

The code is released exclusively for review purposes with the following terms: PROPRIETARY AND CONFIDENTIAL. UNAUTHORIZED USE, COPYING, OR DISTRIBUTION OF THE CODE, VIA ANY MEDIUM, IS STRICTLY PROHIBITED. BY ACCESSING THE CODE, THE REVIEWERS AGREE TO DELETE THEM FROM ALL MEDIA AFTER THE REVIEW PERIOD IS OVER.

We provide code for LIME, LIME/Smooth and LINEX/random explanations and evaluate them. The code is set up for IRIS dataset, but can easily generalize to other tabular, text, and image datasets.

The computation pipeline has 10 components (see the `Tabular` directory) and after running the first notebook `0_create_config_files.ipynb` to create the configuration files, the python scripts can be automatically executed for IRIS dataset by running the shell script `iris_pipeline_all.sh`. The configurations for the experiments are stored in `config/IRIS` directory. The next two notebooks can be executed sequentially to obtain the final result figures.

- `0_create_config_files.ipynb` - Create the necessary configuration files
- `1_data_preproc.py` - Load and preprocess the dataset
- `2_train_bb_model.py` - Train a black box model
- `3_create_base_perturbations.py` - Create the initial base perturbations for all test examples
- `4_create_neighborhood_bb_preds.py` - Estimate the black box predictions for the neighborhoods of all test points
- `5_create_env_perturbations.py` - Create different environments from the base perturbations
- `6_LIME_explanations.py` - Compute LIME and LIME/Smooth explanations
- `7_LINEX_explanations.py` - Compute LINEX explanations
- `8_metrics.ipynb` - Compute all performance measures for the various experiments
- `9_collate_plot_metrics.ipynb` - Aggregate and plot the performances

Tested Environment: Linux Ubuntu 18.04 running in a machine with 56 cores and 242 GB RAM. Used Anaconda Python 3.7.