

Probabilistic coresets

We scale perceived distances of the core-set algorithm by a factor of uncertainty and search for low-confidence configurations, finding significant improvements in sample efficiency across CIFAR10/100 and SVHN image classification.

Installation:

```
pip3 install -r requirements.txt
```

To look at our raw results:

See: - [results/cifar10/main/results.json](#) - [results/cifar100/main/results.json](#) - [results/svhn/main/results.json](#)

To replicate the main findings:

Copy and save the raw result files described in the previous section. Note that the following script deletes them. Do:

```
cd src/

rm ../results/cifar10/main/*
python3 train.py --param_path ../config/cifar10/main.json --download 1

rm ../results/cifar100/main/*
python3 train.py --param_path ../config/cifar100/main.json --download 1

rm ../results/svhn/main/*
python3 train.py --param_path ../config/svhn/main.json --download 1
```

Note that the training script asserts that no corresponding result/log files exist already (to avoid overwriting), so the above script removes them prior to running `src/train.py`.

Key implementations:

See [src/acquire/lc_beam_pweighted_coreset.py](#) and [src/acquire/lc_beam_coreset.py](#) for our efficient, novel implementations. Greedy core-set is implemented in [src/acquire/greedy_coreset.py](#).