Question: In this problem, the goal of using the nearest neighbor programming in the classification of glandular disease It is thyroid. The used data has 100 examples, each of which contains 7 characteristics and one output. The output can be one of the values 0, equivalent to no diagnosis of the disease, 1, 2 or 3, diagnosis of the disease with one of three different classes. The characteristics are as follows:

Age -
ON THYROXINE -
TSH MEASURED -
TSH -
T3 -
TT4 MEASURED -
TT4 -

that characteristics 2, 3 and 6 are binary and the rest of the characteristics are real.

Divide the desired data into 5 equal parts and calculate the effect of changing k from 1 to 10 on the average classification error in the KNN algorithm using Validation-Cross.

We do this in two ways:

A) The state where non-normalized data enters the algorithm

b) The state where the data is normalized in such a way that the average of each of the characteristics becomes zero and their variance becomes one.