twinlab.TrainParams

class **twinlab.TrainParams**(*estimator='gaussian_process_regression',*
estimator_params=<twinlab.params.EstimatorParams object>,
input_explained_variance=None, input_retained_dimensions=None,
output_explained_variance=None, output_retained_dimensions=None, fidelity=None,
dataset_std=None, train_test_ratio=1.0, model_selection=False,
model_selection_params=<twinlab.params.ModelSelectionParams object>, seed=None)

Parameter configuration for training an emulator.

This includes parameters that pertain directly to the training of the model, such as the ratio of training to testing data, as well as parameters that pertain to the setup of the model such as the number of dimensions to retain after decomposition.

Variables:

- **estimator** (*str, optional*) – The type of estimator (emulator) to be trained. Currently only "gaussian_process_regression" is supported, which is the default value.

- **estimator_params** (*EstimatorParams, optional*) – The parameters for the Gaussian Process emulator.

- **input_retained_dimensions** (*Union[int, None], optional*) – The number of input dimensions to retain after applying dimensional reduction. Setting this to an integer cannot be done at the same time as specifying the `input_explained_variance`. The default value is `None`, which means that dimensional reduction is not applied to the input unless `input_explained_variance` is specified.

- **input_explained_variance** (*Union[float, None], optional*) – Specifies what fraction of the variance of the input data is retained after applying dimensional reduction. This must be a number between 0 and 1. This cannot be specified at the same time as `input_retained_dimensions`. The default value is `None`, which means that dimensional reduction is not applied to the input unless `input_retained_dimensions` is specified.

- **output_retained_dimensions** (*Union[int, None], optional*) – The number of output dimensions to retain after applying dimensional reduction. Setting this to an integer cannot be done at the same time as specifying the `output_explained_variance`. The default value is `None`, which means that dimensional reduction is not applied to the output unless `output_explained_variance` is

- **output_explained_variance** (*Union[float, None], optional*) – Specifies what fraction of the variance of the output data is retained after applying dimensional reduction. This must be a number between 0 and 1. This cannot be specified at the same time as `output_retained_dimensions`. The default value is `None`, which means that dimensional reduction is not applied to the output unless `output_retained_dimensions` is specified.
- **fidelity** (*Union[str, None], optional*) – Name of the column in the dataset corresponding to the fidelity parameter if a multi-fidelity model is being trained. The default value is `None`, whereby fidelity information is provided. Fidelity refers to the degree an emulator is able to reproduce the behaviour of the simulated data.
- **train_test_ratio** (*Union[float, None], optional*) – Specifies the fraction of training samples in the dataset. This must be a number beteen 0 and 1. The default value is 1, which means that all of the provided data is used for training. This is good to make the most out of a dataset, but means that it will not be possible to score or benchmark the performance of an emulator.
- **dataset_std** (*Union[Dataset, None], optional*) – A twinLab dataset object that contains the standard deviation of the training data. This is necessary when training a heteroskedastic or fixed noise Gaussian Process.
- **model_selection** (*bool, optional*) – Whether to run Bayesian model selection, a form of automatic machine learning. The default value is `False`, which simply trains the specified emulator, rather than iterating over them.
- **model_selection_params** (*ModelSelectionParams, optional*) – The parameters for model selection, if it is being used.
- **seed** (*Union[int, None], optional*) – The seed used to initialise the random number generators for reproducibility. Setting to an integer is necessary for reproducible results. The default value is `None`, which means the seed is randomly generated each time.

__init__(*estimator='gaussian_process_regression', estimator_params=<twinlab.params.EstimatorParams object>, input_explained_variance=None, input_retained_dimensions=None, output_explained_variance=None, output_retained_dimensions=None, fidelity=None, dataset_std=None, train_test_ratio=1.0, model_selection=False, model_selection_params=<twinlab.params.ModelSelectionParams object>, seed=None*)

Methods

__init__([estimator, estimator_params, ...])