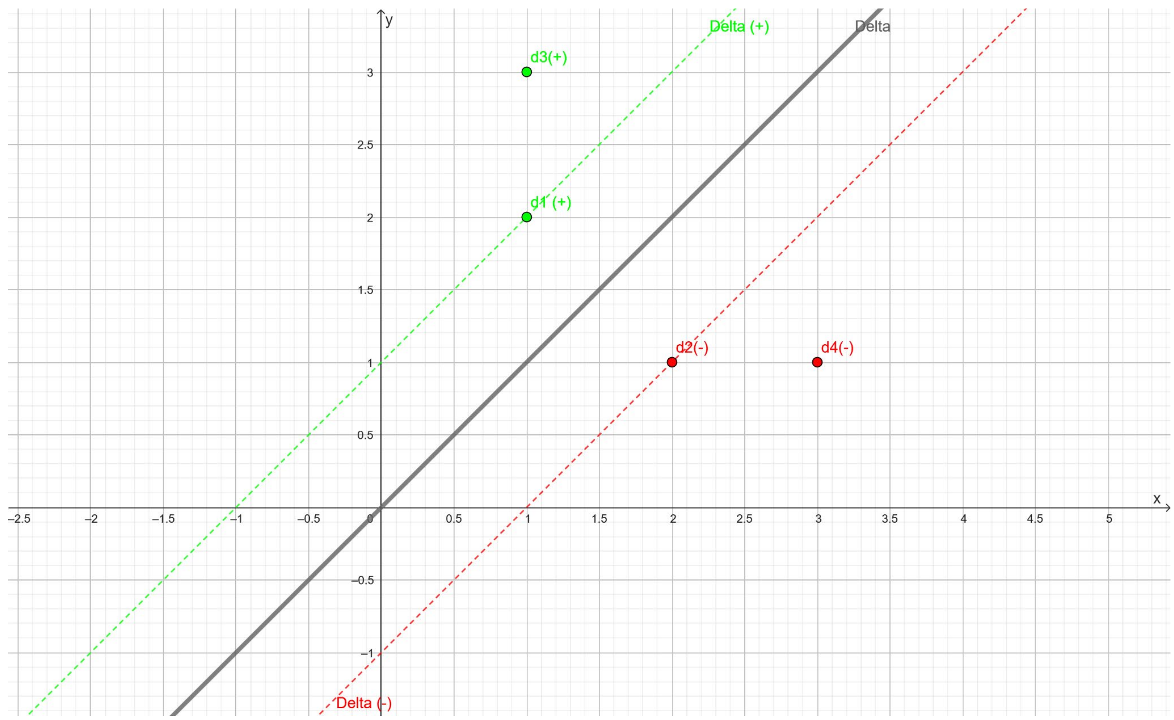
Exercise SVM:

Consider the training dataset for binary classification, consisting of the features x, y and class labels:

x	y	Class
1	2	+
2	1	-
1	3	+
3	1	-

1) Plot the dataset along with the optimal separator line Δ, along with Δ_+ and Δ_-, the two lines parallel to Δ, closest to the positive and negative examples respectively.



The optimal separator line is the line with the maximum margin.

1) Given that the equation of the three separating lines are:

   Δ: $w_1x + w_2y + b = 0$, Δ_+: $w_1x + w_2y + b = +a$ and Δ_-: $w_1x + w_2y + b = -a$.

   Express the margin between the Δ+ and Δ- in terms of the line parameters w1, w2 and b and the offset a.

Take a point $M_+ \in \Delta_+$, the margin is 2*Distance(M^+,Δ).

$Distance(M^+,\Delta) = \frac{|w_1(M_+.x)+w_2(M_+.y)+b|}{\sqrt{w_1^2+w_2^2}} = \frac{a}{\sqrt{w_1^2+w_2^2}}$ **because $M_+ \in \Delta_+$.**

$$Margin = \frac{2a}{\sqrt{w_1^2+w_2^2}}$$

2) Prove that scaling the equations of (Δ, Δ_+, Δ_-) by a nonzero positive factor does not change the separating lines or margin.

Scaling the equations by a nonzero factor does not change the separating lines or the margin.

a. **Scaling the three equations does not change them. This is because if a point satisfies the original line equation, it will also satisfy the scaled equation.**

b. **The margin stays also the same because scaling by a factor k will scale all coefficients (w_1, w_2) by k, and when you compute the margin, the k cancel out.**

$$Margin = \frac{2a'}{\sqrt{w'_1{}^2+w'_2{}^2}} = \frac{2(\frac{a}{k})}{\sqrt{(\frac{w_1}{k})^2+(\frac{w_2}{k})^2}} = \frac{2(\frac{a}{k})}{\sqrt{(\frac{w_1}{k})^2+(\frac{w_2}{k})^2}} = \frac{(\frac{1}{k})\,2a}{(\frac{1}{k})\sqrt{w_1{}^2+w_2{}^2}} = \frac{2a}{\sqrt{w_1{}^2+w_2{}^2}}$$

3) Examine the following equations representing the separating lines Δ: $w_1x + w_2y + b = 0$, Δ_+: $w_1x + w_2y + b = +a$ and Δ_-: $w_1x + w_2y + b = -a$. Does setting the offset $a = 1$ result in any loss of generality? Explain.

