

The initial iteration will be over the linear models.

1. Choose some datasets for benchmarking the regression problem. These need to explore as many of the possible gotchas as we can: Maybe use our generators.

\*\* Shape / simple structure: wide X, tall X, sparse X, etc. \*\* Mathematic structure: conditioning number, local optima, spectrum shape

2. Set up a (pilot) benchmark runner using these datasets. This will slowly build up into a nice speed.pypy -like (but hopefully cleaner) interface so we can monitor the overall performance of the scikit.

\*\* vbench \*\* cProfile / lineprofiler output (or automatically redacted output, ie. top-k worst lines) \*\* buildbot

3. Lose nights obsessing over getting the plot to go lower and lower.

\*\* explore the possibility of cython + openmp on all platforms.