

Manual of Tool Building Project: Configuration Performance Learning

By En Li

1. Prepare experimental data

Copying the directory datasets in the directory where the script lab2_coursework.py is located.

2. Check the software

- (1) Confirm that the Python version is 3.12.9 or above.
- (2) Make sure you have installed all dependencies that are required to run lab2_coursework.py in requirements.pdf.

3. Run lab2_coursework.py

Executing lab2_coursework.py script. This script has two optional parameters:

- d n: n is integer number, means the degree for PolynomialFeatures, default value is 2
- a f: f is float number, means alpha for Ridge function, default value is 1.0

4. Output results

The script lab2_coursework.py will show the statistical information of the results to the screen and save the baseline results and solution results into the file lab2.csv in the current directory.

The column meanings of lab2.csv are as follows:

- (1) SN: line number
- (2) System: the system name in the dataset
- (3) Dataset: the name of the file in the dataset
- (4) LR_MAPE: The MAPE result of Baseline
- (5) LR_MAE: The MAE result of Baseline
- (6) LR_RMSE: The RMSE result of Baseline
- (7) DR_MAPE: The MAPE result of Solution
- (8) DR_MAE: The MAE result of Solution
- (9) DR_RMSE: The RMSE result of Solution

5. lab2_linear.py

The script lab2_linear.py is a script that can output a scatter plot, which can be used to observe whether the performance parameters in the dataset are linearly related to the performance results. The script has two optional parameters:

- f filename: filename is the csv file name of dataset, default value is "ambivert.csv"
- c column: column is column name in the csv file, default value is "jobs"

6. lab2_huber.py

The script lab2_huber.py is used to execute Huber Regressor, without the Min-Max normalization and polynomial regression to get the results in section 5.4 of report. The usage of this script is the same as lab2_coursework.py, except that the output file is lab2_huber.csv.