# Week 10 | Solve With Instructors

Q1) Select the correct statements about Support Vector Machine algorithm

(A) All support vectors lie on the supporting hyperplanes.

(B) Every point on the supporting hyperplanes is a support vector.

(C) $\mathbf{w}$ is a dual variable in the optimization problem $\min_{\mathbf{w}} \max_{\alpha \geqslant 0} \frac{\|\mathbf{w}\|^2}{2} + \sum_{i=1}^{n} \alpha_i \left[1 - \left(\mathbf{w}^T \mathbf{x}_i\right) y_i\right]$

(D) The search space has dimension $d$ for the primal problem.

(E) Large margin classifiers are capable of generalizing better to unseen data

(F) The weight vector can now be seen as a sparse linear combination of the data-points.

(G) Support vectors are the closest data points to the boundary.

**Solution:**

1. All support vectors lie on the supporting hyperplanes.
   - **Correct**. Support vectors are the data points that lie on the margin boundaries or the hyperplanes, and these are crucial for defining the optimal hyperplane.

2. Every point on the supporting hyperplanes is a support vector.
   - **Incorrect**. The support vectors are specifically the data points that lie on the margin, which is defined by the supporting hyperplanes. But it is not necessary that each point lying on the supporting hyperplane is a support vector.

3. $\mathbf{w}$ is a dual variable in the optimization problem
   - **Incorrect**. In the dual formulation of SVM, the dual variables are denoted as $\boldsymbol{\alpha}$, not $\mathbf{w}$. The vector $\mathbf{w}$ represents the weight vector in the primal problem, and it's used to define the decision boundary. The dual formulation involves maximizing the Lagrange multipliers $\boldsymbol{\alpha}_i$.

4. The search space has dimension $d$ for the primal problem.
   - **Correct**. In the primal form of SVM, the weight vector $\mathbf{w}$ resides in a space of dimension $d$, where $d$ is the number of features in the dataset. This space is where the optimization is performed to find the optimal hyperplane.

5. Large margin classifiers are capable of generalizing better to unseen data.
   - **Correct**. One of the key advantages of SVM is that it maximizes the margin between the classes, which helps in generalizing better to unseen data. A large margin indicates less overfitting, resulting in better generalization.

6. The weight vector can now be seen as a sparse linear combination of the data points.
   - Correct. In the dual formulation, the weight vector $\mathbf{w}$ can be expressed as a sparse linear combination of the support vectors. This means the majority of the training data do not contribute to the decision boundary, and only support vectors are important.

7. Support vectors are the closest data points to the boundary.
   - **Correct**. Support vectors are the closest points to the decision boundary (or margin). These points determine the position of the hyperplane and are critical to the SVM's optimization.

Q2) For the dataset given below

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 3 & 3 & -1 & 0 & -3 & -3 \\ 0 & -1 & 2 & 0 & 0 & 1 & -2 & 0 \end{bmatrix}$$

$$\mathbf{y} = \begin{bmatrix} 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \end{bmatrix}^T$$

If the solution to dual problem $\alpha^*$ is $\begin{bmatrix} 0.42 \\ 0.57 \\ 0 \\ 0 \\ 0 \\ 0.81 \\ 0.19 \\ 0 \end{bmatrix}$, then

i) How many support vectors are there?

ii) Find the label of the test data point $\begin{bmatrix} 3 \\ -1 \end{bmatrix}$.

