

$$1. \quad x_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad x_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \quad d_1 = -1, d_2 = 1$$

$$\text{Thus, } X = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \quad d = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \quad w = \begin{bmatrix} -1 \\ 0.5 \end{bmatrix}$$

a) Squared Error loss, thus,

Squared error loss.

$$l(w; X, d) = \|Xw - d\|_2^2$$

$$= \left\| \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ 0.5 \end{bmatrix} - \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\|_2^2$$

$$= \left\| \begin{bmatrix} -0.5 \\ -1.5 \end{bmatrix} - \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\|_2^2$$

$$= \left\| \begin{bmatrix} +0.5 \\ 0.5 \end{bmatrix} \right\|_2^2$$

$$= (+0.5)^2 + (+0.5)^2$$

$$= 0.25 + 0.25 = 0.5$$

$$b) \text{ hinge loss} = \sum_{i=1}^N (1 - d_i x_i^T w)_+$$

$$= (1 - d_1 x_1^T w)_+ + (1 - d_2 x_2^T w)_+$$

$$= \left(1 - (-1) \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ 0.5 \end{bmatrix}\right)_+ + \left(1 - (1) \begin{bmatrix} -1 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ 0.5 \end{bmatrix}\right)_+$$

$$= (1 - (-1)(-0.5))_+ + (1 - (1)(1.5))_+$$

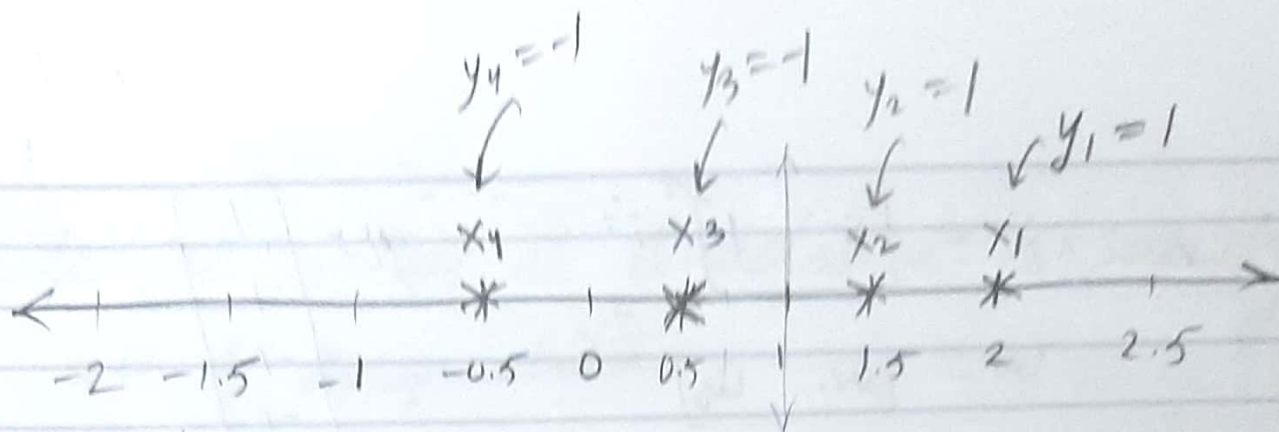
$$= (1 - 0.5)_+ + (1 - 1.5)_+$$

$$= 0.5 + 0$$

$$= \underline{\underline{0.5}}$$



2.



a) The two closest points that are classified differently are  $x_2$  &  $x_3$ .

Thus, the max margin classifier is the midpoint between the two on the real line.

$$\text{Thus, max margin classifier} = \frac{x_2 + x_3}{2} = \frac{0.5 + 1.5}{2} = \frac{2}{2} = 1$$

$$\text{or } \boxed{x = 1}$$

b) This classifier does not make any errors  
(check code later)

c) Zero hinge loss.

