06

Acceleration of feature selection with group regularization

Finding features in data fast and accurately

Abstract

Selecting important feature groups from data is one of the most fundamental tasks in data analysis. However, the computation cost is high when the size of the data is large. In this study, we propose a fast method for feature group selection without degrading the accuracy. It safely skips unnecessary computations and intensively optimizes important groups. As a result, our method is up to 35 times faster than the original method without any loss of accuracy. In addition, our method has no additional hyperparameters and no additional tuning costs. It will be possible to create value from complex and large scale data by speeding up the analysis of data with complex structures such as groups. We create value from a wide variety of data and contribute to society.

Feature group selection

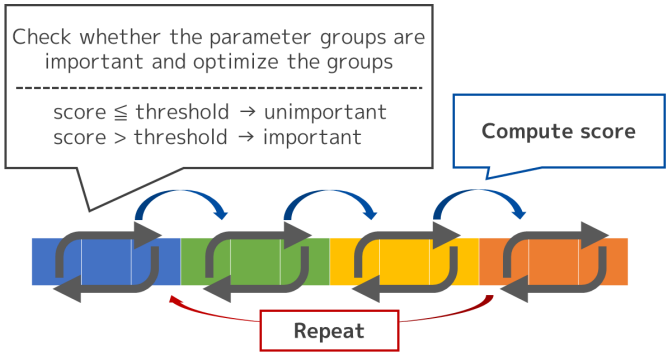
- Selecting important feature groups from data is one of the most fundamental tasks in data analysis.

e.g. finding out which areas are important for Tokyo's weather forecast from the climate data of all areas in Japan.

Which feature groups are important?

Problem: computation cost

- The cost is high when the size of the data is large.

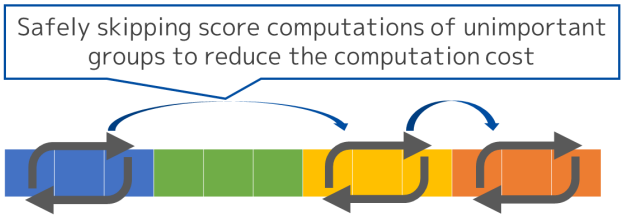


The bottlenecks are "# of score computations" and "# of iterations".

Fast method for feature group selection

- We utilize two ideas:

1. Safely skipping unnecessary computations



We efficiently identify **unimportant groups** by using the **upper bounds** of the scores.

Upper bound of score ≦ threshold → unimportant

2. Intensively optimizing important groups

Intensively optimizing important groups to reduce the value of the objective function

We efficiently identify **important groups** by using the **lower bounds** of the scores.

Lower bound of score > threshold → important

Experiments show that our method is **up to 35 times faster than the original method without any loss of accuracy.**

References

[1] Y. Ida, Y. Fujiwara, H. Kashima, "Fast Sparse Group Lasso," in *Proc. Neural Information Processing Systems (NeurIPS)*, 1700-1708, 2019.

Contact

Yasutoshi Ida / Software Innovation Center
Email: cs-openhouse-ml@hco.ntt.co.jp