**Solution:**

i) As there are $4$ $\alpha_i$'s which are greater than $0$, therefore there are $4$ support vectors.

ii) The weight vector would be

$$\mathbf{w} = \sum_{i=1}^{n} \alpha_i \mathbf{x}_i y_i$$

$$\begin{aligned}\mathbf{w} &= 0.42\begin{bmatrix}1\\0\end{bmatrix} + 0.57\begin{bmatrix}0\\-1\end{bmatrix} + 0\begin{bmatrix}3\\2\end{bmatrix} + 0\begin{bmatrix}3\\0\end{bmatrix} + 0\begin{bmatrix}-1\\0\end{bmatrix} + 0.81\begin{bmatrix}0\\1\end{bmatrix} + 0.19\begin{bmatrix}-3\\-2\end{bmatrix} + 0\begin{bmatrix}-3\\0\end{bmatrix}\\ &= \begin{bmatrix}-0.15\\-0.14\end{bmatrix}\end{aligned}$$

We classify label as $+1$ if $\mathbf{w}^T\mathbf{x} \geqslant 0$, else $0$.

$$\begin{bmatrix}-0.15 & -0.14\end{bmatrix}\begin{bmatrix}3\\-1\end{bmatrix} = -0.31$$

Hence, the label of data point is $-1$.

Q3) Given the vector $\mathbf{w} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, calculate the distance between the line $\mathbf{w}^T\mathbf{x} = 3$ and $\mathbf{w}^T\mathbf{x} = -5$.

**Solution:**

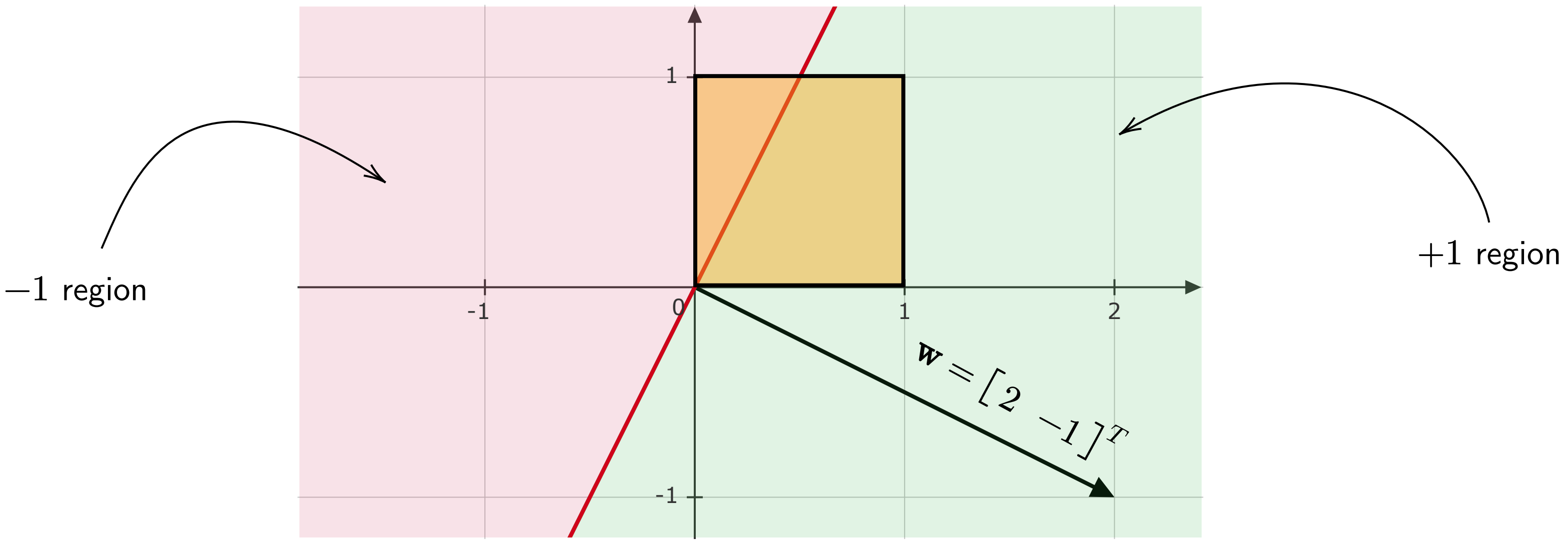
The distance between two hyperplane $\mathbf{w}^T\mathbf{x} = c_1$ and $\mathbf{w}^T\mathbf{x} = c_2$ is given by

