Assignment 4

Computer Science Department, University of Crete

MACHINE LEARNING - CS 577, Fall 2022

Assignment 4

Deadline: 9/11/2022, 23:55 on e-learn (`https://elearn.uoc.gr`).
Deliverable files: Submit a zip file containing a report in PDF with the answers **AND all** Python files (.py) written by you in the scope of the assignment. The final grade will be the result of the quality of your submitted results in your report, together with the correctness of your submitted code. Do not forget to include the plots produced by your code in your report and comment on them.

Exercise 1 (Programming)

You are given the following six tables of the joint probability distribution of two random variables, X, Y, that take values $\{0, 1\}$. Our goal is to check if variable X is independent of Y in each case. The null hypothesis states that X is independent of Y. The significance threshold is 0.05.

	Y	
X	0	1
0	0.05	0.1
1	0.25	0.6

	Y	
X	0	1
0	0.08	0.23
1	0.17	0.52

	Y	
X	0	1
0	0.1	0.2
1	0.2	0.5

	Y	
X	0	1
0	0.1	0.1
1	0.1	0.7

	Y	
X	0	1
0	0.15	0.1
1	0.1	0.65

	Y	
X	0	1
0	0.45	0.05
1	0.05	0.45

Table 1: Joint probability distributions

A. From each joint probability distribution, sample $n = 25$, $n = 100$ and $n = 1000$ values for X and Y.

B. For each joint probability distribution and each sample size :

   i) Compute the X^2 statistic and the p-value (hint: cumulative distribution function). Decide if we should reject the null hypothesis.

   ii) Perform a permutation test to decide if we should reject the null hypothesis. You can try 1000 permutations. Plot the distribution of the p-values. What do you observe ?

Note: Do not use external python packages to compute the X^2 statistic and the permutation test. You have to implement them on your own.