Q2. (a) $f(x,y) = x^4 + xy + x^2$

$$\frac{\partial f}{\partial x} = 4x^3 + y + 2x$$

$$\frac{\partial f}{\partial y} = x$$

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

$\because$ eigen value $\lambda_1 = 6x^2 + (18x^4 + 6x^2 + 1)^{\frac{1}{2}} \cdot \sqrt{2} + 1$

$\lambda_2 = 6x^2 - (18x^4 + 6x^2 + 1)^{\frac{1}{2}} \cdot \sqrt{2} + 1$

$\lambda_2 < 0$ all the time

$\therefore H(f)$ is not PSD

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

let $e = X^T w - y \Rightarrow e_i = X_{ij} w_j - y_i$

$\therefore f(w) = \frac{1}{2n} e^T e = \frac{1}{2n} e_i e_i$

$$\frac{de_i}{dw_m} = X_{ij} \frac{\partial w_j}{\partial w_m} = X_{im}$$

$$\therefore \frac{\partial f}{\partial w_m} = \frac{1}{n} e_i \frac{de_i}{dw_m} = \frac{1}{n} e_i X_{im}.$$

$\because$ It's two-layer. moder

$$\boxed{H = \frac{\partial f}{\partial w_m \partial w_m} = \frac{1}{n} \frac{\partial e_i X_{in}}{\partial w_n} = \frac{1}{n} X_{in} X_{im} \geq 0.}$$

$v = v_n$

$v^T H v = v_m \left(\frac{1}{n} X_{in} X_{im}\right) v_m = \frac{1}{n} X_{in}^2 v_m^2 \geq 0$

Q3. Hyperparameters:

```
hyperparameters = {
    "lr": [0.5, 0.01, 0.001, 0.0001],
    "epoch": [25, 50, 75, 100],
    "batch_size": [10, 30, 50, 100],
    "alpha": [5, 1, 0.5, 0.1]}
```

Training results;

```
--------------------------------------------------------------------------------
best training loss:  100.17866662813911
best test loss:  88.65598290684237
best_learning_rate:  0.001
best_num_of_epoch:  100
best_size_of_batch:  10
best_alpha:  0.5
--------------------------------------------------------------------------------
```

Q4. (a) $\sigma(-x) = \frac{1}{1+e^{x}}$

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

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

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

$$\therefore \sigma'(x) = -(1+e^{-x})^{-2}(-e^{-x})$$

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

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