Practice

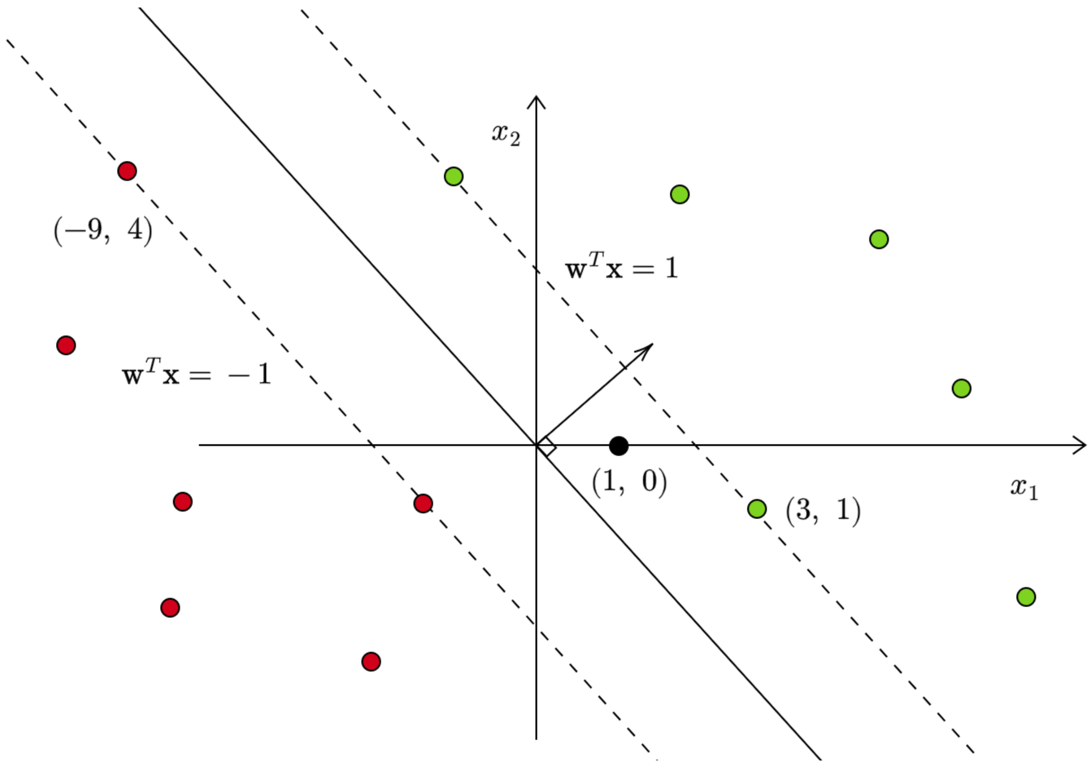
This document has 13 questions.

Common Data for Questions (1) to (6)

Statement

Consider a hard-margin SVM trained on a dataset in $\mathbb{R}^2$ for a binary classification task. Red and green points belong to the training dataset. Red points belong to class -1 and green points belong to class $+1$. The black-point is a test data-point. The dotted lines are the supporting hyperplanes for the SVM.

Note: We don't have access to the test data-point during training; it is given to us *after* the model has been learned on the training dataset.



Question-1

Statement

What is the maximum number of support vectors that the model could have?

Answer

4

Solution

The number of support vectors is upper bounded by the total number of points that lie on the supporting hyperplanes. Recall that support vectors are those points for which $\alpha_i^* > 0$. By complementary slackness, if $\alpha_i^* > 0$ then $(\mathbf{w}^T\mathbf{x}_i)y_i = 1$. So, we can rightfully claim that if $\alpha_i^* > 0$, then that point lies on one of the two supporting hyperplanes.

Now, can we claim that the number of support vectors is exactly equal to 4? This is a bit tricky. The claim we are trying to make here is stronger:

> If a point lies on either of the two supporting hyperplane, then it is a support vector.

Mathematically, this means the following. If $(\mathbf{w}^T\mathbf{x}_i)y_i = 1$, then $\alpha_i^* > 0$. This is something that is not guaranteed by complementary slackness. All that complementary slackness tell us is that for every point in the dataset, $\alpha_i^*[1 - (\mathbf{w}^T\mathbf{x}_i)y_i] = 0$. It could be the case that both $\alpha_i^* = 0$ and $(\mathbf{w}^T\mathbf{x}_i)y_i = 1$. In such a case, the i^{th} point wouldn't qualify as a support vector. To summarize, every support vector lies on one of the two supporting hyperplanes, but every point on the supporting hyperplanes need not be a support vector.

Question-2

Statement

What is the value of the weight vector $\mathbf{w}$? Select all options that are correct.

