

Gitter: <https://gitter.im/scikit-learn/wimlds>

# Crash-Course in Contributing to Open Source Projects

(with some specific instructions  
to sklearn)

Andreas Müller

# Environment

- If you don't have a working python installation, install anaconda.
- If you don't want to install the scikit-learn development version in your main environment, create a separate environment (with conda or virtualenv):

```
conda create -n sklearndev numpy scipy matplotlib  
            pytest sphinx cython ipykernel  
source activate sklearndev  
conda install -c conda-forge sphinx-gallery
```

# Fork sklearn on github

<https://github.com/scikit-learn/scikit-learn>

The screenshot shows the GitHub repository page for `scikit-learn / scikit-learn`. The repository is owned by `scikit-learn` and has 16,863 stars and 9,344 forks. The `Fork` button is circled in red. The repository description is "scikit-learn: machine learning In Python" with a link to <http://scikit-learn.org>. The repository has 21,709 commits, 17 branches, 80 releases, 801 contributors, and is licensed under BSD-3-Clause. The commit history shows the latest commit by `zorroblue` and `agramfort` 3 hours ago, adding a check for `check_X_y` to `lasso_stability_path()`. The commit history also lists several other commits, including fixes for benchmarks, build tools, doc, examples, and sklearn, as well as updates to `.coveragerc`, `.gitattributes`, and `.gitignore`.

scikit-learn / **scikit-learn** Unwatch 1,576 Unstar 16,863 **Fork** 9,344

[Code](#) [Issues 785](#) [Pull requests 493](#) [Projects 5](#) [Wiki](#) [Pulse](#) [Graphs](#) [Settings](#)

scikit-learn: machine learning In Python <http://scikit-learn.org> [Edit](#)

[machine-learning](#) [python](#) [statistics](#) [data-science](#) [data-analysis](#) [Manage topics](#)

21,709 commits 17 branches 80 releases 801 contributors BSD-3-Clause

Branch: master [New pull request](#) [Create new file](#) [Upload files](#) [Find file](#) [Clone or download](#)

**zorroblue** committed with **agramfort** Added check\_X\_y to lasso\_stability\_path() (#7534) Latest commit 8c7ab17 3 hours ago

<a href="#">benchmarks</a>	[MRG+] Fix bench_rcv1_logreg_convergence.py by adding get_max_square...	25 days ago
<a href="#">build_tools</a>	Travis: tweak test_script.sh (#8444)	8 days ago
<a href="#">doc</a>	Correct figure number + matplotlib 2 (#8483)	a day ago
<a href="#">examples</a>	Correct figure number + matplotlib 2 (#8483)	a day ago
<a href="#">sklearn</a>	Added check_X_y to lasso_stability_path() (#7534)	3 hours ago
<a href="#">.coveragerc</a>	coverall added	3 years ago
<a href="#">.gitattributes</a>	MAINT remove .c files from .gitattributes	3 months ago
<a href="#">.gitignore</a>	Ignore py.test generated .cache folder	11 days ago

# Clone the Fork

The screenshot shows the GitHub interface for a forked repository. A green circle highlights the repository name 'amuel / scikit-learn' and the 'Code' button. A red circle highlights the 'Clone or download' button and the 'Clone with SSH' dropdown menu. The dropdown menu shows the option to 'use https' (highlighted in red) and the URL 'git@github.com:amuel/scikit-learn.git'. Below the dropdown menu is a 'Download ZIP' button. The repository statistics show 21,555 commits, 275 branches, 43 releases, and 752 contributors. The commit history shows a list of commits with their messages and dates.

Firefox Web Browser

amuel / scikit-learn  
forked from scikit-learn/scikit-learn

Unwatch 3 Unstar 9 Fork 9,344

Code Pull requests 0 Projects 0 Wiki Pulse Graphs Settings

scikit-learn main repo <http://scikit-learn.sourceforge.net> Edit

New Add topics

21,555 commits 275 branches 43 releases 752 contributors PSD-3-Clause

Branch: master New pull request Create new file Upload files Find file Clone or download

This branch is 154 commits behind scikit-learn:master.

