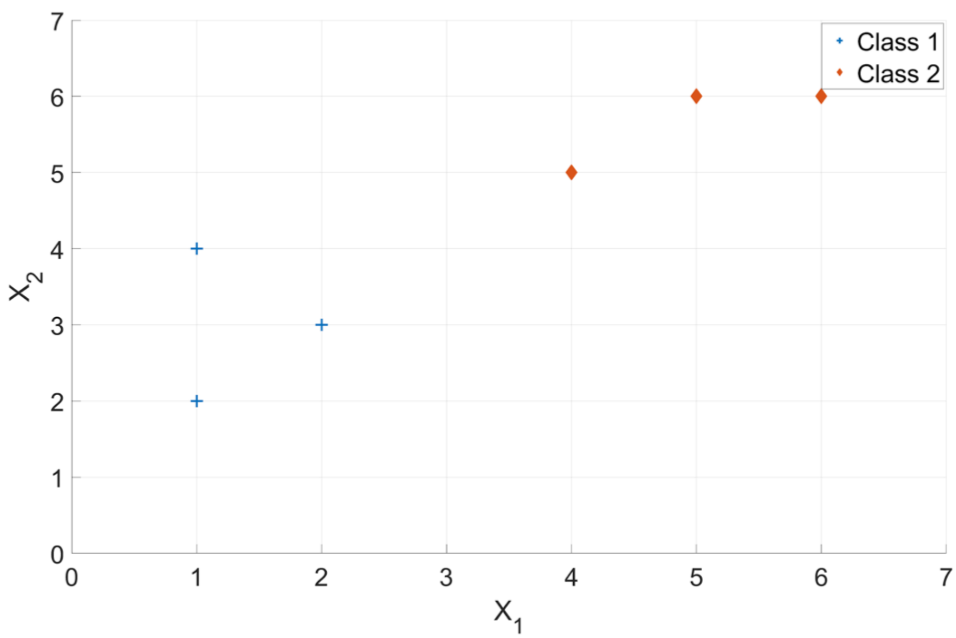
1. Hard Margin SVM (5 Points)



A decision boundary which leads to the largest margin is learnt by the Support Vector Machines. Consider that SVM trains on *Dataset I*, given in the above figure consisting of 6 data points in total as shown. There are two class labels, namely - Class Label 1, which is denoted by a blue plus and Class Label 2, which is denoted by a red diamond.

(a) [1 Points] Given the *Dataset I* above and the support vectors (4, 5) and (2, 3), what is the slope of the decision boundary?

The decision boundary should across (3, 4), prependicular to line between support vectors (4, 5) and (2, 3).
So the slope is -1.

(b) [3 Points] Based on the slope calculated above, find the normal vector (w) and bias (w_0). Based on the above findings, what is the equation to represent the decision boundary? Note - decision boundary should follow $w_0 + w_1x_1 + w_2x_2 = 0$ and the normal vector is of the form (w_1, w_2). Use the point slope form.

From (a), the slope is -1 and cross (3, 4).

$(x_2 - 4) = -1(x_1 - 3)$

$x_2 - 4 = -x_1 + 3$

$x_1 + x_2 = 7$

From this equation, we have $w_1 = w_2$, so:

$w_0 + w_1 2 + w_2 3 = 1$

$w_0 + w_1 4 + w_2 5 = -1$

$w_1 = w_2 = -\frac{1}{2}$

$w_0 = \frac{7}{2}$

So the equation of decision boundary is:

$\frac{7}{2} - \frac{1}{2}x_1 - \frac{1}{2}x_2 = 0$

(c) [1 Points] Based on the decision boundary, provide any other support vectors.

(1, 4)

2. Kernelized SVM (10 Points)

Suppose we have two sets of data points in two-dimensional space. One set represents the positive class (+1) and the second set represents the negative class (-1).

1st Class: {(2,2), (2,-2), (-2,-2), (-2,2)}
2nd Class: {(1,1), (1,-1), (-1,-1), (-1,1)}

By plotting these points on the 2-D plane, we can easily infer that these points are not linearly separable. No problem! We are giving you a Kernel or mapping function that will help you find the decision boundary.

$$\Phi_1\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{cases} \begin{pmatrix} 4 - x_2 + |x_1 - x_2| \\ 4 - x_1 + |x_1 - x_2| \end{pmatrix}, & if \sqrt{x_1^2 + x_2^2} > 2 \\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, & otherwise \end{cases}$$

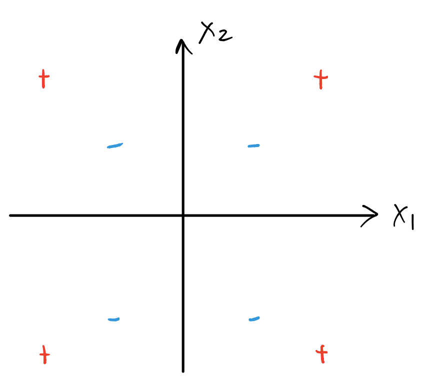
Now just follow these following steps which will enable you to classify the point of your own choice.

