Pattern recognition projects

Nov. 4th, 2019

Data

- 6 datasets, dataset01-dataset06

- 3 artificial problems (1-3), with large number of samples, files *.npz (use **numpy.load()** to read the files);
    - within each file, there is a matrix “X” (samples by rows) and a vector “y” (for labels)

- 3 real-world problems (4-6): for each problem, there are 2 files (*.dat) with the data matrix X (*-train-X.dat) and the labels (*-train-y.dat). Use **numpy.loadtxt()** for reading. Data matrix dimensions: 400x14,883; 450x14,883; 450x14,883 (samples by rows)

Rules

- Model each problem separately, using the classifier(s) of your choice

- Once you believe you have the “right” model, train the final classifier and estimate its performance in terms of (at least): error rate, sensitivity and specificity and provide for each 95% confidence intervals

- The test sets will be released on **Dec. 2nd, 2019**: by the end of that day:

    - you should send me a brief description (1 page/slide per model - **best: an IPython notebook file *.ipynb**) of your modeling approach (a schema would do) and the **claimed performance** (and 95% CIs)

    - you should send me your predictions for each of the 6 problems in the form of a text file with 1 label per row

- On **Dec. 9th, 2019** I will release the evaluations of your models: performance and how well you estimated it.