Key Input Parameters for CLEAR

General input- parameters for CLEAR

`sample_model`	This refers to the UCI dataset / model that is used if running CLEAR with Run_CLEAR_with_sample_model(). There are five options (i) Census (ii) PIMA (iii) Credit Card (iv) BreastC (v) Iris.
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 (default: 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 ECAI 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' (default: Balanced)
CLEAR_path	The working directory for CLEAR
apply_counterfactuals	True/False. Add *b*-counterfactuals to the neighbourhood datasets used for the local regressions (default: True)
counterfactual_weights	The weighting to give to each counterfactual observation in the neighbourhood dataset (default: 9).
generate_regression_files	Create csv file of the neighbourhood dataset named 'local'_date/time.csv eg 'local_20190930.csv' (default: False)
num_iterations	The number of times that CLEAR will evaluate each observation. This is useful if confidence intervals are required. However CLEAR's results have low variance when runs are repeated

	using the *same* model (note that due to random sampling, Run_CLEAR_with_sample_model() generates a different model for each iteration). Default: 1.
`binary_decision_boundary`	The decision boundary value that CLEAR takes to being used by the AI model
`multi_class_focus`	Specifies the reference class to be used with multiclass datasets. For example, with the Iris dataset, if set to 'setosa' CLEAR will explain the AI predictions for an observation with respect to probability of belonging to class 'setosa'. multi_class_focus can also be set to 'All', in which case it will provide explanations for all classes eg setosa, versicolor and virginica

Parameters controlling complexity of CLEAR's regression

no_polynomials	True/False. Only perform simple regressions i.e. no 2nd order or interaction terms (default: False)
interaction_only	True/False. Allow interaction terms bur not 2nd order terms (default: False)
centering	True/False. Center the regressions i.e. force the regression to go through the observation *x* whose prediction is to be explained (default: True).

Parameters controlling features selected in regression

include_all_numerics	True/False. Forces regression model to include all numeric features. For example, with the IRIS dataset this will force the regression to begin: $Y = k_1 + k_2$ SepalW $+ k_3$ PetalL $+k_4$ PetalW $+ k_5$ SepalL
`include_features`	True/False. Applies 'include_features_list' parameter (see below).
include_features_list	Forces regression model to include all the features included in this list