Homework is due Today

Details for the homework (1/2)

- You lose one point every day your are late.
- Count for 50% of the course grade.
- Once you send me the homework, I add your name to a Google document, which is publicly available from the course web page.
- Check that your name is on the list at most two days after you send your homework, and check that your affiliation is correct.

Details for the homework (2/2)

- For each question, you get a score
  - 1: perfect, congratulations;
  - 0.75: correct but one argument is missing;
  - 0.5: important part is missing or incorrect, part is correct;
  - 0.25: there is something correct in the response;
  - 0: response is missing or incorrect.
- The final grade will be a linear function of the previous scores (with appropriate weights).
- I may also award bonus points for original and elegant proofs, or strong results (good constants).

Data challenge (1/3)

The goal of the data challenge is to learn how to implement machine learning algorithms from A to Z, and gain understanding about them. For this reason, the task is fairly simple: digit recognition with a reasonable number of training samples (5000).

- You need to use kernel methods.
- It will happen from March 1st to March 15th. The link will be posted on the course web page on March 1st.
- It will be organized on the platform Kaggle. Each of you need to register. Then, you can form teams.
- Teams may be of 1 to 3 students.
- Two members from a "homework team" cannot be in the same team for the data challenge.
- You need an e-mail address @ens-cachan.fr or @polytechnique.edu to register. I can add other domains on demand.

Data challenge (2/3)

- One important rule will be "Do it yourself"; machine learning libraries (libsvm, liblinear, scikit-learn,...) are not allowed; optimization and linear algebra libraries are allowed; (this is why we are dealing with a simple dataset).
- Two days after the challenge deadline, you need to provide a small report (in pdf format, 11pt, 2 pages A4 max).
- You also need to provide with your report the source code (zip or tar.gz archive), with a simple script "start" (that may be called from Matlab, Python or R) which will reproduce your submission file.
- A leader board will be available during the challenge, which shows the best results per team, as measured on a subset of the test set.
- A different part of the test set will be used after the challenge to evaluate the results.

Data challenge (3/3)

- Using external training data is not allowed.
- The final score will be an unknown function of the performance you obtain, and of what you have done.