Options

(a)

$$\mathbf{w} = \begin{bmatrix} 5/21 \\ 2/7 \end{bmatrix}$$

(b)

$$\mathbf{w} = \begin{bmatrix} 5/3 \\ 2 \end{bmatrix}$$

(c)

$$\mathbf{w} = \begin{bmatrix} 7/2 \\ 21/5 \end{bmatrix}$$

(d)

$$\mathbf{w} = \begin{bmatrix} 1/2 \\ 3/5 \end{bmatrix}$$

Answer

(a)

Solution

Let the weight vector be $\mathbf{w} = \begin{bmatrix} w_1 & w_2 \end{bmatrix}^T$. Then, we have:

$$3w_1 + w_2 = 1$$

$$-9w_1 + 4w_2 = -1$$

Solving this system of equations, we get:

$$\mathbf{w} = \begin{bmatrix} 5/21 \\ 2/7 \end{bmatrix}$$

Question-3

Statement

What is the equation of the decision boundary? Select all options that are correct.

Options

(a)

$$\frac{5}{21}x_1 + \frac{2}{7}x_2 = 0$$

(b)

$$\frac{5}{3}x_1 + 2x_2 = 0$$

(c)

$$7x_1 + 5x_2 = 0$$

(d)

$$\frac{1}{2}x_1 + 5x_2 = 0$$

Answer

(a), (b)

Solution

The equation of the solid line (decision boundary) is given by:

$$\frac{5}{21}x_1 + \frac{2}{7}x_2 = 0$$

We can now multiply both sides by 7 to get:

$$\frac{5}{3}x_1 + 2x_2 = 0$$

Note that this kind of scaling cannot be done with the weight vector! Scaling the weight vector alone would result in a different set of supporting hyperplanes. Moreover, the scaled weight vector would not be a solution to the primal problem.

Question-4

Statement

What is the width of the separation between the two supporting hyperplanes? Enter your answer correct to three decimal places.

Note: the exact value of the width is different from the solution to the optimization problem that is discussed in the lecture.

Answer

5.378

Range: [5.36, 5.39]

Solution

The width is given by:

$$\frac{2}{||\mathbf{w}||} = \frac{2}{\sqrt{(5/21)^2 + (2/7)^2}} \approx 5.378$$

Question-5

Statement

What is the predicted label of the black test-point?

Answer

1

Solution

We can infer this either from the graph or we can use:

$$\text{sign}(\mathbf{w}^T\mathbf{x}) = \text{sign}(5/21 \cdot 1 + 2/7 \cdot 0) = 1$$

Question-6

Statement

Is the following statement true or false?

For any test-point that falls within the region bounded by the supporting hyperplanes, no label can be assigned, as it doesn't satisfy the constraints in the optimization problem.

Options

(a)

True

(b)

False

Answer

(b)

Solution

The decision boundary is given by $\mathbf{w}^T\mathbf{x} = 0$. Though the points on the supporting hyperplanes help in determining the value of $\mathbf{w}$, they don't meddle with the prediction of points that fall within them. This depends purely on the sign of $\mathbf{w}^T\mathbf{x}$.

Question-7

Statement

SVM is a ———

Options

(a)

generative model

(b)

discriminative model

Answer

(b)

Solution

SVM is a discriminative model. There are no explicit probabilities involved, but we can present it this way:

$$P(y=1 \mid \mathbf{x}) = \begin{cases} 1, & \mathbf{w}^T\mathbf{x} \geq 0 \\ 0, & \mathbf{w}^T\mathbf{x} < 0 \end{cases}$$

We aren't really concerned about how the point $\mathbf{x}$ was generated. Given a point, we use the line $\mathbf{w}$ to determine its label. Note that this kind of presentation is similar to what we saw for perceptrons.

Question-8

Statement

Study the similarities and differences between the following models:

(1) perceptron

(2) logistic regression

(3) SVM

Solution

Similarities:

- All three are linear models
- As a result, the decision boundary is given by $\mathbf{w}^T\mathbf{x} = 0$ for all three models
- All three are discriminative models.

