

Practice

This document has 12 questions.

Question-1

Statement

You are given a training dataset of 100 data-points for a classification task. You wish to use a k -NN classifier, with $k = 10$. $d(\mathbf{x}_i, \mathbf{x}_j)$ is a function that returns the Euclidean distance between two points \mathbf{x}_i and \mathbf{x}_j . Calling d on a pair of points corresponds to a single distance computation. If you want to predict the label of a new test point, how many distances would you have to compute?

Answer

100

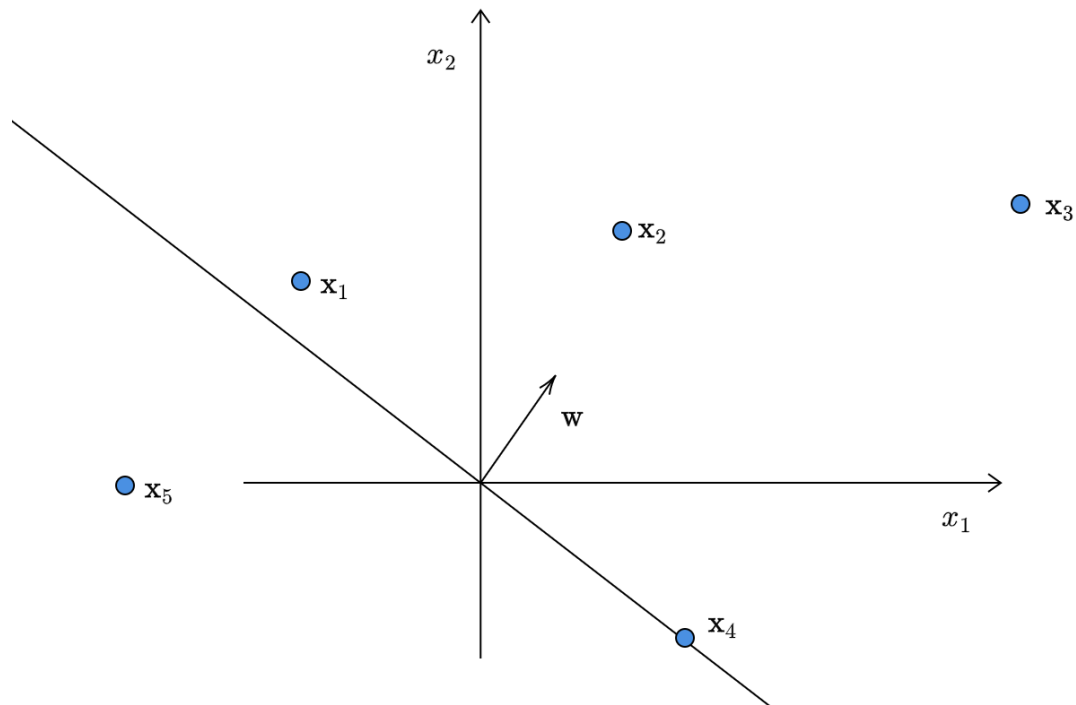
Solution

We would have to compute 100 distances: $d(\mathbf{x}_{\text{test}}, \mathbf{x}_i)$ for $1 \leq i \leq 100$.

Common Data for questions (2) to (3)

Statement

Consider the following data-points in a binary classification problem. \mathbf{w} is the weight vector corresponding to a linear classifier. The labels are $+1$ and -1 .



Question-2

Statement

What is the predicted label for these five points?

Options

(a)

All five points are predicted as $+1$.

(b)

All five points are predicted as -1 .

(c)

$\mathbf{x}_1, \mathbf{x}_2$ and \mathbf{x}_3 are predicted as $+1$ and $\mathbf{x}_4, \mathbf{x}_5$ are predicted as -1 .

(d)

$\mathbf{x}_1, \mathbf{x}_2$ and \mathbf{x}_3 are predicted as -1 and $\mathbf{x}_4, \mathbf{x}_5$ are predicted as $+1$.

Answer

(c)

Solution

For a linear classifier with weight vector \mathbf{w} , the prediction for a test-point \mathbf{x} is given by:

$$y_{\text{pred}} = \text{sign}(\mathbf{w}^T \mathbf{x}) = \begin{cases} 1, & \mathbf{w}^T \mathbf{x} > 0 \\ -1, & \mathbf{w}^T \mathbf{x} \leq 0 \end{cases}$$

The geometric interpretation of this is as follows: a linear classifier divides the space into two half-spaces. This decision is made just by looking at the sign of the dot-product.

Half-space-1: The dot-product is positive.

$$\mathbf{w}^T \mathbf{x} > 0$$

This corresponds to those cases in which the data-point makes an acute angle with the weight vector.

