

Quiz 3: градиентный бустинг

Стакать xГбусты гыгыы brrrrr GOOOOOOOOOOOO

Typical kaggle Chad

Решите все задания. Все ответы должны быть обоснованы. Решения должны быть прописаны для каждого пункта. Рисунки должны быть чёткими и понятными. Все линии должны быть подписаны.

[2] Задание 1

Допустим, мы обучили две композиции решающих деревьев: случайный лес и градиентный бустинг. В обоих случаях мы остановили добавление деревьев, когда ошибка на валидационной выборке перестала убывать. Что произойдёт с ошибкой этих моделей на валидационной выборке, если мы продолжим обучение и добавим в каждую из них по 10000 деревьев? Обоснуйте ваш ответ.

[2] Задание 2

Как решать задачу бинарной классификации с помощью градиентного бустинга (что нужно поменять по сравнению с регрессией)?

[3] Задание 3

У Якоба есть пуассоновская функция потерь $L(y, z) = -yz + \exp z$. Он использует её для обучения градиентного бустинга.

- Запишите формулу для подсчёта сдвигов;
- Выпишите задачу, которая будет решаться при поиске каждого нового базового алгоритма $b_N(x)$.

[3] Задание 4

В чём заключается идея переоббора прогнозов в листьях базовых деревьев в градиентном бустинге? Запишите оптимизационную задачу поиска прогноза в конкретном листе. Для чего это делается?

[0.1] Задание 5

Чем typical kaggle chad отличается от simple ml virgin?