$$\frac{|c_1 - c_2|}{\|w\|}$$

$$\frac{|3 - (-5)|}{\sqrt{1^2 + 1^2}} = \frac{8}{\sqrt{2}} \approx 5.65$$

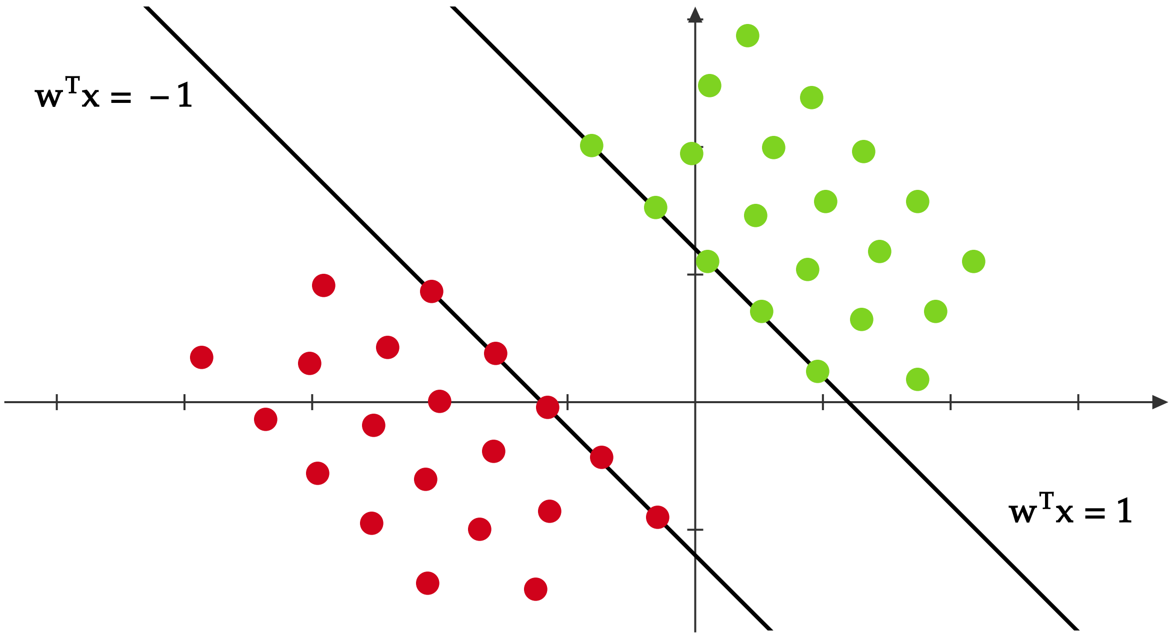
Q4) A SVM has been trained on 2D problem. The feature vector is $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$. It has the following weight vector $\mathbf{w} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$. Labels setup are $+1$ and $-1$. Consider a unit square bounded between the points $\{(0,0),\ (1,0),(0,1),(1,1)\}$. What is the probability that a randomly selected data point from the square region will be selected as $+1$.

**Solution:**




The decision boundary intersect the square at $(0.5, 1)$. So, the area of red part in square is $\dfrac{1}{2} \times 1 \times \dfrac{1}{2} = \dfrac{1}{4}$. So the area of green is $1 - \dfrac{1}{4} = \dfrac{3}{4}$.

$Q5$) Consider the following supporting hyperplanes.




We know that $w$ is is given by.

$$w = \sum_{i=1}^{n} \alpha_i (x_i) y_i$$

For how many data points in the given picture, is $\alpha_i \neq 0$.