Setting a=1 does not lose generality; it is akin to scaling by a positive factor of k=1/a

Δ: $\frac{w_1}{\boldsymbol{a}}x + \frac{w_2}{\boldsymbol{a}}y + \frac{\mathrm{b}}{\boldsymbol{a}} = 0$ which will become Δ: $w_1{}'x + w_2{}'y + b' = 0$

Δ_+: $\frac{w_1}{\boldsymbol{a}}x + \frac{w_2}{\boldsymbol{a}}y + \frac{\mathrm{b}}{\boldsymbol{a}} = +1$ which will become Δ_+: $w_1{}'x + w_2{}'y + b' = +1$

Δ_-: $\frac{w_1}{\boldsymbol{a}}x + \frac{w_2}{\boldsymbol{a}}y + \frac{\mathrm{b}}{\boldsymbol{a}} = -1$ which will become Δ_-: $w_1'x + w_2'y + b' = -1$

$$Margin = \frac{2}{\sqrt{w'_1{}^2 + w'_2{}^2}}$$

In essence, all separators with parameters (k * w1, k * w2, k * b) and offset of (k*a) are equivalent. By fixing a = 1, we simplify the representation without changing relative positions or the margin.

4) To determine the maximum margin separator, present the quadratic function that needs to be minimized along with all the constraints on the parameters (w_1, w_2, b).

a. **The quadratic function to be minimized is the square of the reciprocal of the margin:**

$f(w_1, w_2, b) = (\frac{1}{Margin})^2 = \frac{1}{4}(w_1{}^2 + w_2{}^2)$

Minimizing this function is equivalent to minimizing: $g(w_1, w_2, b) = 2 * f(w_1, w_2, b) = \frac{1}{2}(w_1{}^2 + w_2{}^2)$

Function g replaces f for mathematical simplicity in the optimization process.

b. **All the positive data points should be above Δ_+ and all the negative data points should be bellow Δ_- . This give us the following 4 constraints:**

$w_1(d_1.x) + w_2(d_1.y) + b \geq 1$ **(The positive data point d_1 is above Δ_+)**

$w_1(d_3.x) + w_2(d_3.y) + b \geq 1$ **(The positive data point d_3 is above Δ_+)**

$w_1(d_2.x) + w_2(d_2.y) + b \leq -1$ **(The positive data point d_2 is bellow Δ_-)**

$w_1(d_4.x) + w_2(d_4.y) + b \leq -1$ **(The positive data point d_4 is bellow Δ_-)**

By substituting the coordinates:

$w_1 + 2w_2 + b \geq 1$

$w_1 + 3w_2 + b \geq 1$

$2w_1 + w_2 + b \leq -1$

$3w_1 + w_2 + b \leq -1$

5) Assuming the bias term b is set to zero, sketch the feasible region in the (w_1, w_2) space satisfying the constraints of the previous question.

if b=0, we get the following constraints:

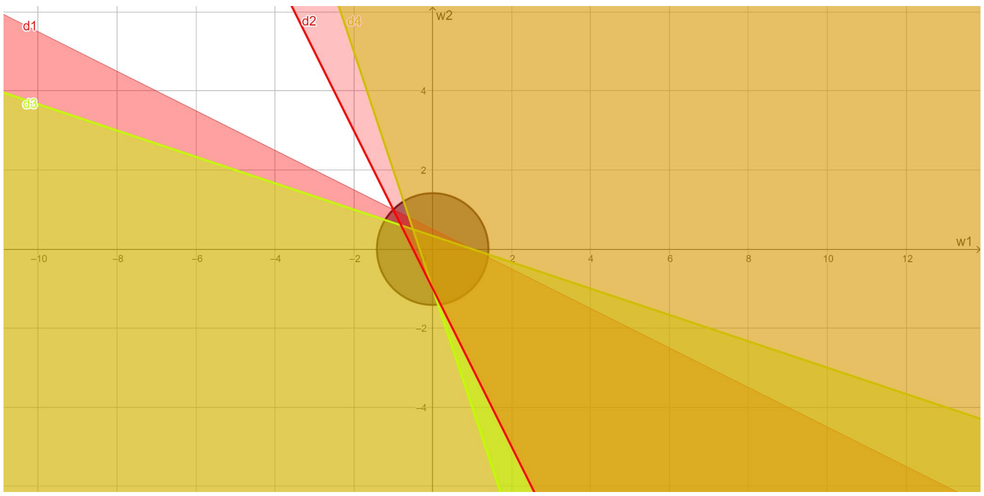
$w_1 + 2w_2 \geq 1$

$w_1 + 3w_2 \geq 1$

$2w_1 + w_2 \leq -1$

$3w_1 + w_2 \leq -1$

6) Visually estimate the values of (w_1, w_2) that maximize the margin between the two classes.



From the graph, we observe the feasible region as an unbounded white space. To minimize the objective function, we need to consider the corner of the triangle where it intersects with the circle defined by $\frac{1}{2}({w_1}^2 + {w_2}^2) = 1$. This intersection occurs at $(w_1 = -1, w_2 = 1, b = 0)$.

7) Using the feasible region, explain graphically why certain data points do not contribute to determining the maximum margin separator.

The data points d_3 and d_4 do not influence the feasible region as their lines lie outside of it. On the other hand, d_1 and d_2 define the feasible region and play a crucial role in determining the solution. These points are called support vectors.