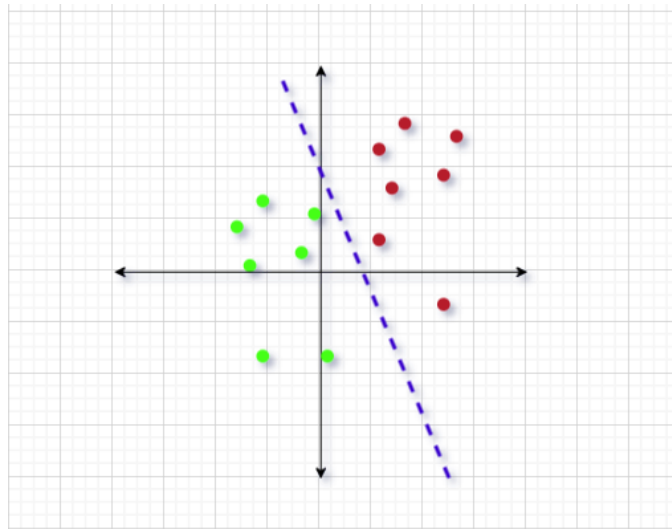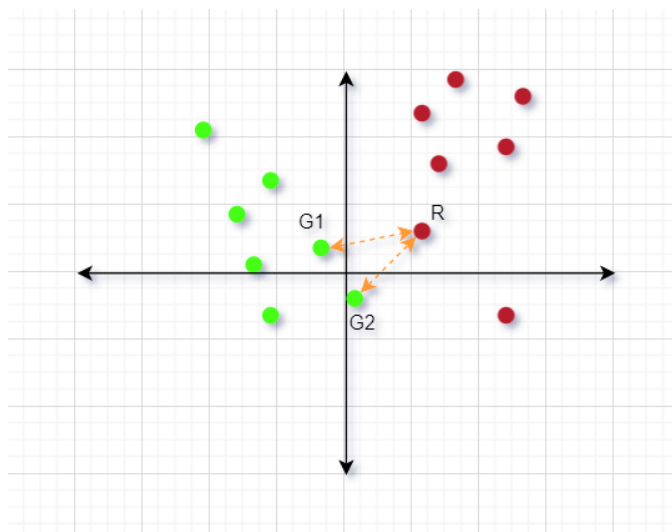# L14_15 SVM Theory

Created by　J　Junaid Mahmud

$$SVM(Support\ Vector\ Machine)\ Theory$$

We will draw a line that will separate red and green points. These are two classes of a dataset.

- We will first calculate the distance between the **negative class and the positive class.**

- Here, the points G1, G2 and R are the closest for the two classes.

- For SVM, we consider the points of different closed classes to draw a boundary using them. These points are called **support vectors**.

- Here:

  - G1 ⇒ S1

  - G2 ⇒ S2

  - R ⇒ S3

Theoretical Steps:

- Let's say the coordinates of the support vectors are:

$$S_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \quad S_2 = \begin{pmatrix} 2 \\ -1 \end{pmatrix} \quad S_3 = \begin{pmatrix} 4 \\ 0 \end{pmatrix}$$

- We will augment a 1 at the end for each vector.

$$\bar{S}_1 = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} \quad \bar{S}_2 = \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} \quad \bar{S}_3 = \begin{pmatrix} 4 \\ 0 \\ 1 \end{pmatrix}$$

- For 3 support vectors, we will have 3 parameters.

$$\alpha_1 \quad \alpha_2 \quad \alpha_3$$

- We will do the following calculations, using the matrix calculation:

$$\bar{S}\bar{S}^T \alpha$$

$$Where, \quad \bar{S} = \begin{pmatrix} \bar{S}_1 \\ \bar{S}_2 \\ \bar{S}_3 \end{pmatrix} \quad and \ \ \bar{S}^T = (\bar{S}_1 \ \bar{S}_2 \ \bar{S}_3)$$

- The result will be as follows:

$$\widetilde{S}\widetilde{S}^T = \begin{bmatrix} \widetilde{S}_1\widetilde{S}_1 & \widetilde{S}_1\widetilde{S}_2 & \widetilde{S}_1\widetilde{S}_3 \\ \widetilde{S}_2\widetilde{S}_1 & \widetilde{S}_2\widetilde{S}_2 & \widetilde{S}_2\widetilde{S}_3 \\ \widetilde{S}_3\widetilde{S}_1 & \widetilde{S}_3\widetilde{S}_2 & \widetilde{S}_3\widetilde{S}_3 \end{bmatrix}$$

(3×3)

- If the aforementioned is multiplied by alpha, then we will get the following:

$$\alpha_1\bar{S}_1\bar{S}_1 + \alpha_2\bar{S}_1\bar{S}_2 + \alpha_3\bar{S}_1\bar{S}_3 = -1$$
$$\alpha_1\bar{S}_2\bar{S}_1 + \alpha_2\bar{S}_2\bar{S}_2 + \alpha_3\bar{S}_2\bar{S}_3 = -1$$
$$\alpha_1\bar{S}_3\bar{S}_1 + \alpha_2\bar{S}_3\bar{S}_2 + \alpha_3\bar{S}_3\bar{S}_3 = +1$$

- We are assigning S1 and S2(Green) as the **negative class**, and S3 (Red) as the **positive class. That's why we are assigning -1 to the first two and +1 to the third one.**

- After putting the values of the augmented vectors:

- We know that, for matrix calculation:

$$if \quad A = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \quad and \quad B = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$$
$$then,$$
$$A.B = a_1b_1 + a_2b_2 + a_3b_3$$