(a) [2 Points] Using this Kernel function, find the new feature representation of each data point from both classes.

1st: $\begin{pmatrix} 2 \\ 2 \end{pmatrix} \begin{pmatrix} 10 \\ 6 \end{pmatrix} \begin{pmatrix} 6 \\ 6 \end{pmatrix} \begin{pmatrix} 6 \\ 10 \end{pmatrix}$

2nd: $\begin{pmatrix} 1 \\ 1 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \begin{pmatrix} -1 \\ -1 \end{pmatrix} \begin{pmatrix} -1 \\ 1 \end{pmatrix}$

(b) [2 Points] Now find the two suitable support vectors and their corresponding augmented support vectors. Given a support vector (x_1, x_2), we define its augmented support vector as (x_1, x_2 , 1) by adding the third coordinate of bias = 1 in support vector.



(2, 2, 1)
(1, 1, 1)

(c) [2 Points] Considering *Two* augmented support vectors where s_1 is for positive class (+1) and s_2 is for negative class (-1), we can use the following two equations to learn a classification function:

$$\left(\alpha_1 \bullet \Phi_1(s_1) + \alpha_2 \bullet \Phi_1(s_2)\right)^T \bullet \Phi_1(s_1) = y_1 = 1$$

$$\left(\alpha_1 \bullet \Phi_1(s_1) + \alpha_2 \bullet \Phi_1(s_2)\right)^T \bullet \Phi_1(s_2) = y_2 = -1$$

Based on the two augmented support vectors obtained from part (b), compute the two parameters α_1 and α_2.

$$\phi_1(s_1) = \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}, \quad \phi_1(s_2) = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\begin{cases} \left(\alpha_1 \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix} + \alpha_2 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\right)^T \cdot \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix} = 1 \\ \left(\alpha_1 \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix} + \alpha_2 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\right)^T \cdot \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = -1 \end{cases}$$

$$\begin{cases} [2\alpha_1 + \alpha_2,\ 2\alpha_1 + \alpha_2,\ \alpha_1 + \alpha_2] \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix} = 1 \\ [2\alpha_1 + \alpha_2,\ 2\alpha_1 + \alpha_2,\ \alpha_1 + \alpha_2] \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = -1 \end{cases}$$

$$\begin{cases} 4\alpha_1 + 2\alpha_2 + 4\alpha_1 + 2\alpha_2 + \alpha_1 + \alpha_2 = 1 \\ 2\alpha_1 + \alpha_2 + 2\alpha_1 + \alpha_2 + \alpha_1 + \alpha_2 = -1 \end{cases}$$

$$\begin{cases} 9\alpha_1 + 5\alpha_2 = 1 \\ 5\alpha_1 + 3\alpha_2 = -1 \end{cases}$$

$$\Rightarrow \begin{cases} \alpha_1 = 4 \\ \alpha_2 = -7 \end{cases}$$

(d) [2 Points] Continuing from part (c), find the hyperplane: $y = wx$ which can be used as the final classifier. Here w can be estimated based on the equation: $w = \sum_i \alpha_i \bullet s_i$, where α_i is estimated from part (c) and s_i denotes an augmented support vector. Provide the estimated parameters w.

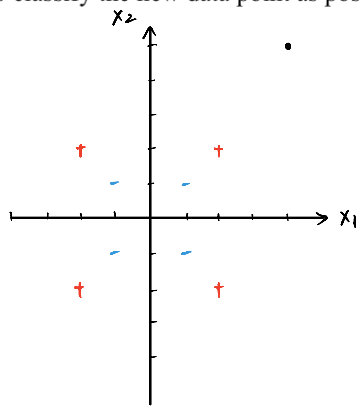
$$w = \sum_i \alpha_i \cdot s_i = \alpha_1 \cdot s_1 + \alpha_2 \cdot s_2$$

$$= 4 \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix} - 7 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 8-7 \\ 8-7 \\ 4-7 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 1 \\ 1 \\ -3 \end{bmatrix}$$

(e) [2 Points] Given a new data point (4, 5), using the kernel function and the hyperplane learned in part (d) to classify the new data point as positive or negative.



The new data is positive.