A. 5
B. 10

C. 0
D. n
E. Cannot conclude with the following information

**ANSWER** : $E$

The complementary slackness condition is as follows

$$\alpha_i\left(1-\left(w^T x_i\right)y_i\right) = 0\ ,\ \forall\ i$$

Here there are 10 data points satisfying $1-\left(w^T x_i\right)y_i = 0$. However, this does not guarentee $\alpha_i \neq 0$.

In simple words, if $\alpha_i \neq 0,$ then $\left(w^T x_i\right)y_i = 1$ but, $\left(w^T x_i\right)y_i = 1$ does not guarentee $\alpha_i \neq 0$.

$Q6)$ Consider the following formulation for $x = (x_1, x_2)$

$$\min_{(x_1,x_2)} \left[3(x_1-2)^2 - 4(x_2+5)^2\right]$$

$$st. \ ||x||^2 \leqslant 5$$

Find the Lagrangian function for this formulation

A. $L(x, \alpha) = \left[3(x_1-2)^2 - 4(x_2+5)^2\right] + \alpha\left(5 - ||x||^2\right)$
B. $L(x, \alpha) = \left[3(x_1-2)^2 - 4(x_2+5)^2\right] + \alpha\left(5 + ||x||^2\right)$
C. $L(x, \alpha) = \left[3(x_1-2)^2 - 4(x_2+5)^2\right] - \alpha\left(5 - ||x||^2\right)$
D. $L(x, \alpha) = \left[3(x_1-2)^2 - 4(x_2+5)^2\right] - \alpha\left(5 + ||x||^2\right)$

**ANSWER** : $C$

For a formultion of the form

$$min_x \; f(x)$$

$$st. \; g(x) \leqslant 0$$

the Lagrangian function is given by

$$L(x,\alpha) = f(x) + \alpha \times g(x)$$

Here, the objective function is

$$f(x,y) = \left[3(x_1-2)^2 - 4(x_2+5)^2\right]$$

and the constraint function is

$$g(x,y) = ||x||^2 - 5$$

Therefore, the Lagrangian function is given by

$$L(x,\alpha) = \left[3(x_1-2)^2 - 4(x_2+5)^2\right] + \alpha\left(||x||^2 - 5\right)$$

This can also be writtern as

$$L(x,\alpha) = \left[3(x_1-2)^2 - 4(x_2+5)^2\right] - \alpha\left(5 - ||x||^2\right)$$

$Q7$) The primal formulation for the previous question is given by :

$$min_x[max_{\alpha \geqslant 0}(L(x, \alpha))]$$

Is the following true?

$$min_x[max_{\alpha \geqslant 0}(L(x, \alpha))] \equiv max_{\alpha \geqslant 0}[min_x(L(x, \alpha))]$$

**ANSWER** : YES

We can say that the following is true because, $f(x) = \left[3(x_1 - 2)^2 - 4(x_2 + 5)^2\right]$ and $g(x, y) = ||x||^2 - 5$ are convex functions.

$$min_x[max_{\alpha \geqslant 0}(L(x, \alpha))] \equiv max_{\alpha \geqslant 0}[min_x(L(x, \alpha))]$$

Q8) Predict the label for $(3, 2)$ assuming $\alpha = \begin{bmatrix} 0, 3 & 1 & 0 & 2 & 0 & 5 & 0 \end{bmatrix}^T$

$$X = \begin{bmatrix} -1 & 0.25 & -1 & 1 & 0.4 & 0.5 & 0 & 0.5 \\ 0 & 0.5 & 1 & 0 & 0.2 & 1 & -1 & -3 \end{bmatrix}$$

