1) We have a dataset of 1000 points for a classification problem using k-NN algorithm. Now consider the following statements:

S1: If $k = 10$, it is enough if we store any 10 points in the training dataset.

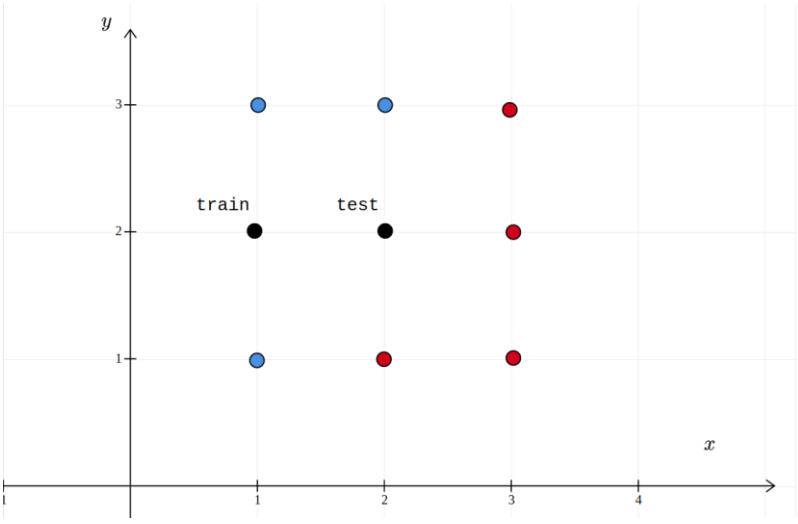
S2: If $k = 10$, we need to store the entire dataset.

S3: The number of data-points that we have to store increases as the size of k increases.

S4: The number of data-points that we have to store is independent of the value of k.

- S1 and S3 are true statements
- S2 and S4 are true statements
- S1 alone is a true statement
- S3 alone is a true statement
- S4 alone is a true statement

2) The blue and the red points belong to two different classes. Both of them are a part of the training dataset. The black point at $(1, 2)$ also belongs to the training dataset, but its true color is hidden form our view. The black point at $(2, 2)$ is a test-point. ***1 point***



How should we recolor the black train point if the test point is classified as "red" without any uncertainty by a k-NN classifier, with $k = 4$? Use the Euclidean distance metric for computing distances.

- blue
- red
- Insufficient information

3) Consider the following feature vectors that make up the training dataset:

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \\ -1 \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} 5 \\ -3 \\ -5 \\ 10 \end{bmatrix}, \mathbf{x}_3 = \begin{bmatrix} 3 \\ 1 \\ 2 \\ 4 \end{bmatrix}, \mathbf{x}_4 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \mathbf{x}_5 = \begin{bmatrix} 10 \\ 7 \\ -3 \\ 2 \end{bmatrix}$$

The labels of these five points are:

$y_1 = 1,\ y_2 = 0,\ y_3 = 1,\ y_4 = 0,\ y_5 = 0$

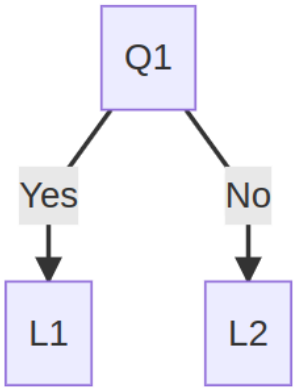
If we use a k-NN algorithm with $k = 3$, what would be the predicted label for the following test point:

$$\mathbf{x}_{\text{test}} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

1

Common Data for Q4 to Q6

Consider the following split at some node in a decision tree:



The following is the distribution of data-points and their labels:

Node	Num of points	Labels
Q1	100	0
Q1	100	1
L1	50	0
L1	30	1
L2	50	0
L2	70	1

For example, L1 has 80 points of which 50 belong to class 0 and 30 belong to class 1. Use $\log_2$ for all calculations that involve logarithms. Use three decimal places for all intermediate quantities.

4) If the algorithm is terminated at this level, then what are the labels associated with L1 and L2?

- [x] L1 : 0
- [ ] L1 : 1
- [ ] L2 : 0
- [x] L2 : 1

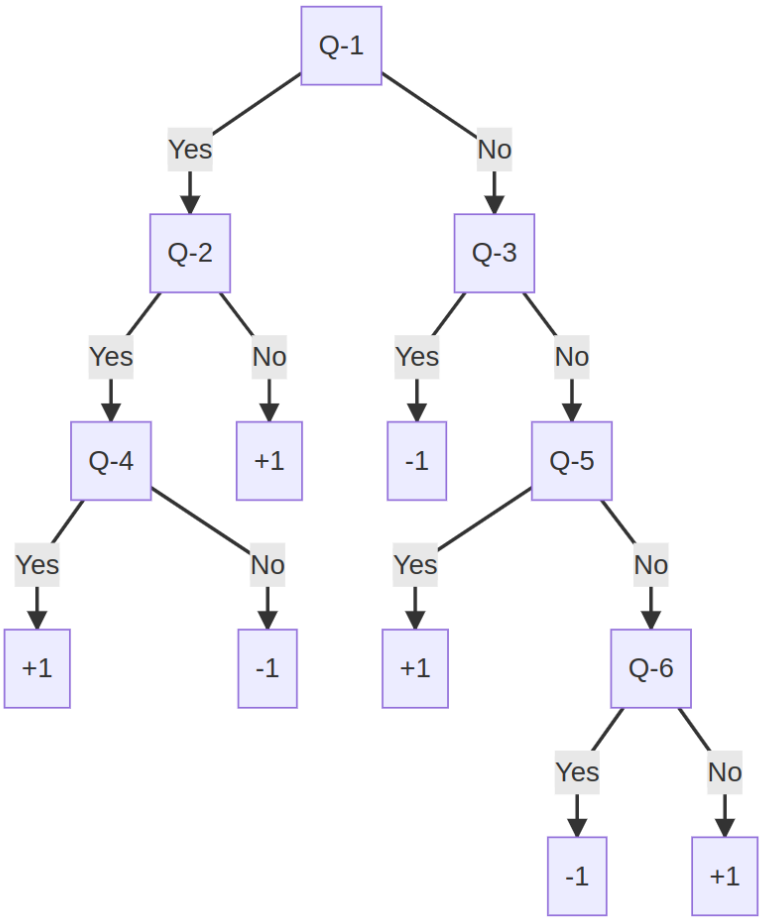
5) What is the impurity in L1 if we use entropy as a measure of impurity? Report your answer correct to three decimal places.

