Data sets

Data set: https://drive.google.com/a/diag.uniroma1.it/file/d/14p_YtIWVqBNxvOMWRPp-qWEops0wAJ1F

Blind test set: https://drive.google.com/a/diag.uniroma1.it/file/d/1uCIIVNvQoa2fXdQQg2e5GHt_5bY3BgOd

Data set contains labelled data as explained in the seminar, blind test set contains only input data to classify after model creation.

Project development

1) Solve the two classification problems: A) optimization prediction, B) compiler prediction.
For each classification problem, realize at least two variants (varying feature extraction, learning algorithm, learning hyper-parameters, etc.).
Note: Use any method at your choice, except neural networks that will be subject of the second homework.
2) Evaluate each variant in a proper way. Find the best model and motivate the choice.
3) For each classification problem, apply the best model to predict output for the blind test set.
4) Write a report (about 10 pages) explaining all the work done: design and implementation choices, evaluation procedure and results. Reports must be individual.

Submission procedure

Submit through classroom the following files:
1) PDF file of the report (no other formats accepted),
2) ZIP file with the code (without data set),
3) CSV file with output on the blind test set, name of the file should be <your_matricola>.csv (e.g., 1234567.csv). If you don't have a matricola, use your last name.

Deadline: 10/11/2019 11:59 PM CET