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

**ANSWER** : 1

We know that the following is true

$$w = XY\alpha$$

$$w = \begin{bmatrix} -1 & 0.25 & -1 & 1 & 0.4 & 0.5 & 0 & 0.5 \\ 0 & 0.5 & 1 & 0 & 0.2 & 1 & -1 & -3 \end{bmatrix} \begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 0 \\ 3 \\ 1 \\ 0 \\ 2 \\ 0 \\ 5 \\ 0 \end{bmatrix}$$

$$w = \begin{bmatrix} -1 & 0.25 & -1 & 1 & 0.4 & 0.5 & 0 & 0.5 \\ 0 & 0.5 & 1 & 0 & 0.2 & 1 & -1 & -3 \end{bmatrix} \begin{bmatrix} 0 \\ 3 \\ -1 \\ 0 \\ 2 \\ 0 \\ -5 \\ 0 \end{bmatrix}$$

$$w = \begin{pmatrix} 0.75 \\ 1.5 \end{pmatrix} + \begin{pmatrix} 1 \\ -1 \end{pmatrix} + \begin{pmatrix} 0.8 \\ 0.4 \end{pmatrix} + \begin{pmatrix} 0 \\ 5 \end{pmatrix}$$

$$w = \begin{pmatrix} 2.55 \\ 5.9 \end{pmatrix}$$

Clearly,

$$\begin{pmatrix} 2.55 \\ 5.9 \end{pmatrix}^T \times \begin{pmatrix} 3 \\ 2 \end{pmatrix} > 0$$

Therefore, the predicted label will be $1$

## Question

Consider a soft-margin Support Vector Machine (SVM) for a binary classification problem with a dataset in a two-dimensional space. The optimization problem for the soft-margin SVM is formulated as:

$$\text{Minimize: } \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i,$$

$$\text{subject to: } y_i(w \cdot x_i + b) \geq 1 - \xi_i \text{ and } \xi_i \geq 0 \text{ for all } i.$$

Where $C$ is a positive constant.

Which of the following statements about the soft-margin SVM is **correct**?

1. When $C = 0$, the optimal value of the objective function of the soft-margin problem is 0.
2. For a dataset with $n$ data points, there are $2n$ constraints for soft-margin SVM.
3. A smaller value of $C$ allows for a larger margin, potentially leading to less misclassifications on the training data.
4. For a dataset with $n$ data points, there are $n$ constraints for soft-margin SVM.

## Solution

The problem involves a soft-margin Support Vector Machine (SVM) for binary classification. The objective function and constraints are given as:

$$\text{Minimize: } \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i,$$

$$\text{subject to: } y_i(w \cdot x_i + b) \geq 1 - \xi_i \text{ and } \xi_i \geq 0 \text{ for all } i.$$

Here, $C > 0$ is a constant that controls the trade-off between maximizing the margin and minimizing the misclassification errors.

1. *When $C = 0$, the optimal value of the objective function of the soft-margin problem is* $0$*.*

When $C = 0$, the second term of the objective function, $C\sum_{i=1}^{N}\xi_i$, becomes 0. The optimization then focuses solely on minimizing $\frac{1}{2}\|w\|^2$ without any penalty for misclassified points. However, this does not imply that the objective function is always 0, as $\|w\|^2$ depends on the data. This statement is therefore **incorrect**.

2. *For a dataset with $n$ data points, there are $2n$ constraints for soft-margin SVM.*

For each data point, there are two constraints:

- $y_i(w \cdot x_i + b) \geq 1 - \xi_i$
- $\xi_i \geq 0$

Thus, for $n$ data points, there are $2n$ constraints. This statement is **correct**.

3. *A smaller value of $C$ allows for a larger margin, potentially leading to less misclassifications on the training data.*

A smaller value of $C$ reduces the penalty for misclassified points, allowing the model to prioritize a larger margin over correctly classifying every point. This leads to more misclassifications on the training data rather than fewer. Therefore, this statement is **incorrect**.

4. *For a dataset with $n$ data points, there are $n$ constraints for soft-margin SVM.*

As discussed above, there are $2n$ constraints for $n$ data points. Hence, this statement is **incorrect**.

