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

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

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

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

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

$$= \dfrac{1 + e^{x} - e^{x}}{1 + e^{x}}$$

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

$$= \sigma(-x)$$

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

Hence proved.

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

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

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

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

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

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

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

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

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 H is convex, we'll find points (z) where $z^T H z < 0$

Let $z = [x, y]^T$

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

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

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

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

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

$\therefore$ 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) $\quad f_{MSE}(w) = \dfrac{1}{2n} (X^T w - y)^T (X^T w - y)$

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

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

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

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

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

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

$\Rightarrow$ Hessian doesn't depend on w

To prove : $v^T H v \geqslant 0$

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

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

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

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

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

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

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

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

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

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

Question 3:

Hyper param list:

```
hyperparams = {
    'num_epoch_list': [25,50,100],
    'lr_list': [0.0005, 0.001, 0.003],
    'batch_size_list': [16,32,64],
    'reg_list': [0.001, 0.01, 0.1],
}
```

Optimal configuration = **lr: 0.001 batch:16 alpha:0.1 num_epoch: 100**

Val Loss for config: lr: 0.001 batch:16 alpha:0.001 num_epoch: 100= **123.82382946217628**

Unregularized MSE Loss on training set: 146.06827263351394

Unregularized MSE Loss on test set: 146.9004202533849