

YiLong 仪陇

Package for While- and Post-Tuning Analysis

10/01/2023

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

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

Class

<u>Class</u>	<u>Purpose</u>
YiLong	<p>Object that reads in a tuning result DataFrame, and can display:</p> <ol style="list-style-type: none">1. Combinations sorted by validation score2. 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:

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

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

Objects:

<u>Objects</u>	<u>Purpose</u>
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

copy