

# XGBoost, Light GBM, Catboost

## XGBoost

- 1) Базовый алгоритм приближает направление с учётом вторых производных.
- 2) Регрессия: Кол-во деревьев
- 3) Критерий информативности зависит от сдвигов.
- 4) Критерий останова зависит от сдвигов

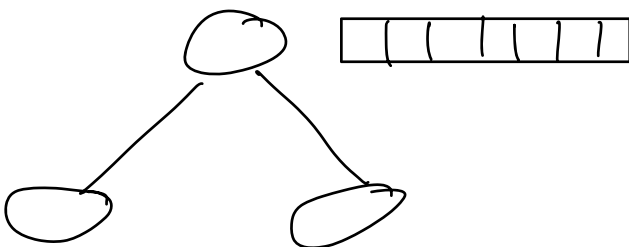
xgboost  
Dmatrix

$$\operatorname{argmin}_{b \in \mathcal{A}} \sum (b(x_i) - s_i)^2$$

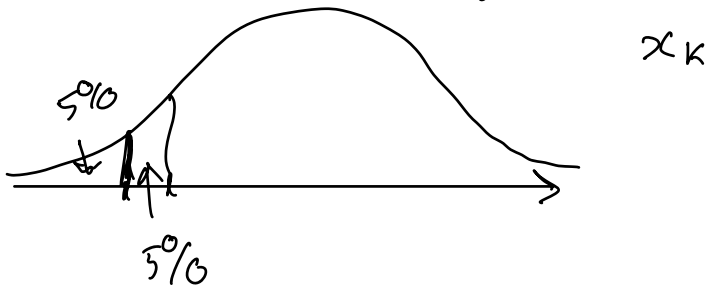
$$\left. \frac{\partial L}{\partial z} \right|_{z = a_{N-1}}$$

$$s_i \left. \frac{\partial L}{\partial z} \right|_{z = a_{N-1}(x_i)} \quad h_i = \left. \frac{\partial^2 L}{\partial z^2} \right|_{z = a_{N-1}(x_i)}$$

$$L_{\text{split}} = \frac{1}{2} \left[ \frac{(\sum_{i \in R_L} s_i)^2}{\sum_{i \in R_L} h_i + \lambda} + \frac{(\sum_{i \in R_R} s_i)^2}{\sum_{i \in R_R} h_i + \lambda} - \frac{(\sum_{i \in R} s_i)^2}{\sum_{i \in R} h_i + \lambda} \right] - \gamma$$



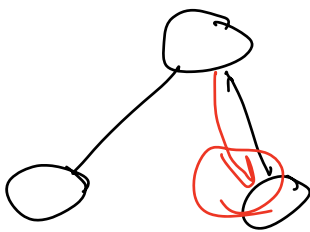
## Публицистический поиск разделения



## Разрешимость.

- 1) Много пропусков
- 2) Арифметическое кодирование
- 3) Много нулей в статистике

Целевая функция



0

0

0

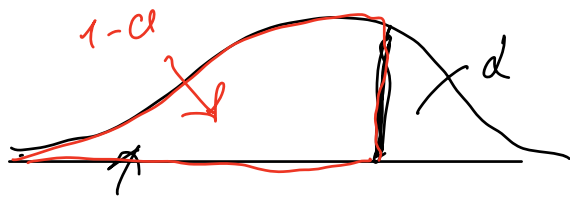
10 мин

0

0

Light GBM

Gradient-based One-side Sampling GOSS



$$\int f(x) dx = 1$$

$\alpha \times 100\%$  объектов с самыми большими градиентами

$b \times 100\%$  случайных объектов из оставшихся и добавим их градиенты на

$$\frac{1-\alpha}{b}$$

$\frac{1-\alpha}{b}$  — это наблюдения с меньшими градиентами

$$\alpha = 0.4$$

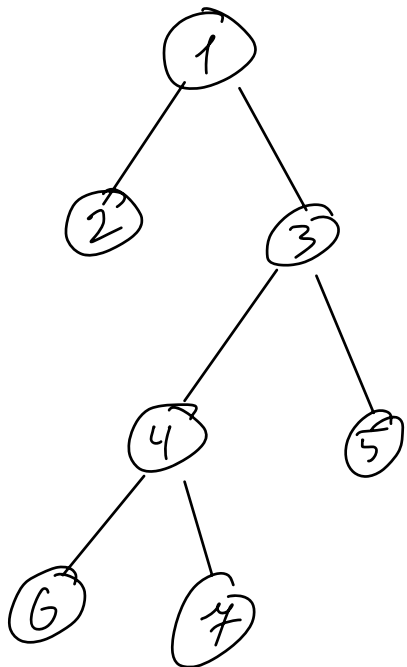
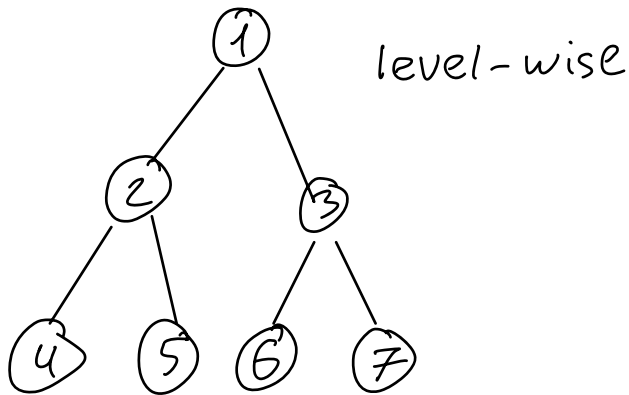
$$b = 0.4$$

Exclusive Feature Bundling.

0	0	0
0	0	0
1	0	1
0	1	0

Рассмотрим меньшее количество признаков, учитывая уникальные кортежи значений

Правила роста дерева:



1) Получаются несбалансиров. деревья

2) Следующим шагом минимизации исходной функции потерь

Catboost

- 1) Обработка категориальных признаков
- 2) Oblivious trees
- 3) Хорошо работает из коробки

- 1) one-hot-size- $N = 2$
- 2) target-encoding
- 3) Комбинированная ком. формул

