We will be holding a sprint in Paris, 25 Feb to Mar 1st. Each day, the sprint will start at 9.30 am until 6.30 pm.

Contact: Guillaume Lemaitre - +33761104782

Location

AXA REV, 7th floor, Immeuble Java, 61 rue Mstislav Rostropovitch, 75017 Paris

Transport

- Metro line 13 (Brochant / Porte de Clichy)
- RER C (Porte de Clichy)
- Tram T3b
- Transilien line L (Pont Cardinet)

Social event

- Monday 25 February: Breakfast at AXA with a short presentation (10 minutes) about `sckit-learn` and sprint objectives.
- Friday 1 March: Aperitif at launch organised by AXA.

People present:

Indicate your expertise level, if you need funding for travel or accommodation and what you want to work on.

- Gaël Varoquaux, core developer, no funding needed
- Guillaume Lemaitre, core developer, no funding needed
- Adrin Jalali, core developer, needs (some) funding
- Alex Gramfort, core developer, no funding needed
- Roman Yurchak, core developer, no funding needed
- Joel Nothman (until 28 Feb), core developer, funding arranged, work_on={pandas issues, sample and feature props, review, too many things...}
- Joan Massich, contributor, no funding needed
- Joris Van den Bossche, core developer, no funding needed
- Andreas Mueller, core developer, no funding needed, working on governance finalization, SLEP process, SLEP reviewing, Roadmap
- Nicolas Hug, contributor, no funding needed
- Thomas Fan, contributor, no funding needed
- Nicolas Goix, contributor, no funding needed
- Albert Thomas, contributor, no funding needed, might not be attending the whole week, merge PR #12827, refactor tests #10027, reviews
- Jérémie du Boisberranger, contributor, no funding needed
- Thomas Moreau, contributor, no funding needed
- Pavel Soriano, new contributor, no funding needed
- William de Vazelhes, contributor, no funding needed, PR #10058 in progress (metric learning)
- Aurélien Bellet, contributor, no funding needed, will probably attend Mon-Tue, PR #8602 and #10058 (metric learning), issue #12228 (graph lasso)
- Romuald Menuet, new contributor, no funding needed
- Olivier Grisel, core developer, no funding need
- Maria Telenczuk, new contributor, no funding needed.
- Bartosz Telenczuk, new contributor, no funding needed.
- Ivan Panico, contributor, no funding needed.
- Oliver Rausch, contributor, funding has been handled.
- Pierre Glaser, contributor, no funding needed.

- Patricio Cerda, contributor, no funding needed (Online NMF).
- Pierre Ablin, contributor, no funding needed (Online NMF).
- Danilo Bzdok, no funding need (KNNImputer, AveragingRegressor, Added estimator checks for pandas object, FrequencyEncoder, Adding explained variances to sparse pca)
- Sébastien Treguer, no funding needed. (joining if space available)
- Assia Benbihi, new contributor, no funding needed.
- Xavier Dupre, new contributor, no funding needed.
- Samuel Ronsin, contributor, no funding needed.
- Julien Jerphanion, potentail new contributor, no funding needed.

Suggested tasks

- The most important tasks are to finish off pull requests, fix bugs and close issues. For this, it can be useful to look at tickets labelled 'easy': https://github.com/scikit-learn/scikit-learn/issues?page=2&q=is%3Aopen+label%3Aeasy

Welcoming new contributors

The sprint is a great time for new contributors to become familiar with the project. We welcome newcomers. Please be sure to read the contributing section of the documentation http://scikit-learn.org/dev/developers/contributing.html, and to have a development environment ready in which you can install scikit-learn from scratch, build it, and use git to push changes to github.

Technical Discussions Schedule

Time	Monday	Tuesday	Wednesday	Thursday	Friday
10:00	Welcome	(logreg) tol/convergence	get_feature_names	pandas handling	efficient GridSearch
16:00	OPTICS? ARM?	Freezing #9397	fit_transform	sample props	beers?

Discussions to add on in spare time:

- GLM support... poisson regression, quantile regression, etc.
- ~~Euclidean distances consistency and stability~~
- ~~Sample props (and feature props?) and their transformation~~
- contrib maintenance (is this the right model? who maintains it? what clear criteria for acceptance?)
- keyword only arguments
- ~~pipeline slicing~~
- search spaces

Explanation / issues:

- Freezing:
- convergence:
- get_feature_names: (includes pipeline slicing because reasons)
- fit_transform: (included imbalance learn interface discussion maybe)
- search spaces: (related to configspace and searchgrid)
- efficient grid-search: (should this include avoiding recomputation of preprocessing steps as well as the warm start logic? - our caching right now is not great and daskML does much better...

Meeting Minutes:

- GridSearch/warm_start: https://docs.google.com/document/d/17aYRi4rfi7KxQnD20NJInhKpELjmjmz8gT6EAdgHGCg/edit?ts=5c7915ad
- Pipeline slicing: https://github.com/scikit-learn/scikit-learn/pull/2568
- Feature names: https://github.com/scikit-learn/scikit-learn/pull/13307
- Resampler API: https://github.com/scikit-learn/enhancement_proposals/pull/15
- Sample properties: https://github.com/scikit-learn/enhancement_proposals/pull/16
- Pandas discussion: https://hackmd.io/4szB-nytQiafWDE6sKf5-g#