Методы регуляризации решающих деревьев

Деревья легко переобучаются и процесс ветвления надо в какой-то момент останавливать.

В основе L1-регуляризации лежит достаточно простая идея. Как и в случае L2-регуляризации, мы просто добавляем штраф к первоначальной функции затрат. Подобно тому, как в L2-регуляризации мы используем L2-нормирование для поправки весовых коэффициентов, в L1-регуляризации мы используем специальное L1-нормирование. L2-регуляризация называется также регрессией Риджа, а L1-регуляризация – лассо-регрессией.

J_{RIDGE} = sum_{i=1}^{N} (y_n - widehat y_n)^2 + lambda parallel wparallel^2_2, 

J_{LASSO} = sum_{i=1}^{N} (y_n - widehat y_n)^2 + lambda parallel wparallel_1. 

L2-регуляризация

Суть в том, что чересчур тяжёлые весовые коэффициенты «отталкивают» нашу линию наилучшего соответствия, построенную на основе минимизации квадрата ошибок, от основной тенденции.

Суть состоит в том, что мы изменяем нашу первоначальную функцию, добавляя «штраф» на большие весовые коэффициенты. Для этого мы добавляем постоянную , умноженный на квадрат *w.*

Когда мы хотим, чтобы модель работала правильно, мы определяем функцию потерь. Эта функция потерь будет определять производительность модели по данным путем расчета потерь. Нам нужно минимизировать потери, чтобы найти нужную модель. Для этого регуляризация добавляет лямбда, чтобы оштрафовать функцию потерь. Мы получаем оптимальное решение от этой методики, так как она отвергает высокие ошибки обучения при меньших значениях лямбда и отвергает более сложные модели с большими значениями лямбда.

2.

**Важность** **признака** **рассчитывается** с учетом увеличения ошибки. Если удаление признака увеличивает ошибку в моде­. ли, этот признак важен.