Detection of the stage of malignant tumors

1. Introduction

Studying medical analyzes of the patients on predicting diseases based on their diagnosis always was a priority task of any doctor. Sometimes, despite the high technological equipment, having test results tests may be not enough. The required number of tests can sometimes be too much, which makes this process quite difficult. In these cases, the specialist is forced to make a diagnosis allowing a certain probability of error. One of such time-consuming tasks is to determine the stage of malignant tumors.

The stage of the majority of malignant tumors can be denoted by the numbers (1, 2, 3, 4, 5) reflecting both the size of the tumor and its spread within the organ.

1 - Critical stage, tumor with metastasis sprouting into neighboring organs
2 - Critical stage with a beginning stage of metastasis
3 - Mobile tumor (from 2 cm) without metastasis
4 - Limited tumor process (up to 2 cm) without affecting the nearest lymph nodes
5 - Tumor not detected.

2. Job Description

We collected anonymous data from the analysis of 8,000 patients who underwent diagnostic tests for the detection of malignant tumors. Overall, 15 results of the patient's analysis were diagnosed, namely:

Blood analysis							Urine analysis			Computer tomography scan				
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15

Note: the F1 ... F15 fields may indicate such things like number of white blood cells and etc., but this does not make much sense within the scope of this task. Each characteristic takes a value from 0 to 10.
The task is to determine the stage of the tumor from these 15 data. The stage of the tumor, as noted above, is determined by the values 1, 2, 3, 4, 5.

Resources:
- train.csv - it stores data (8, 000 lines), which are labeled by patients cancer stage. The first line is a descriptor, i.e. describes each column. Each new line (starting with the second) stores data about the analysis of one unique patient. Each line contains 16 numbers separated by commas, (the first 15 numbers are the results of the analysis, the last one is the stage of the disease).
- test.csv - stores data (without labels) for which it is necessary to correctly classify the stage of the disease (2,000 lines). The same as in the first file here each new line stores data of the analysis of one unique patient.

The format for sending the results:
You need to prepare the archive firstname_secondname.rar containing 2 files:
- results.csv - a file consisting of 2000 lines in each of which one single number is recorded (number from 1 to 5) - probabilistic prediction of disease stages on the corresponding line in the file test.csv.

- **description.doc** / docx / pdf / txt - description of the method by which the participant solved the problem.

3. Criteria for estimating the algorithm
- The results of the algorithm are estimated using the standard formula for calculating accuracy.

4. Notes
1. The application will not be considered if at least one of the sending files is missing or in an incorrect format.
2. In case of same results the winner will be chosen by instructor based on the theoretical knowledge of a student.