Parameter Tuning

02/18/2019 Kazuki Egusa

LightGBM

✓ LEAF-wise tree growth algorithm

✗ DEPTH-wise tree growth algorithm

➔ Converge much FASTER, but OVER-FITTING

Github: https://github.com/Microsoft/LightGBM

LightGBM documentation: https://lightgbm.readthedocs.io/en/latest/

- Parameters: https://lightgbm.readthedocs.io/en/latest/Parameters.html
- Parameters Tuning: https://lightgbm.readthedocs.io/en/latest/Parameters-Tuning.html

Main Parameter

1. num_leaves

Very Important

2. min_data_in_leaf

3. max_depth

- Parameters: https://lightgbm.readthedocs.io/en/latest/Parameters.html

For Faster Speed

- Use bagging by setting bagging_fraction and bagging_freq
- Use feature sub-sampling by setting feature_fraction
- Use small max_bin
- Use save_binary to speed up data loading in future learning
- Use parallel learning, refer to Parallel Learning Guide (https://lightgbm.readthedocs.io/en/latest/Parallel-Learning-Guide.html)

For Better Accuracy

- Use large max_bin (may be slower)
- Use small learning_rate with large num_iterations
- Use large num_leaves (may cause over-fitting)
- Use bigger training data
- Try dart

Deal with Over-fitting

- Use small max_bin
- Use small num_leaves
- Use min_data_in_leaf and min_sum_hessian_in_leaf
- Use bagging by set bagging_fraction and bagging_freq
- Use feature sub-sampling by set feature_fraction
- Use bigger training data
- Try lambda_l1, lambda_l2 and min_gain_to_split for regularization
- Try max_depth to avoid growing deep tree

- Parameters: https://lightgbm.readthedocs.io/en/latest/Parameters.html