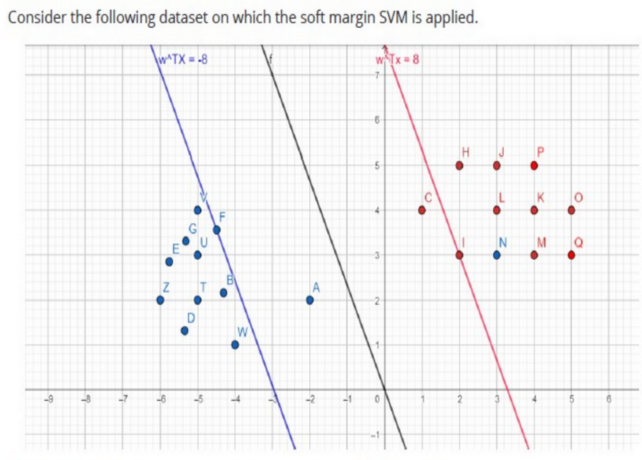



Figure 1: SVM

**Question:**
Consider the following dataset on which the soft margin SVM is applied:
Which of the following statements is/are true about this dataset?
**Options:**

1. Points $F, A, C, I$ are the only support vectors.
2. Points $A, C, N$ are going to be support vectors.
3. Points except $F, A, C, I, N$ do not play any role in determining the optimal weight vector.
4. Points except $F$ and $I$ do not play any role in determining the optimal weight vector.
5. The same dataset can be solved using the hard margin SVM algorithm, and the result would be the same.

**Solution:**
The correct answers are:

- **Option 2:** Points $A, C, N$ are going to be support vectors. ✓
- **Option 3:** Points except $F, A, C, I, N$ do not play any role in determining the optimal weight vector. ✓

Incorrect options:

- **Option 1:** Points $F, A, C, I$ are the only support vectors. **(Incorrect)**
- **Option 4:** Points except $F$ and $I$ do not play any role in determining the optimal weight vector. **(Incorrect)**
- **Option 5:** The same dataset can be solved using the hard margin SVM algorithm, and the result would be the same. **(Incorrect)**

## Explanation

**Statement 1: Points $F, A, C, I$ are the only support vectors.** This statement is **incorrect**. From the diagram:

- For the blue class:
  - $F$: Lies on the margin $w^T x = -8$, so it is a support vector.
  - $A$: Violates the margin constraint (it is on the right side of the decision boundary), so it is a support vector.
  - $I$: Lies inside the margin, so it is a support vector.
- For the red class:
  - $C$: Lies on the margin $w^T x = 8$, so it is a support vector.
  - $N$: Lies inside the margin, so it is a support vector (missing from this statement).

Thus, the support vectors are $F, A, C, I, N$, making this statement incorrect.

**Statement 2: Points $A, C, N$ are going to be support vectors.** This statement is **correct**. The points:

- $A$: Violates the margin.
- $C$: Lies on the red margin $w^T x = 8$.
- $N$: Lies inside the red margin ($w^T x < 8$).

All these points are support vectors.

**Statement 3: Points except $F, A, C, I, N$ do not play any role in determining the optimal weight vector.** This statement is **correct**. In SVM, only support vectors influence the decision boundary and the optimal weight vector $\mathbf{w}$. All other points do not play any role.

**Statement 4: Points except $F$ and $I$ do not play any role in determining the optimal weight vector.** This statement is **incorrect**. The support vectors include $A, C, N$ in addition to $F$ and $I$. These also influence the optimal weight vector $\mathbf{w}$.

**Statement 5: The same dataset can be solved using the hard margin SVM algorithm, and the result would be the same.** This statement is **incorrect**. Hard margin SVM requires that there are no margin violations. However, points like $A$ and $N$ violate the margin constraints, making hard margin SVM infeasible for this dataset.

## Conclusion

The correct statements are:

- **Statement 2:** Points $A, C, N$ are going to be support vectors.
- **Statement 3:** Points except $F, A, C, I, N$ do not play any role in determining the optimal weight vector.