DL HW12

Ans 4 (a) $\sigma(x) = \dfrac{1}{1+e^{-x}}$

$$\boxed{\text{To prove: } \sigma(-x) = \frac{1}{1+e^{x}} = 1 - \sigma(x)}$$

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

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

$$= 1 - \frac{1 \cdot e^{x}}{1+e^{x}}$$

$$= \frac{1 + \cancel{e^{x}} - \cancel{e^{x}}}{1+e^{x}}$$

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

$$= \sigma(-x)$$

$$\therefore \quad \sigma(-x) = 1 - \sigma(x)$$

Hence proved.

(b) $\sigma'(x) = \frac{\partial \sigma(x)}{\partial x} = \sigma(x)\,(1-\sigma(x)) \;\forall x$

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

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

$$= \frac{e^{-x}}{1+e^{-x}} \cdot \frac{1}{1+e^{-x}}$$

$$= \frac{\frac{1}{e^{x}}}{1+e^{-x}} \cdot \frac{e^{x}}{1+e^{x}}$$

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

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

$$= \sigma(x) \cdot \sigma(-x)$$

Using eqn in 4(a)

$$\boxed{\sigma'(x) = \sigma(x)(1-\sigma(x))}$$

Hence proved!

Ans 2 (a) $f(x, y) = x^4 + xy + x^2$

Since it's a polynomial func. the domain of the func. is $\mathbb{R}$

To be a PSD func., the Hessian of the function should always be non-decreasing.

$$f(x, y) = x^4 + xy + x^2$$

$$J(x, y) = \begin{bmatrix} 4x^3 + y + 2x \\ x \end{bmatrix}$$

$$H(x, y) = \begin{bmatrix} 12x^2 + 2 & 1 \\ 1 & 0 \end{bmatrix}$$

To check if $\mathbf{H}$ is convex, we'll find points (z) where $z^T \mathbf{H} z < 0$

Let $\mathbf{z} = [x, y]^T$

$$\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 12x^2 + 2 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} < 0$$

$$\begin{bmatrix} 12x^3+2x+y & x \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} < 0$$

$$\Rightarrow \quad 12x^4 + 2x^2 + xy + xy$$

$$\Rightarrow \quad 12x^4 + 2x^2 + 2xy < 0$$

$$\Rightarrow \quad 6x^3 + x + y < 0$$

$\therefore$ for pt, $x = 1$, $y = -10$

$\Rightarrow$ Let $z = (1, -10)$

$$= \begin{bmatrix} 1 \\ -10 \end{bmatrix}^T \begin{bmatrix} 14 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ -10 \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -10 \end{bmatrix}$$

$$= 2 - 10$$

$$= 8$$

$\Rightarrow$ $f(x, y)$ is non convex as H of the func. is not a PSD.

(b) $$f_{MSE}(w) = \frac{1}{2n} (X^T w - y)^T (X^T w - y)$$

$$J(w) = \frac{1}{2n}\frac{\partial}{\partial w}\left[(X^T w - y)^T (X^T w - y)\right]$$

$$= \frac{1}{2n} \cdot X \cdot 2(X^T w - y)$$

$$= \frac{1}{n} \cdot X(X^T w - y)$$

$$H(w) = \frac{\partial J(w)}{\partial w}$$

$$= \frac{1}{n}\frac{\partial}{\partial w}\left[\underbrace{(X X^T)}_{n \times n \text{ matrix}} w - Xy\right]$$

$$= \frac{1}{n}(X^T X)$$

$\Rightarrow$ Hessian doesn't depend on w

$$\boxed{\text{To prove: } v^T H v \geq 0}$$

$$\frac{1}{n}(v^T X^T X v) \geq 0$$

T.P. $\Rightarrow$ $v^T X^T X v \geq 0$

$$v^T X^T X v = (Xv)^T (Xv)$$

Let $Xv = z$ (a col. vector of size $(q,1)$)

$$\Rightarrow v^T X^T X v = z^T z$$

$$= z_1^2 + z_2^2 + \dots + z_n^2$$

$$= \sum_{i=1}^{q} z_i^2$$

$$\sum_{i=1}^{q} z_i^2 > 0$$

for $z \in \mathbb{R}^{q \times 1}$

Hence $f_{MSE}(w)$ is convex for a two layer linear regression model !