$$1 - y_i x_i^T w \leq 0$$

$$y_i x_i^T w \geq 1$$

$$x_i = [x_i \ 1]^T$$

$$w = [w_1 \ w_0]^T$$

$$y = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}$$

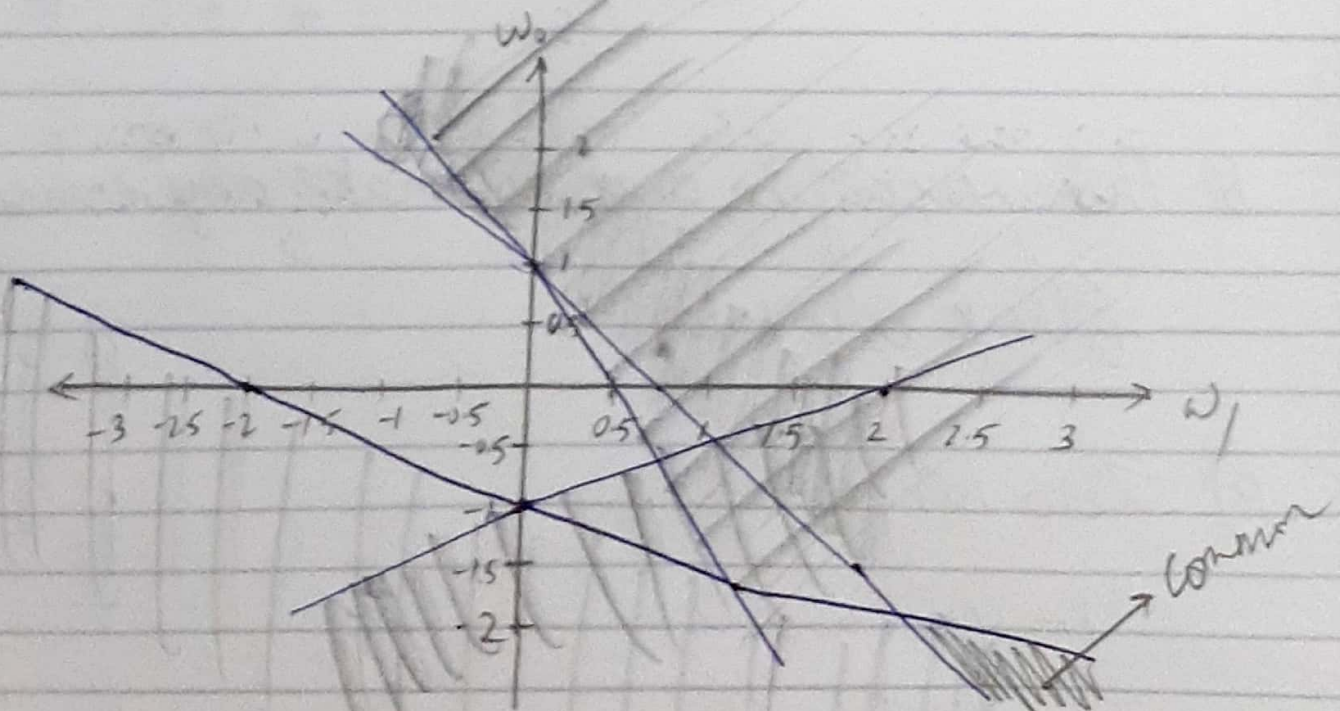
By doing  $y_i x_i^T w$  we get a set of 4 inequalities.

$$1(2w_1 + w_0) \geq 1$$

$$1(1.5w_1 + w_0) \geq 1$$

$$-1(0.5w_1 + w_0) \geq 1$$

$$-1(-0.5w_1 + w_0) \geq 1$$





$$(w_1, w_0)$$

We see, values in Quadrant 4 satisfy all 4 inequalities.

$$\text{Ex. } (w_1 = 100, w_0 = -100)$$

This classifier makes no errors.

d) This classifier makes errors.  
(Find in code).

e) Yes, we can find a classifier with zero hinge loss when  $x_4 = -5$ .

→ It makes no errors.

The misclassified point gives 0 error when classified with hinge loss.

3 a) There are 1213 classification errors with this sum.

b) There are 495 errors with this Squared error classifier.

c) The sum is not affected at all by the newly added points and has the same number of errors as before.

(No change)

d) The error rate increases by a lot after the new points are added.

(nearly 2668 errors)

The sum proves to be much better when data points are far away from the decision boundary.

This is a result of the squared error classifier placing a lot of emphasis on the datapoint's distance from the boundary of separation.