TomDLT committed with ogrisel [MRG+1] Add multiplicative-update solver in NMF, with all beta-diverg...

benchmarks MAING make benchmarks scripts Python 3 compatible

build\_tools CI report which doc files were likely affected (#8032)

doc [MRG+1] Add multiplicative-update solver in NMF, with all beta-diverg... 3 months ago

examples [MRG+1] Add multiplicative-update solver in NMF, with all beta-diverg... 3 months ago

sklearn [MRG+1] Add multiplicative-update solver in NMF, with all beta-diverg... 3 months ago

If you don't have ssh keys set up, use https!

Add main scikit-learn repo as remote called "upstream":

```
git remote add upstream https://github.com/scikit-learn/scikit-learn.git
```

# Build and run tests!

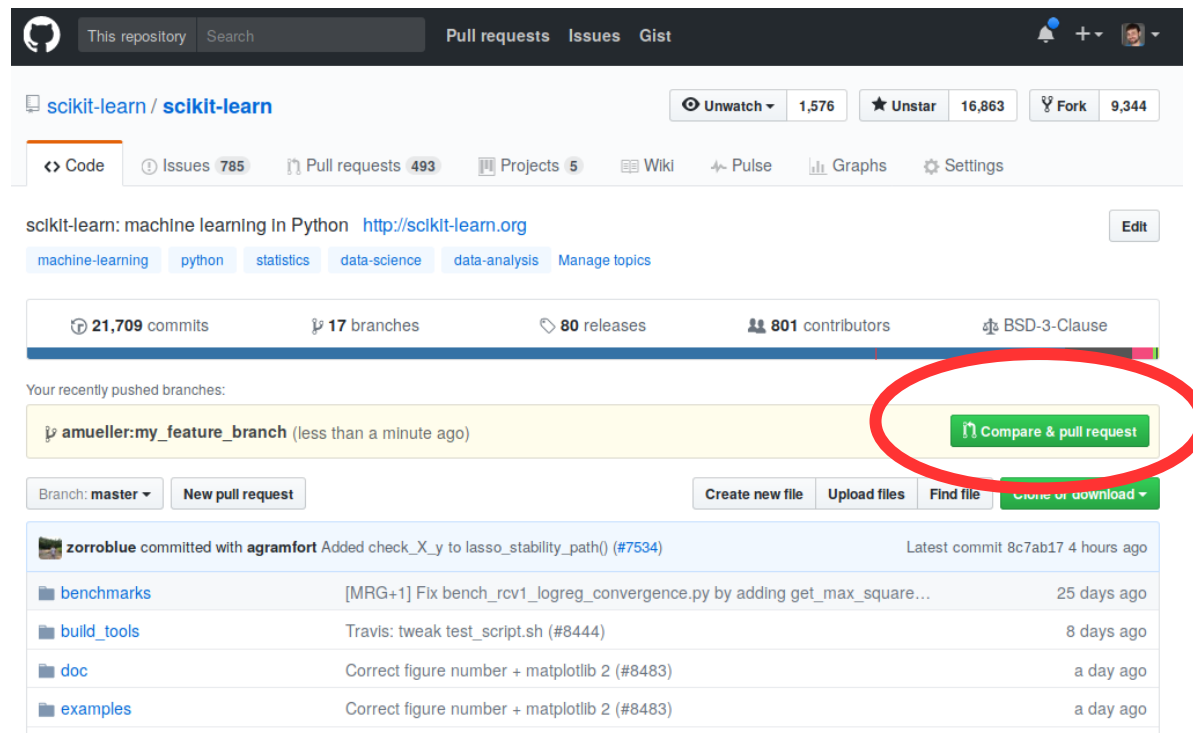
- Easiest way:  
`cd scikit-learn`  
`pip install -e .`
- Will overwrite existing installations!
- Windows: install build tools for visual studio:  
<https://visualstudio.microsoft.com/downloads/#build-tools-for-visual-studio-2019>
- OS X:  
`export SKLEARN_NO_OPENMP=True`