Half-space-2: The dot-product is non-positive.

$$\mathbf{w}^T \mathbf{x} \leq 0$$

This corresponds to those cases in which the data-point makes either a right angle or an obtuse angle with the weight vector. When a point lies on the line (right angle with weight), it could be classified either way. As per convention, we choose -1 .

Question-3

Statement

Which of the following statements are true?

Options

(a)

$$0 < \mathbf{w}^T \mathbf{x}_1 < \mathbf{w}^T \mathbf{x}_2 < \mathbf{w}^T \mathbf{x}_3$$

(b)

$$0 < \mathbf{w}^T \mathbf{x}_3 < \mathbf{w}^T \mathbf{x}_2 < \mathbf{w}^T \mathbf{x}_1$$

(c)

$$\mathbf{w}^T \mathbf{x}_4 = 0$$

(d)

$$\mathbf{w}^T \mathbf{x}_5 > 0$$

(e)

$$\mathbf{w}^T \mathbf{x}_5 < 0$$

Answer

(a), (c), (e)

Solution

The farther away a data-point is from the decision boundary, the larger the magnitude of its projection onto the weight vector. Hence:

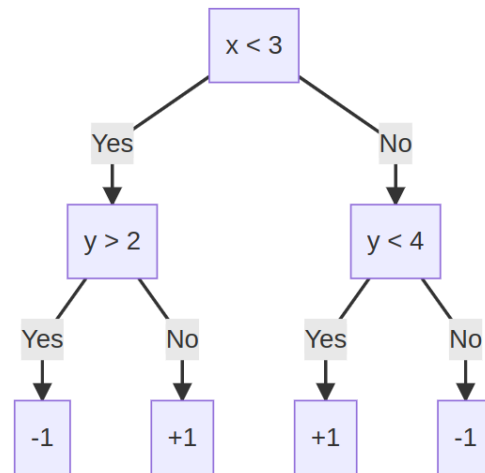
$$|\mathbf{w}^T \mathbf{x}_3| > |\mathbf{w}^T \mathbf{x}_2| > |\mathbf{w}^T \mathbf{x}_1|$$

Since all these three points lie on the positive half-plane, these dot products will be positive and we can remove the modulus sign. As for point \mathbf{x}_4 , its projection on \mathbf{w} will be zero as it is orthogonal to it. Finally, $\mathbf{w}^T \mathbf{x}_5$ will be negative as \mathbf{x}_5 lies in the negative half-plane.

Comprehension Type (4 - 5)

Statement

Consider the following decision tree for a binary classification problem that has two features: (x, y) .



Question-4

Statement

Which of the following statements are true?

Options

(a)

$(1, 0)$ is classified as -1

(b)

$(0, 1)$ is classified as $+1$

(c)

$(4, 2)$ is classified as -1

(d)

$(5, 7)$ is classified as -1

Answer

(b), (d)