0.955

6) What is the information gain for this split? Report your answer correct to three decimal places. Use at least three decimal places in all intermediate computations.

0.020

7) Consider the following decision tree. Q-i corresponds to a question. The labels are $+1$ and -1.



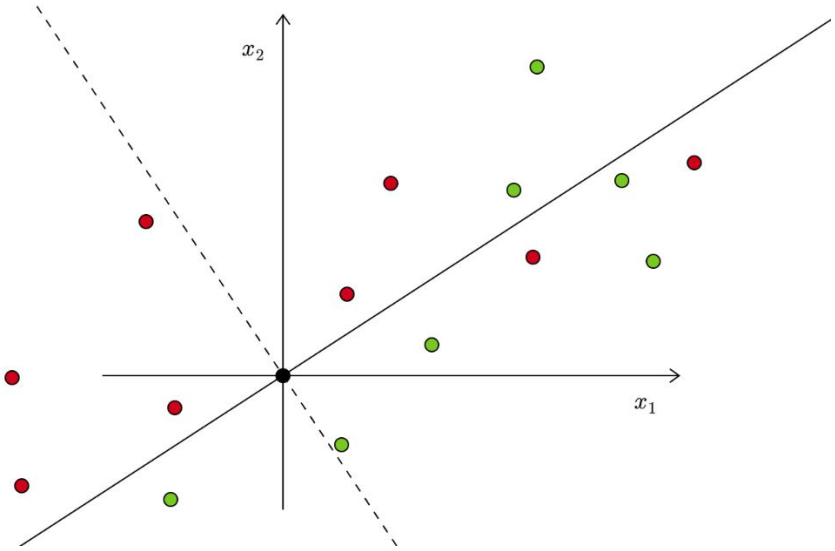
If a test-point comes up for prediction, what is the minimum and maximum number of questions that it would have to pass through before being assigned a label?

- [ ] $\min = 1$
- [x] $\min = 2$
- [ ] $\min = 3$
- [ ] $\max = 3$
- [x] $\max = 4$

8) p is the proportion of points with label 1 in some node in a decision tree. Which of the following statements are true? [MSQ]

- [ ] As the value of p increases from 0 to 1, the impurity of the node increases
- [ ] As the value of p increases from 0 to 1, the impurity of the node decreases
- [ ] The impurity of the node does not depend on p
- [x] $p = 0.5$ correspond to the case of maximum impurity

9) Consider a binary classification problem in which all data-points are in $\mathbb{R}^2$. The red points belong to class $+1$ and the green points belong to class -1. A linear classifier has been trained on this data. The decision boundary is given by the solid line. ***1 point***



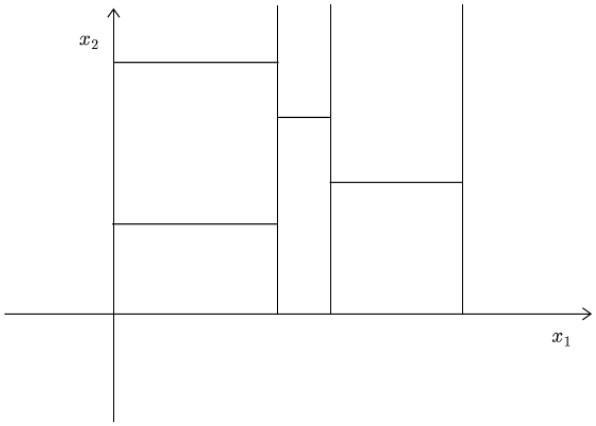
This classifier misclassifies four points. Which of the following could be a possible value for the weight vector?

- [ ] $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$
- [x] $\begin{bmatrix} -1 \\ 2 \end{bmatrix}$
- [ ] $\begin{bmatrix} -1 \\ -2 \end{bmatrix}$
- [ ] $\begin{bmatrix} 1 \\ -2 \end{bmatrix}$

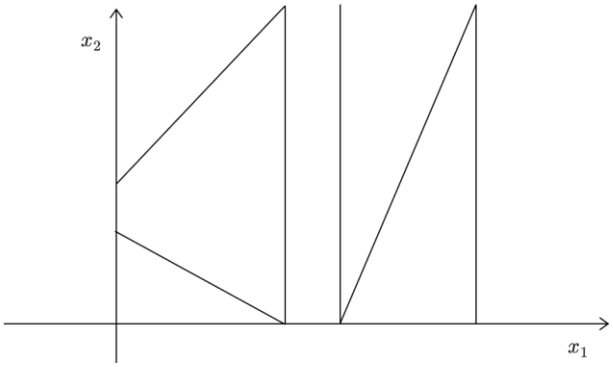
10) Which of the following are valid decision regions for a decision tree classifier for datapoints in $\mathbb{R}^2$? The question in every internal node is of the form $f_k \leq \theta$. Both the features are positive real numbers. ***1 point***

Option-1

Option-2



Option-3



Option-4