \boxed{a.}$

日期: /

(b) ① Forward,

$$a = W^{(1)} x + b^{(1)} = \begin{pmatrix} -3.2 \\ 1.3 \end{pmatrix}$$

$$h = \sigma(a) = \max\{a, 0\} = \begin{pmatrix} 0 \\ 1.3 \end{pmatrix}$$

$$\hat{y} = W^{(2)} h + b^{(2)} = 1.33$$

we have $x = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$, we get $\boxed{\hat{y} = 1.33} \triangleq f(x;\theta)$

(forward value)

② Backward,

$$L = (f(x;\theta) - y)^2/2. = (\hat{y} - y)^2/2.$$

$$\frac{\partial L}{\partial \hat{y}} = \hat{y} - y = 1$$

$w^1, b^1, w_2, b^2.$

$$\frac{\partial L}{\partial w^2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w^2} = \frac{\partial L}{\partial \hat{y}} \cdot h = \begin{pmatrix} 0 \\ 1.3 \end{pmatrix}$$

$$\frac{\partial L}{\partial b^2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b^2} = 1$$

$$\frac{\partial L}{\partial h} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial h} = \frac{\partial L}{\partial \hat{y}} \cdot W^2 = \begin{pmatrix} 0.6 \\ 1.1 \end{pmatrix}$$

$$\frac{\partial L}{\partial a} = \frac{\partial L}{\partial h} \cdot \frac{\partial h}{\partial a} = \frac{\partial L}{\partial h} \cdot I(a>0) = \begin{pmatrix} 0 \\ 1.1 \end{pmatrix}$$

$$\frac{\partial L}{\partial W_1} = \frac{\partial L}{\partial a} \cdot \frac{\partial a}{\partial W^1} = \frac{\partial L}{\partial a} \cdot x = \begin{pmatrix} 0 & 0 & 0 \\ 1.1 & 2.2 & 3.3 \end{pmatrix}$$

$$\frac{\partial L}{\partial b_1} = \frac{\partial L}{\partial a} \cdot \frac{\partial a}{\partial b^1} = \frac{\partial L}{\partial a} \cdot 1 = \begin{pmatrix} 0 \\ 1.1 \end{pmatrix}$$

Results:
$$\boxed{\frac{\partial L}{\partial W_2} = \begin{pmatrix} 0 \\ 1.3 \end{pmatrix}, \quad \frac{\partial L}{\partial b^2} = 1, \quad \frac{\partial L}{\partial W_1} = \begin{pmatrix} 0 & 0 & 0 \\ 1.1 & 2.2 & 3.3 \end{pmatrix} \quad \frac{\partial L}{\partial b_1} = \begin{pmatrix} 0 \\ 1.1 \end{pmatrix}.}$$

日期: /

2. $R(f) = E_{(X,Y)}\{\rho_\tau(Y - f(X))\}$, assume $h_{X,Y}(x,y)$ is joint pdf.

$$= \iint [\tau - I(y - f(x) < 0)](y - f(x))\, h(x,y)\, dx\, dy$$

$$= \iint \tau (y - f(x))\, h(x,y)\, dx dy - \int_{-\infty}^{f(x)} \int_{-\infty}^{+\infty} (y - f(x))\, h(x,y)\, dx\, dy$$

$$\frac{dR(f)}{df} = \iint -\tau\, h(x,y)\, dx\, dy + \int_{-\infty}^{f(x)} \int_{-\infty}^{+\infty} h(x,y)\, dx\, dy$$

$$= -\tau + H_Y(f(x))$$

$$= 0$$

$\Leftrightarrow \quad H_Y(f(x)) = \tau$

$\Leftrightarrow \quad \underline{P(Y \le f(X)) = \tau}$, the discrete version is

$\underline{\frac{1}{n} \operatorname{card}(\{Y_i \le f(X_i)\}) = \tau.}$

$P(Y \le f(X)) = \tau$ is the τ-th quantile of Y given $X = x$. $\square$

日期:　　/

3. Lemma 3.1 $f(y) \le f(x) + \nabla f(x)^T (y-x) + \frac{L}{2}\|y-x\|^2$.

we plug $\theta^{k+1} = \theta^k - \frac{1}{L}\nabla f(\theta^k)$ into it, we get

$$f(\theta^{k+1}) \le f(\theta^k) - \frac{1}{L}\|\nabla f(\theta^k)\|_2^2 + \frac{L}{2}\|\frac{1}{L}\nabla f(\theta^k)\|_2^2$$

$$\le f(\theta^k) - \frac{1}{2L}\|\nabla f(\theta^k)\|_2^2$$

$$\Longleftrightarrow \quad \|\nabla f(\theta^k)\|_2^2 \le 2L[f(\theta^k) - f(\theta^{k+1})]$$

$$\Longrightarrow \quad \min_{1\le k\le T} \|\nabla f(\theta^k)\|_2^2 \le 2L[f(\theta^k) - f(\theta^{k+1})]$$

we set $k=1,2,\cdots T$, and sum the inequality, we get

$$\min_{1\le k\le T} \|\nabla f(\theta^k)\|_2^2 \le \frac{2L(f(\theta^0) - \bar{f})}{T}. \quad \square$$