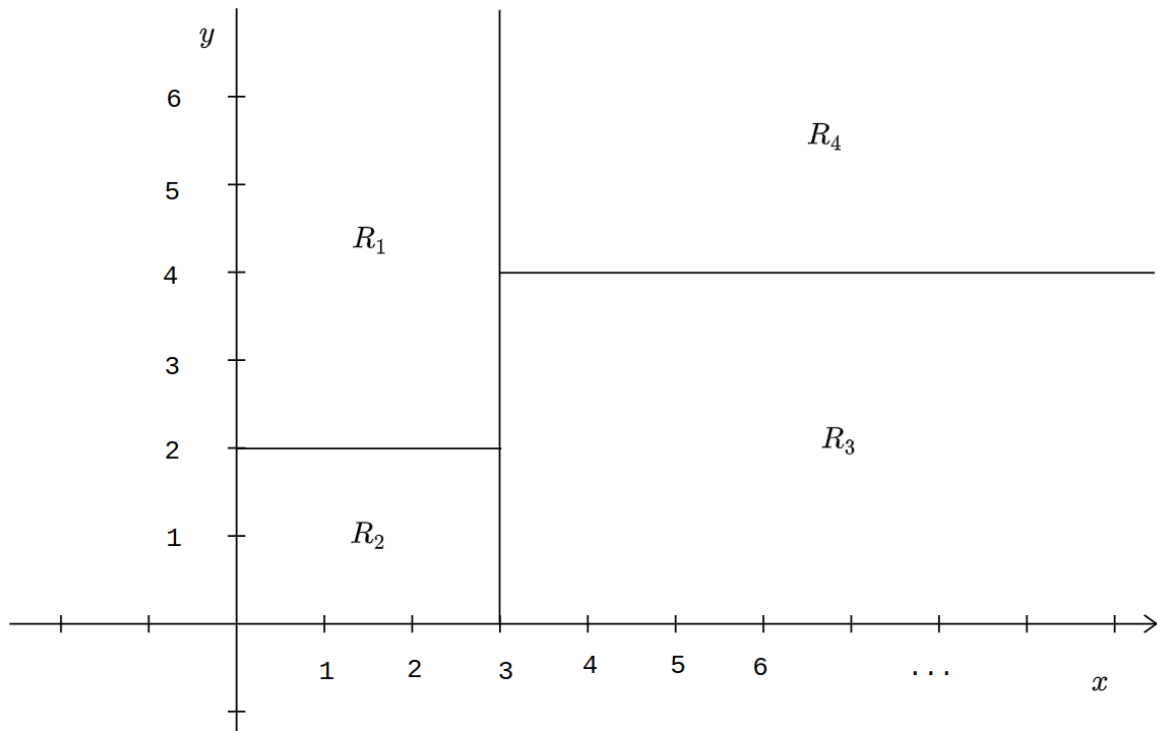
Solution

Start at the root of the decision tree and keep traversing the internal nodes until you end up with a prediction/leaf node.

Question-5

Statement

This decision tree partitions the feature space into four regions as shown below:



Assume that $x, y \geq 0$ for all points. What can you say about the predicted labels of points that fall in the four regions?

Options

(a)

$$R_1 : +1$$

$$R_2 : -1$$

$$R_3 : +1$$

$$R_4 : -1$$

(b)

$$R_1 : -1$$

$$R_2 : +1$$

$$R_3 : +1$$

$$R_4 : -1$$

(c)

$$R_1 : -1$$

$$R_2 : +1$$

$$R_3 : -1$$

$$R_4 : +1$$

(d)

$$R_1 : +1$$

$$R_2 : +1$$

$$R_3 : -1$$

$$R_4 : -1$$

Answer

(b)

Solution

The tree partitions the space into four regions: R_1, R_2, R_3, R_4 . These four regions correspond to the four leaves from left to right.

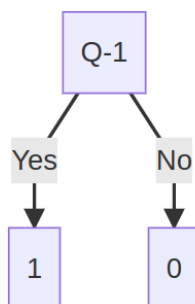
Question-6

Statement

Consider the following training dataset for a binary classification task:

x	y
2	1
10	0
8	0
-4	1
0	1
9	0

The following decision tree cleanly separates the two classes, such that the resulting leaves are pure.



Q-1 is of the form $x < v$. How many possible integer values can v take?

Answer

6

Solution

v can take any one of the values from this set: $\{3, 4, 5, 6, 7, 8\}$. It can also take the value 8 as $x < 8$ will go to the right branch for the data-point 8.

Question-7

Statement

p is the proportion of points with label $+1$ in some leaf in a decision tree. For what values of p will this node be assigned a label of $+1$? Select the most appropriate option.

Options

(a)

$$0 \leq p < 0.5$$

(b)

$$0.5 < p \leq 1$$

(c)

$$0 \leq p \leq 1$$

Answer

(b)

Solution

In a vote, we need more points to belong to class 1 than class 0. Therefore, $p > 0.5$

Question-8

Statement

Q is some internal node (question node) of a decision tree that splits into two leaf nodes (prediction nodes) L and R . The proportion of data-points belonging to class 1 in each of the three nodes is the same and is equal to 0.7. What is the information gain for the question corresponding to node Q ?

Answer

0

Solution

Let p be the proportion of ones in each node. Then, all three nodes have the same entropy as entropy only depends on p . Call this E . If γ is the ratio into which the original dataset is partitioned by this question, then:

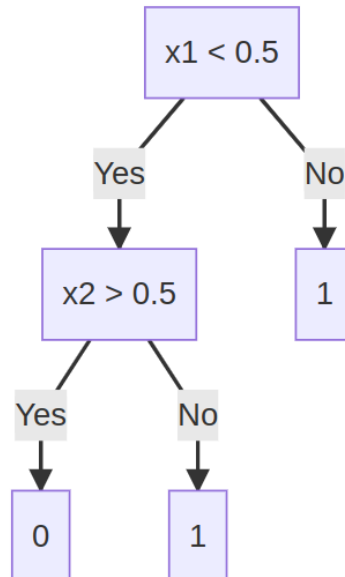
$$IG = E - [\gamma E + (1 - \gamma)E] = 0$$

We gain nothing because of this split. Intuitively this makes sense because in both leaf nodes the status quo prevails. To put it in another way, none of the child nodes have become purer than the parent node. They have the same level of impurity as their parent.