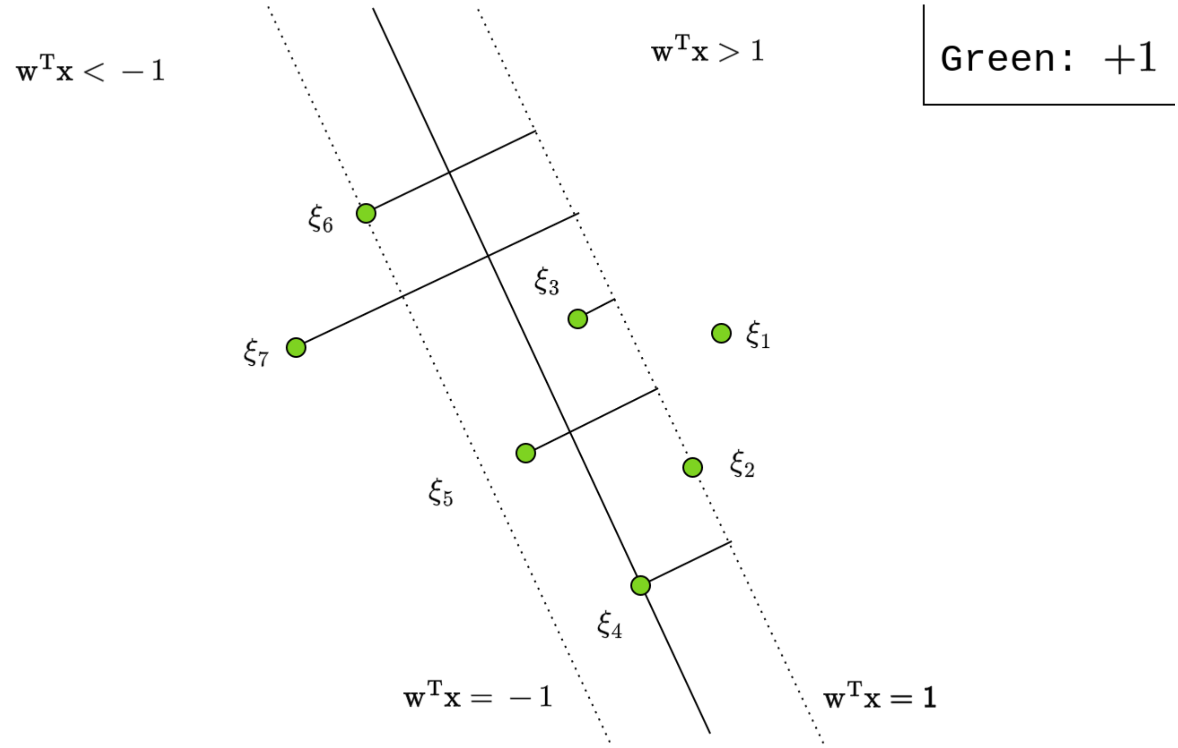
Differences:

- Logistic regression associates an explicit probability for each data-point. Farther apart a point is from the decision boundary, greater is the confidence with which LR predicts its label. This is not the case with perceptron and SVM.
- SVM uses a more principled approach compared to perceptrons. Max-margin classifiers will generalize better than perceptrons.
- The soft-margin formulation is robust to outliers, this is not the case with perceptrons which is guaranteed to converge only in the linearly separable case. Logistic regression is also more forgiving with outliers compared to perceptrons.

Common Data for Questions (9) and (13)

Statement

Consider a soft-margin SVM that has been trained on a dataset in $\mathbb{R}^2$. A subset of the data-points and the decision boundary (solid line) is shown below:



For each point, consider ξ_i to be the minimum bribe that has to be paid to take it to the correct supporting hyperplane.

Solution

For all the problems here, the basic idea is to start with the inequalities for the slack-variables:

$$\xi_i \geq 0$$
$$\xi_i \geq 1 - (\mathbf{w}^T\mathbf{x}_i)y_i$$

These correspond to the two inequalities in the soft-margin formulation for each point. Combining these two inequalities, we get:

$$\xi_i \geq \max(0, 1 - (\mathbf{w}^T\mathbf{x}_i)y_i)$$

Since we have been asked for the minimum bribe, it is going to be:

$$\xi_i = \max(0, 1 - (\mathbf{w}^T\mathbf{x}_i)y_i)$$

For points which are either on the right supporting hyperplane or beyond it, we have $(\mathbf{w}^T\mathbf{x}_i)y_i \geq 1$. As a result, we see that $\xi_i = 0$. That is, these points do not have to pay any bribe as they are already on the right side. Other points have to pay a non-zero, positive bribe. The value of the bribe can be computed by studying the geometry of the three parallel lines: $\mathbf{w}^T\mathbf{x} = 0$, $\mathbf{w}^T\mathbf{x} = -1$ and $\mathbf{w}^T\mathbf{x} = 1$.

Question-9

Statement

What is the value of ξ_1?

Answer

0

Question-10

Statement

What is the value of ξ_2?

Answer

0

Question-11

Statement

What is the value of ξ_4?

Answer

1

Question-12

Statement

What is the value of ξ_6?

Answer

2

Question-13

Statement

Select all true statements.

Options

(a)

$\xi_3 < 0$

(b)

$0 < \xi_3 < 1$

(c)

$0 < \xi_5 < 1$

(d)

$1 < \xi_5 < 2$

(e)

$\xi_7 > 2$

(f)

$\xi_7 < 0$

Answer

(b), (d), (e)