Assignment 6

Computer Science Department, University of Crete

MACHINE LEARNING - CS 577, Fall 2024

<u>Deadline:</u> 27/1/2025, 23:55 on e-learn (`https://elearn.uoc.gr`).

Deliverable files: Submit a pdf file containing your answers.

Exercise 1 [50 points]

In the table below you are provided with a training data of a 1-norm, soft-margin SVM as well as the (fictional) Lagrange multipliers α that stem from training the model with cost $C = 10$. The data are fictional with values that make the computations easier. With lower x are denoted the input training vectors and capital X the input variables, while y is the class. The kernel employed is the full polynomial quadratic of degree 2: $K(x, z) = (x \cdot z + 1)^2$. The dot stands for the inner product between two vectors.

1. Explain why the Lagrange multipliers cannot really be the solution to an 1-norm, soft-margin SVM problem
2. How are the features in feature space relate to the input variables X_i.
3. What is the weight vector w and the intercept term b that defines the decision surface $f(x_{test}) = sign(w \cdot x_{test} + b)$.
4. Write the same classification function f without using w but only using the kernel
5. Describe as clearly as possible all the regions in feature space in comparison to the SVM margin where the input vectors x_i fall for $i = 1, \ldots, 5$.

Sample	a	y	X1	X2
x1	1	1	1	0
x2	10	-1	0	-1
x3	10	-1	1	1
x4	0	1	-1	0
x5	0	-1	0	-1

Exercise 2 [50 points]

Draw a Bayesian Network with 5 nodes and 10 edges; each node should have at least 2 neighbors and there should be at least 3 colliders in the graph. Assume your data distribution is faithful to the network. Assume that you have enough samples and perfect statistical tests that correctly identify all dependencies and independencies in your data. Name one of the nodes as the target T. T should have parents, children, and spouse nodes.

1. Which variables will be returned by the feature selection algorithm that selects all variables dependent with T.

2. Show a possible trace of the Forward-Backward Feature Selection algorithm. In the trace show the conditional independence tests that the algorithm performs at each step, and the selections that it makes. Use the d-separation criterion to explain why the algorithm could have made the selections you indicate. (hint: create the network in such a way that the algorithm finishes in the minimal number of steps, or you'll never finish the computations).

3. Do the same for the Forward-Backward with Early Dropping algorithm with one run.

4. Compare the execution of the two algorithms on the data from your network: which algorithm is computationally faster and which has better quality of results.