Question-9

Statement

Consider the following decision tree for a classification problem in which all the data-points are constrained to lie in the unit square in the first quadrant. That is, $0 \leq x_1, x_2 \leq 1$.



If a point is picked at random from the unit square, what is the probability that the decision tree predicts this point as belonging to class 1?

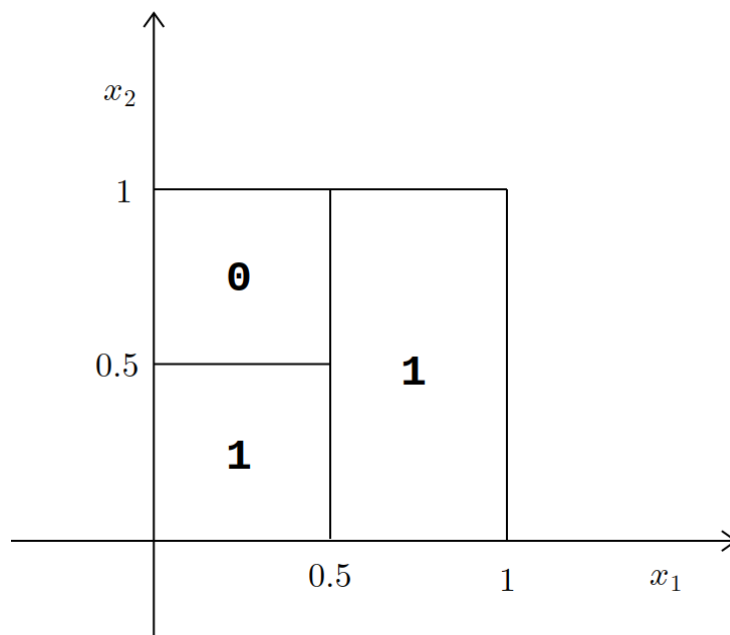
Answer

0.75

Range: [0.74, 0.76]

Solution

The decision regions will look as follows:

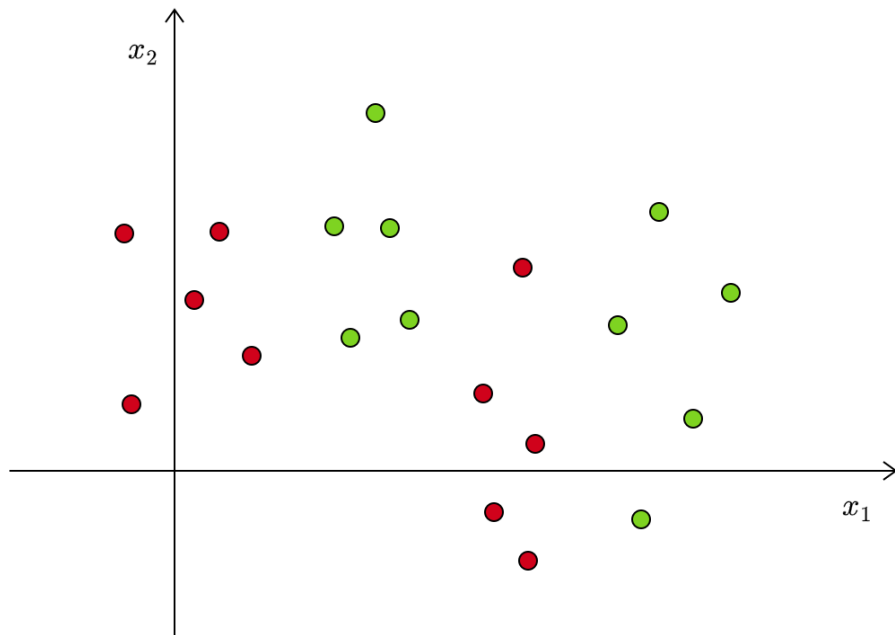


We see that 75% of the region belongs to class 1, hence the required probability is 0.75.