- Finally, after calculation, we have the following 3 equations:

$$6\alpha_1 \; + \; 4\alpha_2 \; + \; 9\alpha_3 \; = \; -1 \; -----(i)$$
$$4\alpha_1 \; + \; 6\alpha_2 \; + \; 9\alpha_3 \; = \; -1 \; -----(ii)$$
$$9\alpha_1 \; + \; 9\alpha_2 \; + \; 17\alpha_3 \; = \; 1 \; -----(iii)$$

- Now we solve for the alpha values and get the following:

$$\alpha_1 = -3.25$$
$$\alpha_2 = -3.25$$
$$\alpha_3 = 3.5$$

⇒ Since we are working with **Linear SVM**, the **decision boundary will be linear. Hence**

$$y \; = \; mx \; + \; c$$

- From the machine learning perspective, the equation becomes:

$$y \; = \; WX \; + \; b$$

- We have to **maximize** the margin D. Because we want to separate the classes using the decision boundary. If it's close to the classes, then there is a chance of overlap. So the distance of the **decision boundary** from the classes should be **as high as** possible.
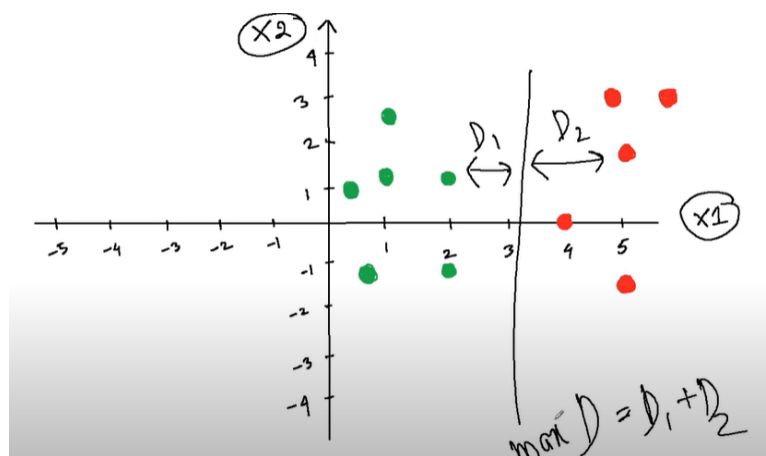
- Finally, the value of Weight is as follows:

$$\bar{W} \; = \; \sum \alpha_i \, \bar{S}_i$$

$$\bar{W} = \alpha_1 \bar{S}_1 + \alpha_2 \bar{S}_2 + \alpha_3 \bar{S}_3$$

$$= -3.25 \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} + -3.25 \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} + -3.5 \begin{pmatrix} 4 \\ 0 \\ 1 \end{pmatrix}$$

$$= \begin{pmatrix} -6.50 \\ -3.25 \\ -3.25 \end{pmatrix} + \begin{pmatrix} -6.50 \\ 3.25 \\ -3.25 \end{pmatrix} + \begin{pmatrix} 14.0 \\ 0 \\ 3.5 \end{pmatrix}$$
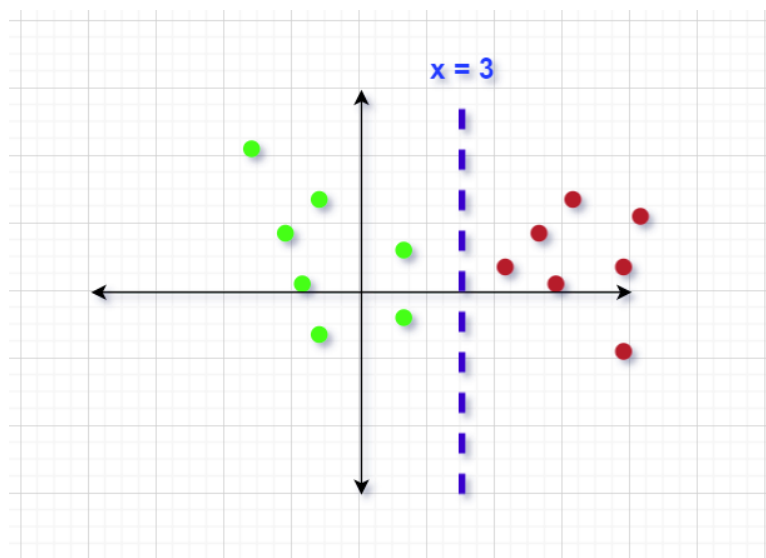
$$\bar{W} = \begin{pmatrix} 1 \\ 0 \\ 3 \end{pmatrix}$$

- This is the **augmented W**. The last value -3 was a bias. So finally:

$$\bar{W} = \begin{pmatrix} 1 \\ 3 \\ -3 \end{pmatrix} \quad So, \quad W = \begin{pmatrix} 1 \\ 3 \end{pmatrix} \quad and \quad b = -3$$

- The equation of the decision boundary becomes:

$$y = W\ x\ -\ 3$$
$$=> y = \begin{pmatrix} 1 \\ 0 \end{pmatrix} x\ -\ 3$$

- As b = -3, **x = 3, will be positive**. If b were positive, x would be negative. The decision boundary/hyperplane will be:

- For a new point (1, 0), the following will be the calculation to get the class:

$$y = \begin{pmatrix} 1 \\ 0 \end{pmatrix} x - 3$$

$$y = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} - 3$$

$$= (1 + 0) - 3$$

$$= 1 - 3$$

$$So, \ y = -2$$

- As the value is negative, the new point belongs to the **negative class.** If the result were positive, the point would belong to the positive class.

Linear SVM also works with non-linear data, which is a great strength.