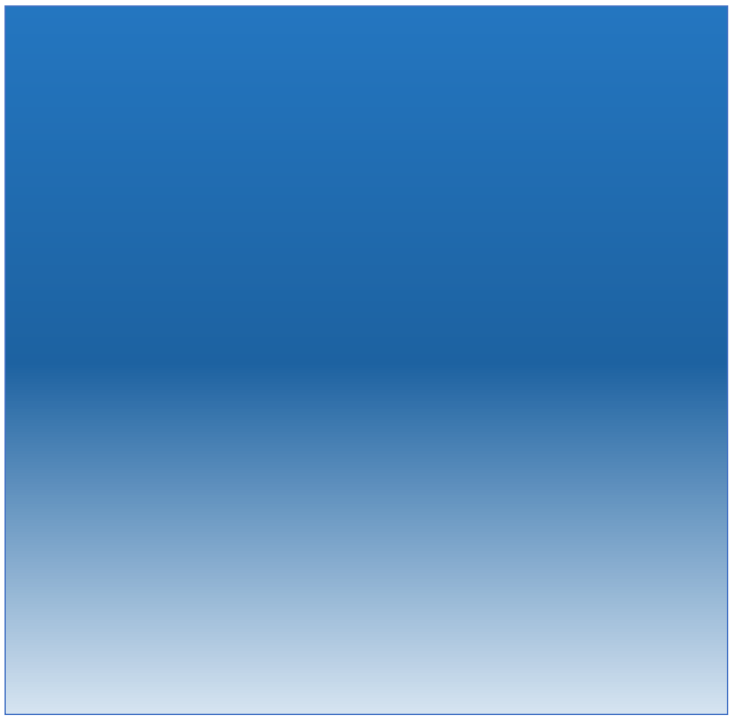

YiLong 仪陇

Package for While- and Post-Tuning Analysis

24/06/2023

Background

The purpose of this package is to cooperate with tuning packages such as JiXi and YangZhou to provide real-time analysis of tuning results, as well as post-tuning analysis.

The package performs two main functions in viewing the tuning data sorted in descending order by the validation score, and also viewing the means of evaluation metrics for combinations containing each individual value of a hyperparameter – for each hyperparameter.

Functions

Function	Purpose
`combine_tuning_results(tuning_results, output_address)`	Takes in multiple tuning output DataFrames, merging and exporting it into one file Parameters: tuning_results – list of DataFrames – at least length 2, all DataFrames must have same column name output_address – str – does not need to contain '.csv'

Class

Class	Purpose
`YiLong`	Object that reads in a tuning result DataFrame, and can display: 1. Combinations sorted by validation score 2. Means of train, validation, test scores for combinations containing each individual value of a hyperparameter – for each hyperparameter. 3. Groups of train, validation, test scores grouped by combinations containing each individual value of a hyperparameter – for each hyperparameter

Methods:

Methods	Purpose
YiLong(type)	Initialisation must input type Parameters: type – str - either 'Classification' or 'Regression'
read_tuning_result(address, extra_to_discard_columns = None)	Reads in Tuning Result DataFrame using address as guide to csv Parameters: address – str – include '.csv' extra_to_discard_columns – list – optional, default None. List of column names which are not useful hyperparameters nor evaluation metrics
read_sorted_full_df(interested_statistic = None, ascending = False)	View the Tuning Result DataFrame in sorted evaluation metric (default Validation Score, decreasing) order. Displays top 60 and bottom 60 (if len(DataFrame) <= 120 than may be overlap; if len(DataFrame) <= 60 then displays top and bottom len(DataFrame) combinations, where top and bottom dataframe is exactly the same Also displays and returns the best combination according to the evaluation metric Parameters:

	interested_statistic – str – must be a valid evaluation metric of the model type ascending – bool – default False
`read_mean_val_scores()`	View the means of evaluation metrics for combinations containing each individual value of a hyperparameter – for each hyperparameter.
`read_grouped_scores()`	View all evaluation metrics for combinations grouped by containing each individual value of a hyperparameter – for each hyperparameter If any of the individual values of a hyperparameter exceeds 60, then sample down to 60 without replacement, using seed 19861201

Objects:

Objects	Purpose
`clf_type`	Str – either ‘Regression’ or ‘Classification’
`tuning_result`	DataFrame
`hyperparameters`	List
`regression_extra_output_columns` & `classification_extra_output_columns` & `GLM_Regression_extra_output_columns`	Lists containing column names that are used internally by YiLong
`discard_columns`	List

Dependencies

pandas