Question-10

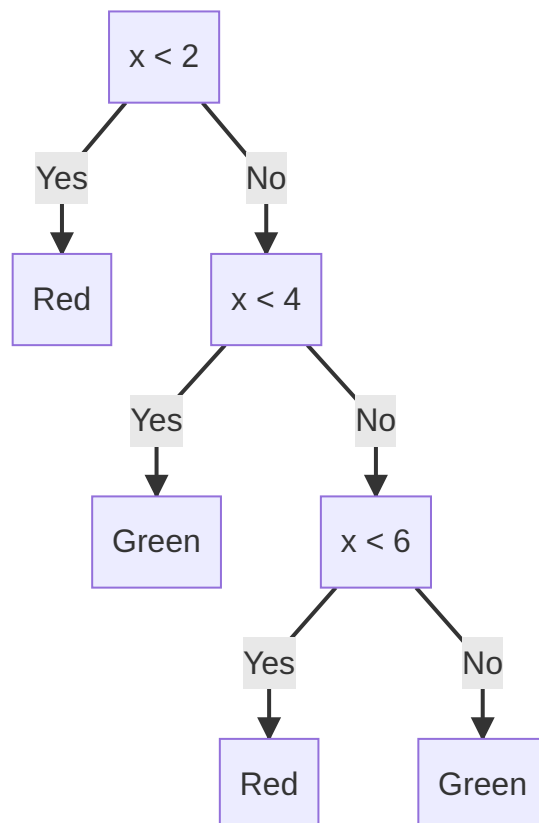
Statement

Consider the following dataset. By visually inspecting the data-points, construct a decision tree.



Solution

The structure of the data seems clear. There are four blobs of red and green points. From left to right, we have blob-1, blob-2, blob-3 and blob-4. They take the colors red, green, red and green respectively. Assume that the lines parallel to the y-axis that separate consecutive blobs are $x = 2$, $x = 4$ and $x = 6$. Think about why we need only three lines. One tree that we could construct has the following form:



Question-11

Statement

$H(p)$ is the entropy of a node. p is the proportion of points that belong to class 1. q is the proportion of points that belong to class 0. Which of the following statements is true?

Options

(a)

$$H(p) > H(q)$$

(b)

$$H(p) < H(q)$$

(c)

$$H(p) = H(q)$$

(d)

Insufficient information. We need the number of data-points that are in this node.

Answer

(c)

Solution

We notice the symmetry in the entropy function about the axis $p = 0.5$. Therefore, $H(p) = H(1 - p)$:

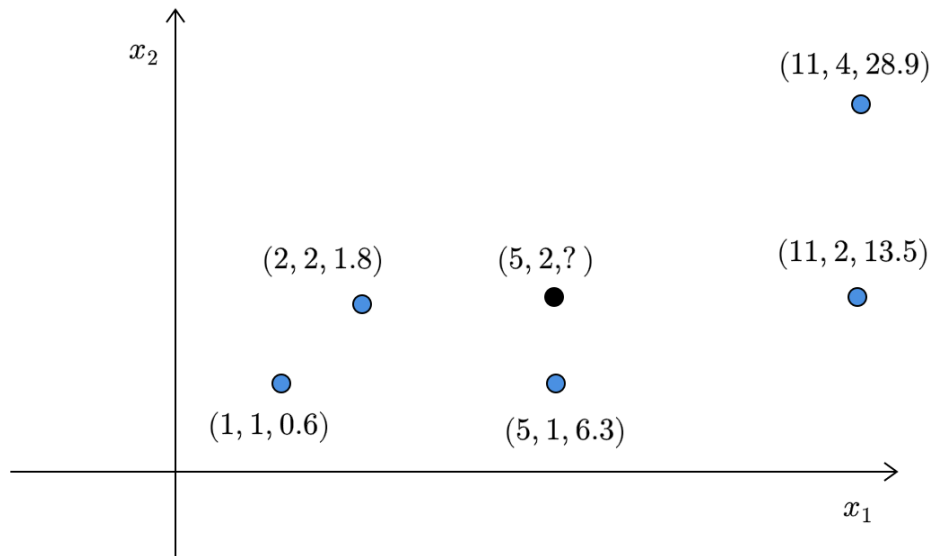
$$H(1 - p) = -(1 - p) \log(1 - p) - p \log p = H(p)$$

Since $q = 1 - p$, we have $H(p) = H(q)$. So, for determining the impurity of a node, we could either use the proportion of ones or the proportion of zeros.

Question-12

Statement

The k -NN algorithm can be adapted to perform regression giving us what is called the k -NN regressor. Rather than looking at the majority vote among the k nearest neighbors, a k -NN regressor computes the mean of the labels of the k nearest neighbors and returns this as the prediction. Consider the following dataset:



All blue points belong to the training dataset. The annotation for each point should be understood as (x_1, x_2, y) . What is the predicted label for the black test-point? Use $k = 3$.

Answer

2.9

Range: [2.8, 3]

Solution

The three nearest neighbors are:

- (1, 1)
- (5, 1)
- (2, 2)

The prediction for the test-point is the average of the labels corresponding to these three points:

$$\frac{1.8 + 0.6 + 6.3}{3} = 2.9$$