Graded

This document has 8 questions.

Question-1 [1 point]

Statement

Consider a linearly separable dataset for a binary classification problem in $\mathbb{R}^d$. Three linear classifiers have been trained on this dataset. All three pass through the origin and have the following property:

$$(\mathbf{w}_j^T \mathbf{x}_i) \cdot y_i \geq 1, \quad 1 \leq i \leq n$$

Here, $\mathbf{w}_j$ is the weight vector corresponding to the j^{th} classifier. Note that the above property is satisfied for each of the n data-points. If $\mathbf{w}_1$ is the weight vector corresponding to a hard-margin SVM, which of the following statements is always true? You can assume that the norms of all three weights are different from each other.

Options

(a)

$$||\mathbf{w}_1|| > ||\mathbf{w}_2|| > ||\mathbf{w}_3||$$

(b)

$$||\mathbf{w}_1|| < ||\mathbf{w}_2|| < ||\mathbf{w}_3||$$

(c)

$$||\mathbf{w}_1|| < ||\mathbf{w}_2|| \text{ and } ||\mathbf{w}_1|| < ||\mathbf{w}_3||$$

(d)

$$||\mathbf{w}_1|| > ||\mathbf{w}_2|| \text{ and } ||\mathbf{w}_1|| > ||\mathbf{w}_3||$$

Answer

(c)

Solution

$\mathbf{w}_1$ will have the smallest norm (maximum margin) among the three classifiers. The three weight vectors are feasible points for the primal. Among them, $\mathbf{w}_1$ is optimal.

Common Data for questions (2) to (4)

Statement

Common data for questions (2) to (4)

Consider the following training dataset for a binary classification problem in $\mathbb{R}^2$. Each data-point is represented by $\mathbf{x} = [x_1 \quad x_2]^T$ whose label is y.

Index	x_1	x_2	y
1	1	0	1
2	-1	0	-1
3	5	4	1
4	-5	-4	-1

We wish to train a hard-margin SVM for this problem. $\mathbf{w} = [w_1 \quad w_2]^T$ represents the weight vector. The index i is for the i^{th} data-point. α_i is the Lagrange multiplier for the i^{th} data-point.

Question-2 [1 point]

Statement

Select all primal constraints from the options given below.

Options

(a)

$$w_1 \geq 1$$

(b)

$$w_1 \leq 1$$

(c)

$$5w_1 + 4w_2 \geq 1$$

(d)

$$5w_1 + 4w_2 \leq 1$$

Answer

(a), (c)

Solution

Because of the symmetry in the problem, we effectively have only two constraints even though there are 4 data-points:

$$\begin{aligned} w_1 &\geq 1 \\ 5w_1 + 4w_2 &\geq 1 \end{aligned}$$

But in order to remain consistent with our formulation, let us list it down in the following manner:

$$\begin{aligned} w_1 &\geq 1 \quad (1) \\ w_1 &\geq 1 \quad (2) \\ 5w_1 + 4w_2 &\geq 1 \quad (3) \\ 5w_1 + 4w_2 &\geq 1 \quad (4) \end{aligned}$$

Question-3 [2 points]

Statement

Which of the following is the objective function of the dual problem? In all options, $\boldsymbol{\alpha} = [\alpha_1 \quad \alpha_2 \quad \alpha_3 \quad \alpha_4]^T$ and $\mathbf{1} = [1 \quad 1 \quad 1 \quad 1]^T$.

Options

(a)

$$\boldsymbol{\alpha}^T\mathbf{1} - \frac{1}{2} \cdot \boldsymbol{\alpha}^T\boldsymbol{\alpha}$$

(b)

$$\boldsymbol{\alpha}^T\mathbf{1} - \frac{1}{2} \cdot \boldsymbol{\alpha}^T \begin{bmatrix} 1 & 1 & 5 & 5 \\ 1 & 1 & 5 & 5 \\ 5 & 5 & 41 & 41 \\ 5 & 5 & 41 & 41 \end{bmatrix} \boldsymbol{\alpha}$$