# Starting on Issues

- Go to <https://github.com/WiMLDS/nyc-2019-scikit-sprint/projects/1>
- Comment on issue saying “I’m working in this”.
- Update local master branch from main sklearn repo (`git pull upstream master`).
- Fetching someone else’s PR:  
`git fetch https://github.com/theirusername/reponame.git theirbranch:ourbranch`
- Create feature branch  
`git checkout -b <branchname>`
- Commit changes to branch, run tests  
`pytest sklearn`  
(or individual test files)
- Run flake8 on changed files

# Creating a PR


- Push changes to your fork  
`git push origin <branchname>`
- Create PR using github UI:



# Describing PR

## Open a pull request

Create a new pull request by comparing changes across two branches. If you need to, you can also [compare across forks](#).



base fork: **scikit-learn/scikit-learn** ▼

base: **master** ▼

...

head fork: **amueller/scikit-learn** ▼

compare: **my\_feature\_branch** ▼

✓ **Able to merge.** These branches can be merged together.

Please review the [guidelines for contributing](#) to this repository.



[MRG] minor pep8

Write

Preview

AA ▼

B

i

“

<>

🔗

⋮

≡

✓

✎

↶

@

🌟

<!--

Thanks for contributing a pull request! Please ensure you have taken a look at the contribution guidelines: <https://github.com/scikit-learn/scikit-learn/blob/master/CONTRIBUTING.md#Contributing-Pull-Requests>

-->

### Reference Issue

<!-- Example: Fixes #1234 -->

Attach files by dragging & dropping, [selecting them](#), or pasting from the clipboard.

☒ **Allow edits from maintainers.** [Learn more](#)

