

- Пусть уже построено $(m - 1)$ деревьев в ансамбле
- $y_i = yTrain_i$ - правильный ответ на x_i объекте
- $answers_i$ - ответ $(m - 1)$ моделей на x_i
- Строим m -ое дерево, фитим его на $[x_i, sign(y_i - answers_i)]$
- $h_i = h(x_i)$ - ответ m -ого дерева на x_i объекте

Хотим

$$\sum_{i=1}^{i=n} |y_i - answers_i - \rho * h_i| \rightarrow \min_{\rho}$$

- Пусть R_{km} - k -ый лист m -ого дерева,
- $leafValue(R_{km})$ - значение k -ом листе m -ого дерева

$$h_i = \sum_{R_{km}} [I(x_i \text{ in } R_{km}) * leafValue(R_{km})]$$

То есть

$$\sum_{i=1}^{i=n} |y_i - answers_i - \rho * h_i| = \sum_{i=1}^{i=n} |y_i - answers_i - \rho * \sum_{R_{km}} [I(x_i \text{ in } R_{km}) * leafValue(R_{km})]|$$

Минимизация суммы таких модулей равносильна минимизации суммы модулей в каждом листе m -ого дерева, т.е. для каждого листа получим задачу:

$$\sum_{j=1}^{j=l} |a_j - b| \rightarrow \min_b$$

$$\frac{\partial}{\partial b} \left[\sum_{j=1}^{j=l} |a_j - b| \right] = \sum_{j=1}^{j=l} [sign(a_j - b)] = 0 \rightarrow$$

\rightarrow половина слагаемых равна единице, а другая - минус единице \rightarrow

\rightarrow половина слагаемых больше b и половина меньше либо равна $b \rightarrow$

$\rightarrow b = median(a_1, a_2, \dots, a_l)$

В нашем случае $a = (y_i - answers_i)$

То есть для каждого листа R_{km} присваиваем

$$leafValue(R_{km}) = \underset{i: x_i \in R_{km}}{median}(y_i - answers_i)$$