ex4 ass 1

Task 1 Plot the six training points

0.1 Task 1: Plot the six training points

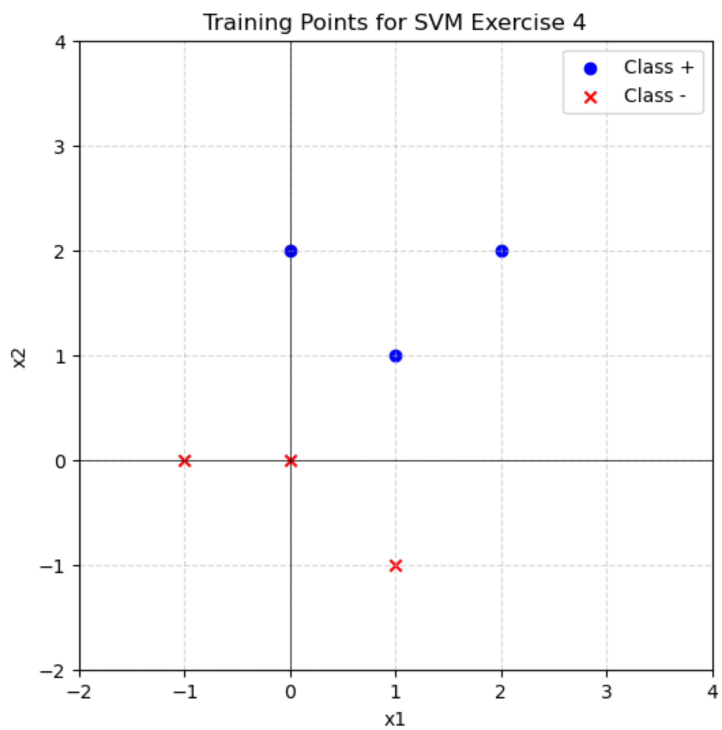

Figure 1:

Task 1 Are the classes linearly separable?

By observing the plot it looks like the two classes are linearly separable

Task 2: Construct the weight vector of the maximum margin hyperplane by inspection and identify the support vectors.

$$\mathbf{w}^T\phi + w_0 = 0$$

where
$\mathbf{w}$ is the weight vector (normal to the Hyperplane)
w_0 is the bias term
ϕ is a point in the feature space
Hyperplane's equation

$$\mathbf{w}^T\phi + w_0 = 0$$

where:
$\mathbf{w}$ is the weight vector (normal to the Hyperplane)
w_0 is the bias term
ϕ is a point in the feature space

$$|y(\phi)| = |\mathbf{w}^T\phi + w_0| \geq 1,$$

i.e.
Normalized equation

$$|y(\phi)| = |\mathbf{w}^T\phi + w_0| \geq 1,$$

i.e.

$$\text{class + (blue):} \quad \text{if} \quad \mathbf{w}^T\phi + w_0 \geq 1$$
$$\text{class - (red):} \quad \text{if} \quad \mathbf{w}^T\phi + w_0 \leq -1$$

Let ϕ^- be the closest point on the "minus" margin, thus $\phi^- = (0,0)$

$$\mathbf{w}^T\phi + w_0 = -1$$
$$\mathbf{w}^T(0,0) + w_0 = -1$$
$$\Rightarrow w_0 = -1$$

Let ϕ^+ be the closest point on the "plus" margin, thus $\phi^+ = (1,1)$

$$\mathbf{w}^T\phi + w_0 = 1$$
$$\mathbf{w}^T(1,1) + w_0 = 1$$
$$\Rightarrow w_1 + w_2 + w_0 = 1$$

with $w_0 = -1$ and equal weights $\Rightarrow w_1 = 1$ and $w_2 = 1$

Then,$\phi^+ = \phi^- + \lambda\mathbf{w}$ for a scalar λ. Find $\lambda(1,1) = (0,0) + \lambda(1,1) \Rightarrow \lambda = 1$

Margin

$$M = ||w|| = 1\sqrt{(1)^2 + (1)^2} = \sqrt{2}$$

Margin

$$M = ||w|| = 1\sqrt{(1)^2 + (1)^2} = \sqrt{2}$$

Support Vectors

$$\phi^+ = \begin{bmatrix} \mathbf{1} \\ \mathbf{1} \end{bmatrix} \qquad \phi^- = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}$$

