Name	Student's ID	Score

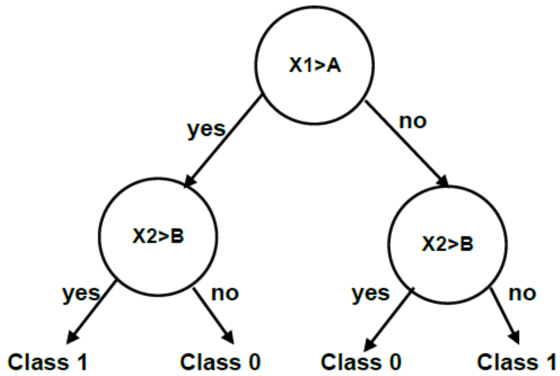
DECISION SUPPORT SYSTEM 2017-2018
THE EXAMINATION

1. What happens when the number of nodes in the hidden layer of a two-layer perceptron is increased?
    a. It will be capable of representing more complicated decision boundaries
    b. It will be less prone to overfitting
    c. It will generalize better
    d. It will converge faster

2. Assume we have the decision tree in which classifies two dimensional vectors

$$(X_1;X_2)\in\mathbf{R}^2\text{-}\{A;B\}.$$

In other words, the values A and B are never used in the inputs. Can this decision tree be implemented as a 1-NN? If so, explicitly say what are the values you use for the 1-NN (you should use the minimal number possible). If not, either explain why or provide a counter example.

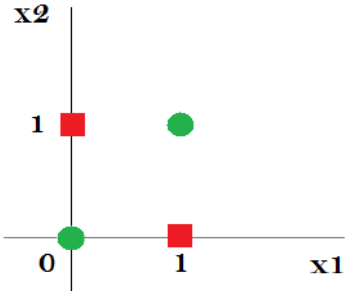


3. Assume that you are given observations $(x_1; x_2) \in \mathrm{R}^2$ in the following order

	x_1	x_2	*class*
1	10	10	+1
2	0	0	-1
3	8	4	+1
4	3	3	-1
5	4	8	+1
6	0.5	0.5	-1
7	4	3	+1
8	2	5	+1

Show the action of the perceptron algorithm for the above sequence of observations. We start with an initial set of weights w = (1, 1) and bias b = 0.

4. The following figure illustrates exclusive-OR with two inputs (x_1 and x_2), where the squares are labeled as class 1 and the circles are labeled as class 0.



As the figure illustrates and as we have discussed in the class, the VC dimension of a **linear classifier** in 2D is 3.

   a. Assuming that we would like to correctly shatter any set of 4 points with linear decision boundaries, what would you do with the input points to allow successful classification?

   b. What is the VC dimension of the classifier proposed in question a)?

5. Our goal is to construct a decision tree classifier for predicting flight delays on the base of 6 features: Rain, Wind, Summer, Winter, Day, Night. We have collected data for a few months and a summary of the data is provided in the following table:

ID	**Rain**	**Wind**	**Summer**	**Winter**	**Day**	**Night**	**Delayed?**
1	Yes	No	Yes	Yes	No	Yes	*Delayed*
	…	…					…
80	No	No	Yes				*Not*

The contingence tables of these six features are as follow:

		Rain		**Wind**		**Summer**		**Winter**		**Day**		**Night**	
		Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Delayed	Yes	30	10	25	15	5	35	20	20	20	20	15	25
	No	10	30	15	25	35	5	10	30	20	20	10	30

   a. Based on the table, which feature should be at the root of the decision tree (briefly explain, no need to provide exact values for information gain)?

   b. Based on the table, which feature should be on the second level (the level just beneath the root) of the decision tree (briefly explain)?

6. Consider the 3-means algorithm on a set S consisting of the following 6 points in the plane: a=(0,0), b=(8,0), c=(16,0), d=(0,6), e=(8,6), f=(16,6). The algorithm uses the Euclidean distance metric to assign each point to its nearest centroid; ties are broken in favor of the centroid to the left/down. A starting configuration is a subset of 3 starting points from S that form the initial centroids. A 3-partition is a partition of S into 3 subsets; thus {a,b,e}, {c,d}, {f} is an example of 3-partition; clearly any 3-partition induces a set of three centroids in the natural manner. A 3-partition is stable if repetition of the 3-means iteration with the induced centroids leaves it unchanged.
    a. How many starting configurations are there?

    b. What are the stable 3-partitions?

    c. What is the number of starting configurations leading to each of the stable 3-partitions in (b) above?

    d. What is the maximum number of iterations from any starting configuration to its stable 3-partition?

7. You have a dataset that involves four features. Feature D's values are in [0,100]. For the other three features, all of their possible values appear in this dataset.

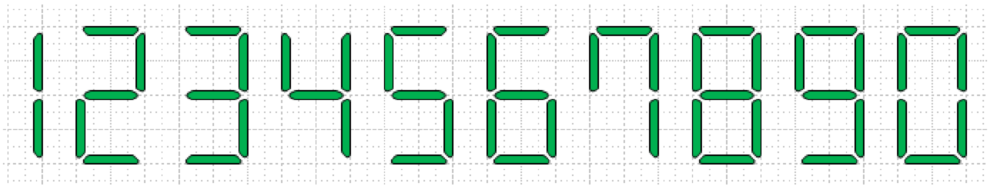
	A	B	C	D	Category
Ex1	T	X	F	75	true
Ex2	F	Y	T	20	false
Ex3	T	Z	T	10	true
Ex4	F	Y	T	35	false
Ex5	F	X	T	90	false
Ex6	T	Z	F	50	false

    a. How much information about the category is gained by knowing the value of feature B ?

    b. What are the three nearest-neighbors to Example 6? Explain. If this example was in the test set instead of the training set, would k-NN predict it correctly (using k=3)?

    c. Draw the best decision tree for this data set

8. Each digit in Digital Clock is made of a certain number of dashes, as shown in the image below. Each dash is displayed by a LED (light-emitting diode). Propose a decision table to store the information about those digits and use the rough set methods to solve the following problems:



a) Assume that we want to switch off some LEDs to save the energy, but we still want to recognise the parity of the digit presented by the remaining dashes. What is the minimal set of dashes you want to display?

b) The same question for the case we want to recognise all digits