(c)

$$\boldsymbol{\alpha}^T\mathbf{1} - \frac{1}{2} \cdot \boldsymbol{\alpha}^T \begin{bmatrix} 5 & 5 & 1 & 1 \\ 5 & 5 & 1 & 1 \\ 1 & 1 & 10 & 41 \\ 1 & 1 & 41 & 10 \end{bmatrix} \boldsymbol{\alpha}$$

(d)

$$\boldsymbol{\alpha}^T\mathbf{1} - \frac{1}{2} \cdot \boldsymbol{\alpha}^T \begin{bmatrix} 1 & 1 & 30 & 30 \\ 1 & 1 & 30 & 30 \\ 5 & 5 & 10 & 10 \\ 5 & 5 & 10 & 10 \end{bmatrix} \boldsymbol{\alpha}$$

Answer

(b)

Solution

The objective function corresponding to the dual is:

$$\boldsymbol{\alpha}^T\mathbf{1} - \frac{1}{2} \cdot \boldsymbol{\alpha}^T(\mathbf{Y}^T\mathbf{X}^T\mathbf{X}\mathbf{Y})\boldsymbol{\alpha}$$

We have:

$$\mathbf{X} = \begin{bmatrix} 1 & -1 & 5 & -5 \\ 0 & 0 & 4 & -4 \end{bmatrix}, \mathbf{Y} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

Therefore:

$$\mathbf{X}^T\mathbf{X} = \begin{bmatrix} 1 & -1 & 5 & -5 \\ -1 & 1 & -5 & 5 \\ 5 & -5 & 41 & -41 \\ -5 & 5 & -41 & 41 \end{bmatrix}$$

And:

$$\mathbf{X}^T\mathbf{X}\mathbf{Y} = \begin{bmatrix} 1 & 1 & 5 & 5 \\ -1 & -1 & -5 & -5 \\ 5 & 5 & 41 & 41 \\ -5 & -5 & -41 & -41 \end{bmatrix}$$

Finally:

$$\mathbf{Y}^T\mathbf{X}^T\mathbf{X}\mathbf{Y} = \begin{bmatrix} 1 & 1 & 5 & 5 \\ 1 & 1 & 5 & 5 \\ 5 & 5 & 41 & 41 \\ 5 & 5 & 41 & 41 \end{bmatrix}$$

Question-4 [1 point]

Statement

What is the optimal weight vector, $\mathbf{w}^*$?

Hint: Plot the points and try to compute the answer using geometry; do not try to solve the dual algebraically!

Options

(a)

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

(b)

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

(c)

$$\begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

(d)

$$\begin{bmatrix} 0 \\ -1 \end{bmatrix}$$

Answer

(a)

Solution

We see that the x_2 axis has to be the optimal separator as that is the one which has maximum margin. So, the equation of the decision boundary is $x_1 = 0$. This implies, $w_2 = 0$. Therefore, the weight vector becomes $[w_1 \quad 0]^T$. The choice of w_1 can be found out by noticing that $x_1 = 1$ and $x_1 = -1$ are the supporting hyperplanes. Therefore, $w_1 \cdot 1 = 1 \implies w_1 = 1$.

Question-5 [1 point]

Statement

Consider a kernel-SVM trained on a dataset of 100 points with polynomial kernel of degree 2. If $\boldsymbol{\alpha}^*$ is the optimal dual solution, what is the predicted label for a test-point $\mathbf{x}_{\text{test}}$?

Options

(a)

$$\sum_{i=1}^{100} \alpha_i^* \cdot \mathbf{x}_{\text{test}}^T \mathbf{x}_i \cdot y_i$$

(b)

$$\text{sign}\left(\sum_{i=1}^{100} \alpha_i^* \cdot \mathbf{x}_{\text{test}}^T \mathbf{x}_i \cdot y_i \right)$$

(c)

$$\sum_{i=1}^{100} \alpha_i^* \cdot (1 + \mathbf{x}_{\text{test}}^T \mathbf{x}_i)^2 \cdot y_i$$

(d)

$$\text{sign}\left(\sum_{i=1}^{100} \alpha_i^* \cdot (1 + \mathbf{x}_{\text{test}}^T \mathbf{x}_i)^2 \cdot y_i \right)$$

Answer

(d)

Solution

The optimal weight vector is given by:

$$\mathbf{w}^* = \sum_{i=1}^{100} \alpha_i^* \cdot \phi(\mathbf{x}_i) \cdot y_i$$