Support Vectors

$$\phi^+ = \begin{bmatrix} \mathbf{1} \\ \mathbf{1} \end{bmatrix} \qquad \phi^- = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}$$

Weight vector of the maximum margin

$$w = \begin{bmatrix} \mathbf{1} \\ \mathbf{1} \end{bmatrix} \qquad w_0 = -1$$

Weight vector of the maximum margin

$$w = \begin{bmatrix} \mathbf{1} \\ \mathbf{1} \end{bmatrix} \qquad w_0 = -1$$

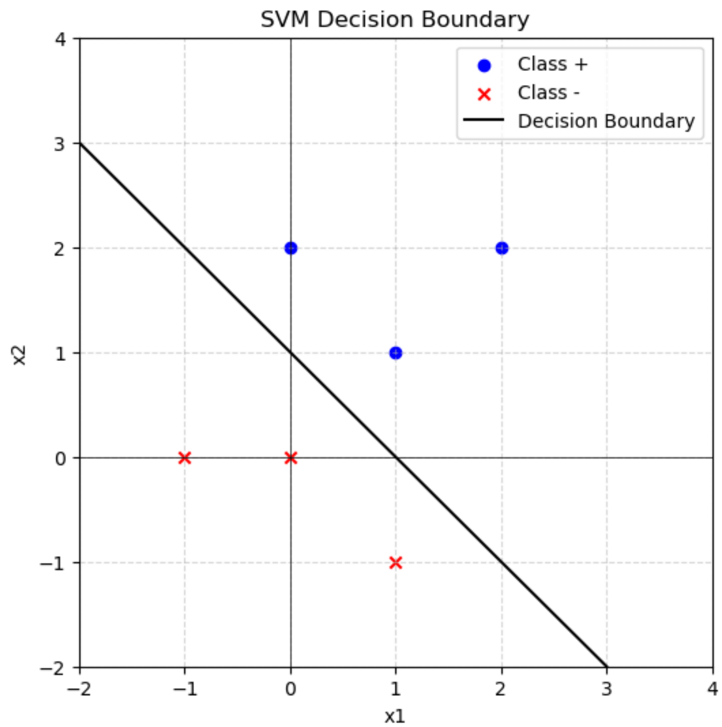

Figure 2:

Task 3 If you remove one of the support vectors, does the size of the optimal margin decrease, stay the same, or increase?

If I remove one support vector, the size of the optimal margin stays the same since there are in both classes other points that have the same distance from the hyperplane like our current support vectors. For example points $(0,2)$ for class $+$ and $(1,-1)$ for class $-$.

Task 3: If you remove one of the support vectors, does the size of the optimal margin decrease, stay the same, or increase?

If I remove one support vector, the size of the optimal margin stays the same since there are in both classes other points that have the same distance from the hyperplane like our current support vectors. For example points $(0,2)$ for class $+$ and $(1,-1)$ for class $-$.

Task 4 Is your answer to (3) also true for any dataset? Provide a counterexample or give a short proof.

No, my answer isn't true for any dataset.

Let's take the dataset from the task before and modify it such that there are no more the points (0,2) and (1,1). The Support Vectors are still the same as before

$$\phi^+ = \begin{bmatrix} \mathbf{1} \\ \mathbf{1} \end{bmatrix} \qquad \phi^- = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}$$

Thus the weight vector and the optimal margin remain the same.

The new Data Set is:

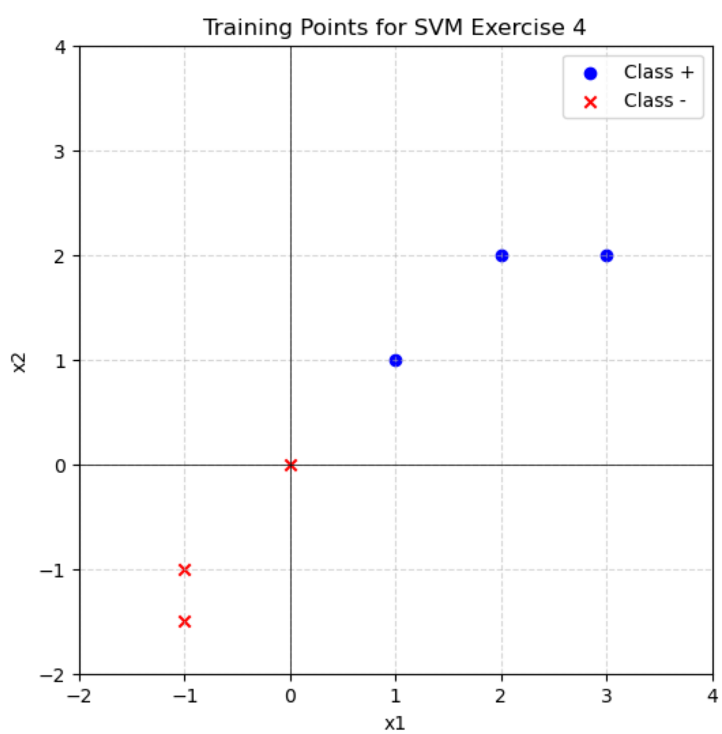

Figure 3:

Now we remove the Support Vectors:

$$\phi^+ = \begin{bmatrix} \mathbf{1} \\ \mathbf{1} \end{bmatrix} \qquad \phi^- = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}$$

from the dataset and we recompute the maximum margin hyperplane.

By inspection we can observe that the closest point of the two classes are now (-1,-1) and (2,2)

Let ϕ^- be the closest point on the "minus" margin, thus $\phi^- = (-1, -1)$

$$\mathbf{w}^T \phi + w_0 = -1$$

$$\mathbf{w}^T (-1, -1) + w_0 = -1$$

$$\Rightarrow -w_1 - w_2 + w_0 = -1$$

Let ϕ^+ be the closest point on the "plus" margin, thus $\phi^+ = (2, 2)$

$$\mathbf{w}^T \phi + w_0 = 1$$

$$\mathbf{w}^T (2, 2) + w_0 = 1$$

$$\Rightarrow 2w_1 + 2w_2 + w_0 = 1$$

With two equations:

$$-w_1 - w_2 + w_0 = -1 \quad (1)$$

$$2w_1 + 2w_2 + w_0 = 1 \quad (2)$$

Solve: From (1):

$$w_0 = w_1 + w_2 - 1.$$

Substitute in (2):

$$2w_1 + 2w_2 + (w_1 + w_2 - 1) = 1$$

$$3w_1 + 3w_2 = 2$$

Assuming equal weights $w_1 = w_2$, substitute:

$$3w_1 + 3w_1 = 2 \Rightarrow 6w_1 = 2 \Rightarrow w_1 = \frac{1}{3}, w_2 = \frac{1}{3}$$

$$w_0 = \frac{1}{3} + \frac{1}{3} - 1 = -\frac{1}{3}$$

Recalculate Margin

Then,

$$\lambda = \frac{2}{w^T w}$$

Find λ

$$\lambda = \frac{2}{\begin{bmatrix} \mathbf{\frac{1}{3}} & \mathbf{\frac{1}{3}} \end{bmatrix} \begin{bmatrix} \mathbf{\frac{1}{3}} \\ \mathbf{\frac{1}{3}} \end{bmatrix}} \qquad \Rightarrow \lambda = \frac{2}{\frac{1}{9} + \frac{1}{9}} = \frac{2}{\frac{2}{9}} = 9.$$

Margin

$$M = ||\lambda \mathbf{w}|| = 9\sqrt{\left(\frac{1}{3}\right)^2 + \left(\frac{1}{3}\right)^2} = 9 * \frac{\sqrt{2}}{3} = 3\sqrt{2} \approx 4.24.$$

Support Vectors

$$\phi^+ = \begin{bmatrix} \mathbf{2} \\ \mathbf{2} \end{bmatrix} \qquad \phi^- = \begin{bmatrix} \mathbf{-1} \\ \mathbf{-1} \end{bmatrix}$$

Weight Vector of the Maximum Margin

$$\mathbf{w} = \begin{bmatrix} \mathbf{\frac{1}{3}} \\ \mathbf{\frac{1}{3}} \end{bmatrix} \qquad w_0 = -\frac{1}{3}$$

Conclusion: If before the margin was $\sqrt{2}$, now removing the old support vectors the margin has increased to 4.24 because there aren't any data points with the same distance between them as the ones before.