Midterm 2

1.

$$\hat{y}_i = \sum_{j=1}^{M} \frac{a_j}{1+e^{-b_j x_i}} \qquad J = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$

a)

$$\frac{\partial J}{\partial a_j} = \sum_{i=1}^{N} \frac{\partial J}{\partial \hat{y}_i} \frac{\partial \hat{y}_i}{\partial a_j} = 2\sum_{i=1}^{N} (\hat{y}_i - y_i) \frac{1}{1+e^{-b_j x_i}}$$

$$\frac{\partial J}{\partial \hat{y}_i} = -2(y_i - \hat{y}_i)$$

$$\frac{\partial \hat{y}_i}{\partial a_j} = \frac{1}{1+e^{-b_j x_i}}$$

$$\frac{\partial J}{\partial b_j} = \sum_{i=1}^{N} \frac{\partial J}{\partial \hat{y}_i} \frac{\partial \hat{y}_i}{\partial b_j} = 2\sum_{i=1}^{N} (\hat{y}_i - y_i) \frac{\left(a_j x_i e^{-b_j x_i}\right)}{(1+e^{-b_j x_i})^2}$$

$$\hat{y}_i = \sum_{j=1}^{M} a_j (1+e^{-b_j x_i})^{-1}$$

$$\frac{\partial \hat{y}_i}{\partial b_j} = -a_j (1+e^{-b_j x_i})^{-2} (-x_i e^{-b_j x_i})$$

$\cancel{N},M \quad N,M$

b)

```
def Jeval(a, b, X, y):    M,1
    A = 1/(1+ np.exp(-b[None,:] * x[:,None]) # shape: (N,M)
    yhat = np.sum( a[None,:] * A, axis=1) # shape : (N, )
    J = np.sum((y - yhat)**2)

    yerr = yhat - y

    dJ_dyhat = 2* yerr

    Jgrada = A.T. dot(dJ_dyhat) # (M,N)·(N,1) = (M,1)
    B1= a[None,:]* x[:,None] * np.exp(-b[None,:]* x[:,None] #(N,M)
    B2 = 1/(1+ np.exp(-b[None,:] * x[:,None]))**2 # (N,M)
    Jgradb = (B1 * B2).T. dot(dJ_dyhat) # (M,N)·(N,1) = (M,1)
    return J, Jgrad a, Jgradb
```

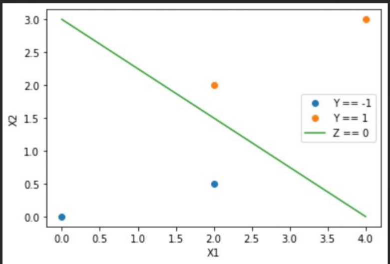
2.

a)

```python
x1 = np.array([0, 2, 2, 4])
x2 = np.array([0, 0.5, 2, 3])
y = np.array([-1, -1, 1, 1])
z = 0.75*x1 + x2 - 3
yhat = (z >= 0)

# if z = 0
# x2 = -0.75x1 + 3
p = np.arange(0, 5)

plt.plot(x1[y==-1], x2[y==-1], 'o')
plt.plot(x1[y==1], x2[y==1], 'o')
plt.plot(p, -0.75 * p + 3)
plt.xlabel("X1")
plt.ylabel("X2")

plt.legend(['Y == -1', 'Y == 1', 'Z == 0'])
```

✓ 0.1s

```
<matplotlib.legend.Legend at 0x17a3387d9a0>
```



b)

$$\epsilon = [0, 0, 0.5, 0]$$

c) minimum distance on $X = (2, 2)$

$$d = \frac{y_i \cdot z_i}{\sqrt{w_1^2 + w_2^2}} = \frac{0.5}{\sqrt{0.75^2 + 1^2}} = 0.4$$

$$W = [0.75, 1]$$

d) The boundary will be shifted to right since (0,0) is a supporting vector. How do we know? Because (2,2) has hinge loss >0 and hinge loss of (0,0) = 0

3.

a)

Hidden layer

$$Z_H = W^H X + b^H = \begin{bmatrix} W_0^H X + b_0^H \\ W_1^H X + b_1^H \end{bmatrix}$$

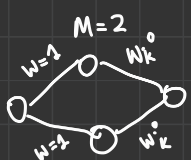
Activation : $$U_H = \max(0, Z_H) = \begin{bmatrix} \max(0, W_0^H X + b_0^H) \\ \max(0, W_1^H X + b_1^H) \end{bmatrix}$$

Output layer

$$Z_O = W^O U_H + b^O = \begin{bmatrix} W_0^O \max(0, W_0^H x + b_0^H) + b^O \\ W_1^O \max(0, W_1^H x + b_1^H) + b^O \end{bmatrix}$$

$$\hat{y} = \sum_{j=1}^{M} Z_{O,j}$$

b)



$$W_j^H = 1$$

$$Z_{H,j} = X + b_j^H$$

$$U_{H,j} = \max(0, X + b_j^H)$$

$$Z_{O,K} = W_K^O \max(0, X + b_j^H) + b^O$$

$$y = W_0^O \max(0, x + b_0^H) + b^O + W_1^O \max(0, X + b_2^H) + b^O$$

a)

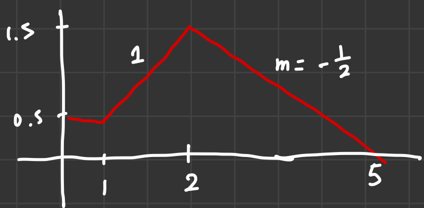
$f_2(X)$ cannot be the output since there's a jump which is not possible for a summation of two ReLU shaped inputs

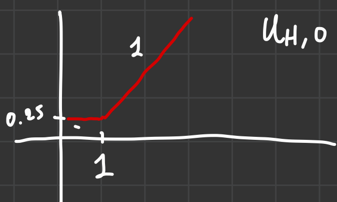
$f_3(x)$ cannot be the output since there's 3 slopes when we only have two hidden units so two slopes.

Find θ for $f_1(x)$ so $b_0^H, b_1^H, W_0^O, W_1^O, b^O$

$x = [0, 1, 2, 5]$
$y = [0.5, 0.5, 1.5, 0]$

$$y = W_0^O \max(0, x + b_0^H) + b^O + W_1^O \max(0, x + b_2^H) + b^O$$



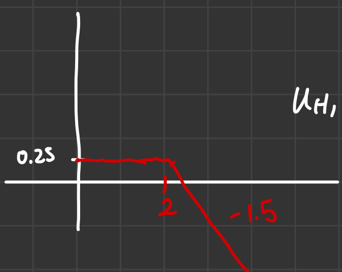


$u_{H,0}$

$W_0^O = 1$ $\quad b_o = 0.25$

$b_0^H = -1$

+



$u_{H,1}$

$W_1^O = -1.5$ $\quad b_o = 0.25$

$b_0^H = -2$