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

For Hessian, we double derivate the function with x, y to form the matrix.

$$f'(x) = 4x^3 + y + 2x \qquad \frac{df}{dy\,dx} = 1$$
$$f''(x) = 12x^2 + 2$$

$$f'(y) = x \qquad \frac{df}{dx\,dy} = 1$$
$$f''(y) = 0$$

Hessian matrix will be:-

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

We need to check for the eigenvalue sign for checking for PSD.

$$(H - \lambda I) = 0$$

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

$$\therefore \lambda = \frac{12x^2+2 \pm \sqrt{(12x^2+2)^2+4}}{2}$$

Clearly, $\sqrt{(12x^2+2)^2+4} > \sqrt{(12x^2+2)^2} = 12x^2+2 > 0$

$\therefore \lambda$ can be < 0 for all $x \in \mathbb{R}$

Consider a point for $x = 1$, $y = -11$

$\therefore H = \begin{bmatrix} 12x^2+2 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 14 & 1 \\ 1 & 0 \end{bmatrix}$

$\therefore$ If $v^T H v < 0$, then matrix is not PSD if $v \in \mathbb{R}^2$ also:-

~~Consider v =~~ consider $v = \begin{bmatrix} 1 \\ -11 \end{bmatrix}$

$\cancel{v^T H v = [1 \ -6] \begin{bmatrix} 14 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ -6 \end{bmatrix} = \begin{bmatrix} 8 \\ 1 \end{bmatrix} [1 \ -6] \begin{bmatrix} 8 \\ 1 \end{bmatrix}}$

$$v^T H v = [1 \ -11] \begin{bmatrix} 14 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ -11 \end{bmatrix} = [1 \ -11] \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$
$$= -8 < 0$$

$\therefore$ Hessian isn't PSD at $(1, -8)$,,

Q-2
(b) The half MSE function can be given as:-

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

Now taking Hessian of f_{MSE}, i.e. basically derivating twice wrt w.

$$\nabla(f_{MSE}(w)) = X(X^T w - y)$$

$$\nabla(\nabla f_{MSE}(w)) = X^T X$$

$$\boxed{\nabla^2(f_{MSE}(w)) = X^T X} \boxed{= H}$$

Q-2 (b)

~~To prove Hessian~~

MSE function can be given as:-

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

Now taking Hessian of f_{MSE}, i.e. basically derivative twice wrt w.

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

$$\nabla^2 f_{MSE} (w)) = \boxed{\frac{1}{2n} X^T X}$$

Ignore $\cancel{X^T X}$ $\frac{1}{2n}$ as its a ~~tr~~ scalar so will contribute to PSD.

~~Now to prove, that $v^T H v \geq 0$~~

~~v~~

$\therefore$ Now to prove, that $v^T H v \geq 0$

$v^T H v \geqslant 0$

$v^T (X^T X) v \geqslant 0$

~~The dot~~

$(vX)^T (Xv) \geqslant 0$

The dot product of any thing with itself will never be negative

Thus we can say that H is positive definite and the half MSE function is convex in the 2-layer neural network setting.

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

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

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

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

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

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

Hence proved,,

Q-4
(2) $\sigma'(x) = \dfrac{\partial \sigma}{\partial x}(x) = \sigma(x)(1-\sigma(x))$

$$\frac{d}{dx}\sigma(x) = \frac{d}{dx}\left[\frac{1}{1+e^{-x}}\right] = \frac{d}{dx}(1+e^{-x})^{-1}$$

$$= -1[1+e^{-x}]^{-2}(-e^{-x})$$

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

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

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

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

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

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

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

$$= \sigma(x) \left[1 - \sigma(x) \right]_{,,}$$

Hence proved,

Q-3 $\boxed{f_{MSE} = 81.8860}$