Scalable Machine Learning: IA317

Core infos

- Where : Telecom Paris
- Volume : 24hrs
- When : 8 weeks (one week for exam)

Teaching staff

Nom	Etablissement	Email
P. BIANCHI	Télécom ParisTech	pascal.bianchi@telecom-paristech.fr
R. GOWER	Télécom ParisTech	robert.gower@telecom-paristech.fr
T. BONALD	Télécom ParisTech	thomas.bonald@telecom-paristech.fr
F. d'ALCHE	Télécom ParisTech	florence.dalche@telecom-paristech.fr

Syllabus

The course will introduce techniques and the fundamental mathematics for handling large dimensional data within the context of machine learning. This includes
1) dimension reduction tools, that allow one to compress large scale data and still preserve desirable properties
2) large scale classification techniques based on random forests, gradient boosting and XGboost
3) large scale unsupervised learning through clustering techniques.
At the end of the course, the students will be aware of the costs of handling large scale data, the underlining theory and be able to implement the techniques in Python.

Evaluation

- **Labs**. 2 Labs with Jupyter graded (40% of the final grade).
- **Exam**. 1.5h Exam (60% of the final grade).

To sum up :

Eval type	% final grade	Remarks
Lab1	20%	e-mail to gowerrobert@gmail.com by 22:00 on 21/Nov
Lab2	20%	e-mail to pascal.bianchi@telecom-paris.fr by 22:00 on 04/Dec
Exam	60%	Final course exam

Detailed program

Subject		Prof	Date	Time
Sparse matrix format, Dimension reduction, advanced LA, JL lemma	Leçon - Amphi Rose Dieng-Kuntz	Robert	07-Nov	9:00 - 12:15
Python notebook on Dimension reduction	TP- 1A201 and 1A207	Robert	14-Nov	9:00 - 12:15
Random Forest, Gradient boosting, xgboost	Leçon - Amphi Rose Dieng-Kuntz	Pascal	21-Nov	9:00 - 12:15
Python notebook on Random Forest, Gradient boosting, xgboost	TP- 1A201 and 1A207	Pascal	28-Nov	9:00 - 12:15
Large-scale clustering, hashing, sketching	Leçon - Amphi 2	Thomas	05-Dec	9:00 - 12:15
Python notebook on clustering, hashing, sketching	TP - 1A201 and 1A207	Thomas	12-Dec	9:00 - 12:15
Random fourier features and Kernel methods	Leçon - Amphi Rose Dieng-Kuntz	Florence	19-Dec	9:00 - 12:15
Exam	Leçon - Amphi Rose Dieng-Kuntz		09-Jan-2020	9:00 - 12:00

Bilan

- Pascal : 3h*1.5 lectures + 3h TP = 8h
- Robert : 3h*1.5 lectures + 3h TP = 8h
- Thomas : 3h*1.5 lectures + 3h TP = 8h
- Florence : 3h*1.5h lectures = 5h