Input Parameters for CLEAR

General input- parameters for CLEAR

case_study	The name of the UCI dataset that will be used to create the MLP (the explanandum). There are four options (i) Census (ii) Pima Indians Diabetes (iii) Credit Card (iv) BreastC.
max_predictors	The maximum number of independent variables to be used in a regression. The actual number used may be smaller if CLEAR determines that adding additional variables does not improve the score of its stepwise regression.
first_obs, last_obs	The index numbers of the test dataset observations whose predictions are to be explained. For example, if 'first_obs'=1 and 'last_obs' =5 then the predictions for the first five observations of the test dataset will be explained. In order to explain a single prediction, set 'first_obs' and 'last_obs' to the same number eg if 'first_obs'= 7, 'last_obs'=7 then the prediction for observation 7 will be explained.
num_samples	The total number of synthetic data points to be generated. The default value is 50,000.
regression_type	Set to either 'multiple' or 'logistic'
score_type	For logistic regression this can be set to either 'prsquared' (for McFadden's pseudo R-squared) or to 'AIC'. For multiple regression this can be set to 'adjR' (adjusted R-squared) or to AIC. In the Neurips paper prsquared and adjR were used.
regression_sample_size	The number of synthetic observations to use in a local regression (default 200)
neighbourhood_algorithm	'Balanced' or 'Unbalanced'
CLEAR_path	The working directory for CLEAR
apply_counterfactuals	True/False. Add w-counterfactuals to the neighbourhood datasets used for the local regressions
counterfactual_weights	The weighting to give to each counterfactual observations in the neighbourhood dataset
generate_regression_files	Create csv file of the neighbourhood dataset named 'local'_date/time.csv eg 'local_20190930.csv'
with_indicator_feature	True/False. Whether to include an indicator variable in the regressions
feature_with_indicator	Feature that is to have an indicator variable
indicator_threshold	The threshold for the indicator variable.

'perturb_one_feature'	True/False. Only generate w-counterfactuals for one feature. This was created just for the Census dataset, where only 'age' was perturbed. The only other numeric features was 'hoursPerweek' which did not not have sufficient variance to provide a basis for calculating w-counterfactuals.
only_feature_perturbed	This accompanies 'perrturb_one_feature' and is set to 'age' for the Census dataset.

Parameters created to demonstrate the incremental benefits of CLEAR's functionality

no_polynomials	True/False. Only perform simple regressions i.e. no 2^{nd} order or interaction terms
interaction_only	True/False. Allow interaction terms bur not 2^{nd} order terms.
no_centering	True/False. Do not center the regressions i.e. force the regression to go through the observation $\boldsymbol{x}$ whose prediction is to be explained.
LIME_comparison	True/False. Runs CLEAR using LIME's algorithms for creating the neighbourhood datasets and generating regressions