Methods of classification and dimensionality reduction 05.03.2023

BONUS TASK

Paweł Lorek

DEADLINE: 17.06.2023, 23:59
BENEFITS FOR BEST SOLUTIONS: TBA

1 Data prediction – problem statement

In Moodle a file `data_train.csv` is posted. It is an `csv` (separated with semicolons "`;`") file with the following format:

- First column – row number
- First 34 columns (0 – 33) are the features, last column is the class $\in \{0, 1\}$.

In a file `data_test_public_no_classes.csv`, there are only features (34 columns) of test data. **THE TASK** is to predict the class (0 or 1) for each row.

Submission format:

- File must be named in the following format: `2023_MCaDR_bonus_task_IndexNr.csv`
- It should be a `csv` file separated with semicolons "`;`"
- It should contain as many rows as `data_test_public_no_classes.csv` and only two columns:
  - First column = row number
  - Second column = your prediction
- Sample submission file `2023_MCaDR_bonus_task_random_12345678.csv` is posted in Moodle. Its few first rows:

```
0;1
1;0
2;0
3;0
4;1
5;1
6;1
7;1
```

Metric. Say that y_i is the "true" value of the next number in row i, and $\hat{y}_i$ is your prediction. I will use accuracy to evaluate your submission (R = number of rows):

$$ACC = \frac{1}{R}\sum_{i=0}^{R-1} \mathbf{1}(y_i = \hat{y}_i).$$