Create pull request

Reviewers

No reviews—request one

Assignees

No one—assign yourself

Labels

None yet

Milestone

No milestone

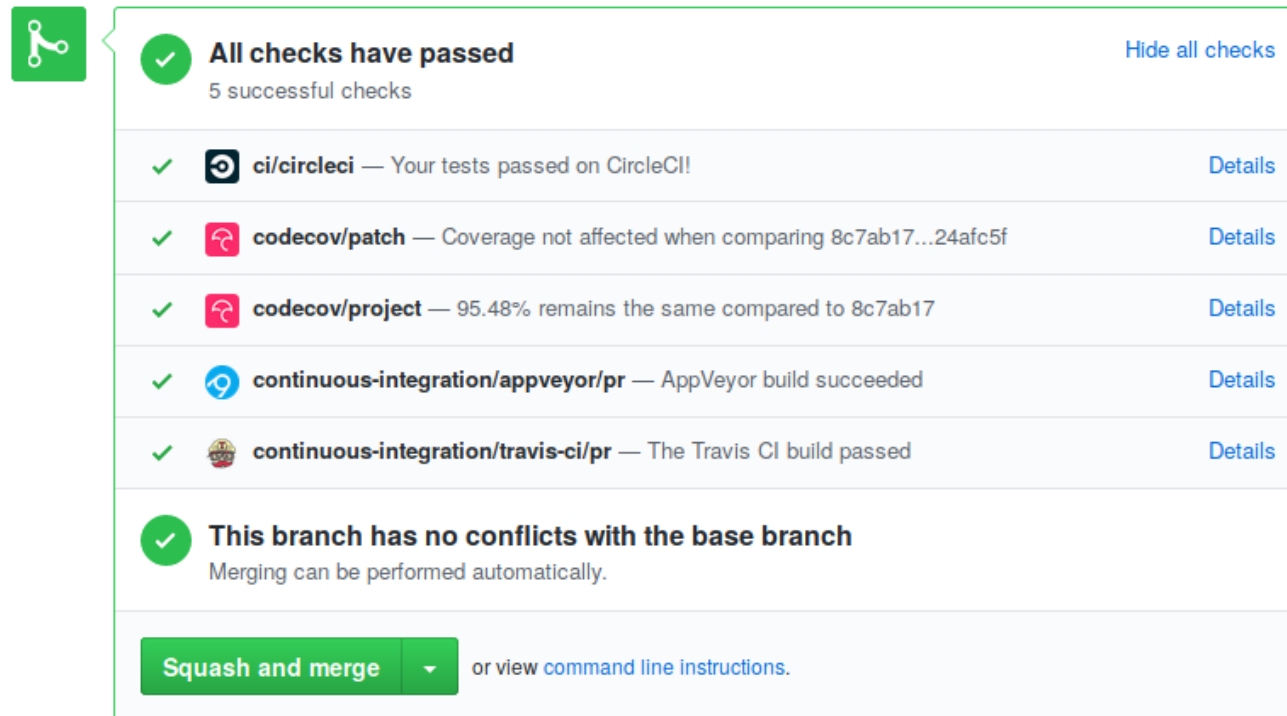
[WIP] = work in progress

[MRG] = ready to merge









# Regression tests

- Are mandatory! For everything (except documentation changes)!
- Make sure continuous integration passes:




The screenshot shows a GitHub pull request status bar with a green checkmark icon on the left. The main status is "All checks have passed" with a link to "Hide all checks". Below this, there are five individual check items, each with a green checkmark, an icon, a description, and a "Details" link. The checks are: 1. CircleCI (ci/circleci) - "Your tests passed on CircleCI"; 2. Codecov (codecov/patch) - "Coverage not affected when comparing 8c7ab17...24afc5f"; 3. Codecov (codecov/project) - "95.48% remains the same compared to 8c7ab17"; 4. AppVeyor (continuous-integration/appveyor/pr) - "AppVeyor build succeeded"; 5. Travis CI (continuous-integration/travis-ci/pr) - "The Travis CI build passed". At the bottom, there is a green "Squash and merge" button with a dropdown arrow, followed by the text "or view command line instructions."

 **✓ All checks have passed** [Hide all checks](#)  
5 successful checks

✓	 <b>ci/circleci</b> — Your tests passed on CircleCI	<a href="#">Details</a>
✓	 <b>codecov/patch</b> — Coverage not affected when comparing 8c7ab17...24afc5f	<a href="#">Details</a>
✓	 <b>codecov/project</b> — 95.48% remains the same compared to 8c7ab17	<a href="#">Details</a>
✓	 <b>continuous-integration/appveyor/pr</b> — AppVeyor build succeeded	<a href="#">Details</a>
✓	 <b>continuous-integration/travis-ci/pr</b> — The Travis CI build passed	<a href="#">Details</a>

**✓ This branch has no conflicts with the base branch**  
Merging can be performed automatically.

**Squash and merge**  or view [command line instructions](#).

# What's next?

- Wait for reviews (be patient).
- Address review comments in the same branch.
- Pushing to your fork will update the PR
- Reviewers will “approve” PR or change title to [MRG + 1]
- You need two approvals for a merge.

# Finding Issues

- Check “need contributor”, “easy” and “sprint” issues.
- Something unclear in the docs? Fix it!
- Can’t fix something that’s unclear? Open an issue!
- Problem that you keep running into: Open an issue!
- Find stalled PRs (the author didn’t address reviews for  $\sim >1$  month) and continue them!
- <https://github.com/WiMLDS/nyc-2019-scikit-sprint/projects/1>
-

# Reviewing

- You can review PRs and issues!
- Some bugs are not confirmed. See if you can confirm them and under what conditions?
- You can review documentation PRs for language and whether they are clear to you.
- You can review code changes on whether they address the issue (might be a bit tricky).
- Don't be afraid to ask clarifying questions!

# Final words

- Pick something TRIVIALY SIMPLE as the first contribution.
- You can do the cool stuff afterwards!
- There might be interesting issues that are not appropriately tagged.

Thank you for your help!  
Enjoy!