

Workshop paper result discussion

Wednesday, May 15, 2024

8:49 PM

If convert more numerical features to categorical, it will have a much bigger impact on the second output variable, NP

Multi-task learning will reduce uncertainty, not necessarily the accuracy

simpleGP outperforms random forest

LMGP with 3 categorical variable will make the 1st output variable performance worse, 2nd one better

Interesting combinations

rf-no-x-scale-no-y-scale-no-cate

simpleGP-gpytorch-x-minmax-y-minmax-no-cate

multi-task-single-output-output_rank_1-x-minmax-y-minmax-no-cate

multi-task-single-output-output_rank_2-lik_rank_3-x-minmax-y-minmax-no-cate

simpleGP-gpytorch---x-minmax-y-minmax-n_cate_var_3-cate_transform_label

Output 1 accuracy is largely improved

multi-task-single-output-output_rank_2-lik_rank_0-x-minmax-y-minmax-n_cate_var_3-cate_transform_label

Uncertainty is improved from 'simpleGP-gpytorch---x-minmax-y-minmax-n_cate_var_3-cate_transform_label'

multi-task-single-output-output_rank_2-lik_rank_0-x-minmax-y-minmax-n_cate_var_3-cate_transform_LMGP

Uncertainty is further improved, but accuracy became worse

multi-task-single-output-output_rank_2-lik_rank_0-x-minmax-y-minmax-n_cate_var_1-cate_transform_partial_LMGP

Only reduce categorical variable, the accuracy of the second output is the best, the accuracy of 1st is also better than other LMGP transform

multi-task-single-output-output_rank_2-lik_rank_0-x-minmax-y-minmax-n_cate_var_2-cate_transform_partial-LMGP

Both outputs improved from the previous one

multi-task-single-output-output_rank_2-lik_rank_0-x-minmax-y-minmax-n_cate_var_3-cate_transform_partial-LMGP

Stopped to improve from previous

multi-task-single-output-output_rank_2-lik_rank_3-x-minmax-y-minmax-n_cate_var_3-cate_transform_partial-LMGP

Output 1 improved from 'multi-task-single-output-output_rank_2-lik_rank_0-x-minmax-y-minmax-n_cate_var_2-cate_transform_partial-LMGP'

multi-task-single-output-output_rank_2-lik_rank_3-x-minmax-y-minmax-n_cate_var_2-cate_transform_partial-LMGP

Output 1 is worse than 'multi-task-single-output-output_rank_2-lik_rank_3-x-minmax-y-minmax-n_cate_var_3-cate_transform_partial-LMGP'

Group by round index

multi-task-multi-output-output_rank_2-task_rank_4-lik_rank_3-x-minmax-y-minmax-n_cate_var_1-cate_transform_label

Output 1 uncertainty is very small, R2 is high

Groupby 2nd output because its range is big

Within folder opensource-splitby2ndoutput-multitask-multiout
multi-task-multi-output-output_rank_2-task_rank_3-lik_rank_1-x-minmax-y-minmax-n_cate_var_1-cate_transform_label largely reduced the 2nd output uncertainty

multi-task-multi-output-output_rank_2-task_rank_3-lik_rank_3-x-minmax-y-minmax-n_cate_var_2-cate_transform_partial-LMGP
Training R2 for both variables are the best
Average test uncertainty is also low for both outputs

Regularization, high correlation not numerical stable

Create an artificial dataset

Ask synthetic data to zhiyuan's dataset

Send PDF to Dan, abstract to ostic, full abstract submit to workshop,
Remind Zhiyuan to send manuscript to Zhiyuan's boss