Here, $\phi(\mathbf{x}_i)$ is the vector in the transformed space. First we compute the dot-product:

$$\mathbf{w^*}^T \phi(\mathbf{x}_{\text{test}}) = \sum_{i=1}^{100} \alpha_i^* \cdot \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_{\text{test}}) \cdot y_i$$

$$= \sum_{i=1}^{100} \alpha_i^* \cdot k(\mathbf{x}_i, \mathbf{x}_{\text{test}}) \cdot y_i$$

$$= \sum_{i=1}^{100} \alpha_i^* \cdot (1 + \mathbf{x}_{\text{test}}^T \mathbf{x}_i) \cdot y_i$$

Finally, the prediction is:

$$\text{sign}\left(\mathbf{w^*}^T\phi(\mathbf{x}_{\text{test}})\right) = \text{sign}\left(\sum_{i=1}^{100} \alpha_i^* \cdot (1 + \mathbf{x}_{\text{test}}^T \mathbf{x}_i)^2 \cdot y_i\right)$$

Common Data for questions (6) and (7)

Statement

Common data for questions (6) and (7)

Consider the transformation $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}^6$ associated with the polynomial kernel with degree 2:

$$\phi(\mathbf{x}) = \begin{bmatrix} 1 \\ x_1^2 \\ x_2^2 \\ \sqrt{2}x_1x_2 \\ \sqrt{2}x_1 \\ \sqrt{2}x_2 \end{bmatrix}$$

A kernel-SVM is trained on a dataset with the above kernel. The optimal weight vector is as follows:

$$\mathbf{w}^* = \begin{bmatrix} -25 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

You can assume that the dataset is linearly separable in the transformed space.

Question-6 [1 point]

Statement

What is the shape of the decision boundary in $\mathbb{R}^2$?

Options

(a)

It is a parabola of the form $x_2 = 25x_1^2$

(b)

It is a parabola of the form $x_2 = -25x_1^2$

(c)

It is a straight line

(d)

It is a circle

Answer

(d)

Solution

The decision boundary in $\mathbb{R}^2$ is given by:

$$\mathbf{w^*}^T\phi(\mathbf{x}) = 0$$

$$x_1^2 + x_2^2 = 25$$

It is a circle centered at the origin with radius 5.

Question-7 [2 points]

Statement

Which of the following training data-points are certainly **not** support vectors?

Options

(a)

$[-1 \quad 5]^T$

(b)

$[3 \quad 5]^T$

(c)

$[-4 \quad -5]^T$

(d)

$[2\sqrt{5} \quad -2]^T$

(e)

$[\sqrt{30} \quad \sqrt{6}]^T$

Answer

(b), (c), (e)

Solution

All the support vectors will lie on the two supporting hyperplanes:

$$\mathbf{w}^{*T}\phi(\mathbf{x}) = \pm 1$$

This gives two curves:

$$x_1^2 + x_2^2 = 24 \quad \text{OR} \quad x_1^2 + x_2^2 = 26$$

These curves are two circles, one smaller than the decision boundary and one larger than the decision boundary. From the points given here, there are two points that could be support vectors:

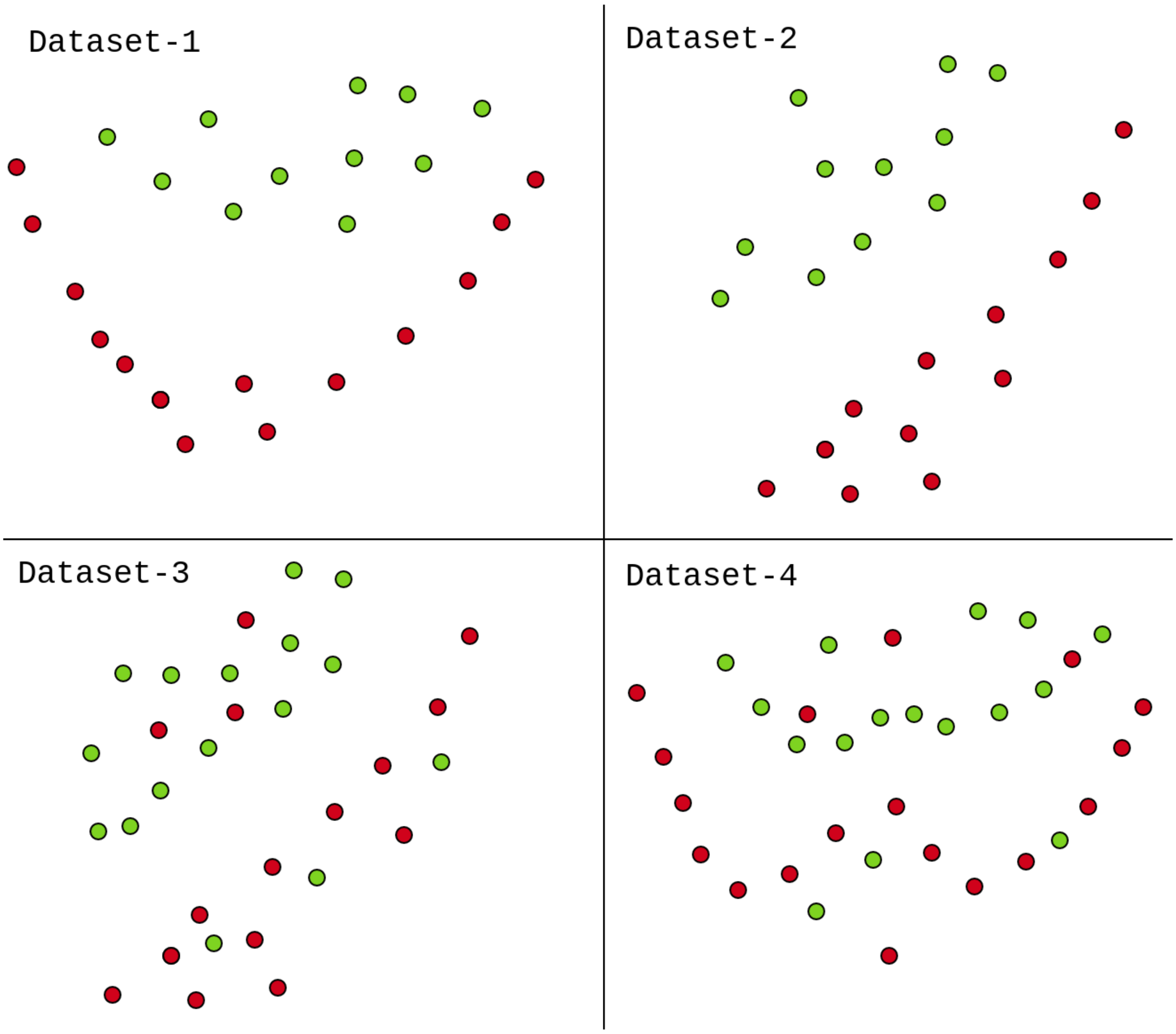
- $[-1 \quad 5]^T$: this point lies on $x_1^2 + x_2^2 = 26$
- $[2\sqrt{5} \quad -2]^T$: this point lies on $x_1^2 + x_2^2 = 24$

Note that these two are "potential" support vectors. Every support vector lies on the supporting hyperplanes. But every point on the supporting hyperplanes need not be a support vector.

Question-8 [1 point]

Statement

Match the following classification datasets with the most appropriate choice of adjectives:



Options

(a)

Dataset-1: kernel, hard-margin

Dataset-2: hard-margin

Dataset-3: soft-margin

Dataset-4: hard-margin

(b)

Dataset-1: kernel, hard-margin

Dataset-2: hard-margin

Dataset-3: soft-margin

Dataset-4: kernel, soft-margin

(c)

Dataset-1: soft-margin

Dataset-2: hard-margin

Dataset-3: kernel

Dataset-4: soft-margin

Answer

(b)

Solution

- Dataset-1: the decision boundary is non-linear. The structure is non-linear and the problem is linearly separable in some high dimensional space.
- Dataset-2: this is a clear case of hard-margin SVM
- Dataset-3: the boundary is linear. The presence of outliers suggests that this should be solved using a soft-margin SVM
- Dataset-4: The boundary is non-linear. In addition, the dataset has outliers. So, this should involve both a kernel and a soft-margin formulation.