# 1 Теория

Предположим, что у нас есть переменная $Y$, стохастически зависящая от вектора переменных $X_1, \dots, X_n$, функции $G_1(x)$, $G_2(x)$, детерминировано зависящие от вектора переменных $X_1, \dots, X_n$.

Имеется выборка $\tilde{S} = \{(y_j, x_j, G_1(x_j), G_2(x_j)), j = \overline{1, m}\}$, где

- $y_j$ - значение переменной $Y$ для объекта с номером $j$
- $x_j = (x_{j1}, \dots, x_{jn})$ - вектор значений признаков $X_1, \dots, X_n$ для объекта с номером $j$
- $G_1(x_j)$ - значение функции $G_1$ в точке $x_j$
- $G_2(x_j)$ - значение функции $G_2$ в точке $x_j$

Предлагается построить дерево $T(x)$, для которого достигается минимум функционала:

$$\Phi(\tilde{S}, T) = \Sigma_{j=1}^{m}\{\gamma_1[T(x_j) - y_j]^2 + \gamma_2[T(x_j) - G_2(x_j)]^2 - \mu[T(x_j) - G_1(x_j)]^2\}$$

где $\gamma_1 + \gamma_2 = 1; \quad \gamma_1, \gamma_2, \mu \in [0, 1]$

Как видно из структуры функционала $\Phi$, дерево $T$, соответствующее его минимуму, будет аппроксимировать связь $Y$ с переменными $X_1, \dots, X_n$ при $\gamma_1 > 0$.

Одновременно дерево $T(x)$ будет удаляться от зависимости $G_2(x)$ при возрастании $\mu$ и приближаться к зависимости $G_1(x)$ при возрастании $\gamma_2$.

## 1.1 Реализация дерева

При построении дерева был использован "жадный" метод оптимизации целевого функционала: на каждом шагу к дереву добавляется узел, обеспечивающий наибольшее снижение используемого функционала $\Phi$.

Предположим, что на каком-то шаге дерево $T_k$ содержит $k$ концевых узлов, которым соответствуют концевые выборки $S_1^k, \dots, S_k^k$.

Новое дерево $T_{k+1}$ строится через добавление к дереву $T_k$ дополнительного узла $u$.

Узел $u$ получается из некоторого концевого узла $g$ с помощью порогового правила вида $X_u > \delta_u$, где $X_u$ и $\delta_u$ признак и порог к нему соответственно.

Правило $X_u > \delta_u$ расщепляет выборку $S_g^k$ на две подвыборки.

Признак $X_u$ и порог $\delta_u$ ищутся из условия максимизации разности $\Phi(\tilde{S}, T_k) - \Phi(\tilde{S}, T_{k+1})$.

Процесс построения может быть прекращен при выполнении одного из перечисленных условий:

- На очередном шаге не удается уменьшить функционал
- На очередном шаге произошло изменение функционала, меньшее чем некоторое пороговое значение
- Кол-во объектов внутри узла меньше некоторого порогового значения

## 1.2 Вывод формул

### 1.2.1 Выбор оптимального предсказания для терминального узла дерева

Для вывода формул обобщим функцию потерь, представив её как взвешенную сумму $MSE$ для разных векторов $y$. Все различные вектора $y$ запишем в столбцы матрицы $G$. $\alpha$ - вектор весов, значения которых могут быть отрицательными, но сумма всех коэффициентов должна быть положительной.

$$\Phi(X,T) = \Sigma_{k=1}^{n}\Sigma_{j=1}^{m}\alpha_k(G_{jk} - T(X_j))^2$$

Пусть по некоторому из правил текущий узел определен как терминальный. Найдем минимизирующее функционал предсказание.

$$\Phi(T) = \Sigma_{k=1}^{n}\Sigma_{j=1}^{m}\alpha_k(G_{jk} - T)^2$$

$$\Phi'(T) = 2\Sigma_{k=1}^{n}\Sigma_{j=1}^{m}\alpha_k(T - G_{jk}) =$$

$$= 2\Sigma_{k=1}^{n}\Sigma_{j=1}^{m}\alpha_k T - 2\Sigma_{k=1}^{n}\Sigma_{j=1}^{m}\alpha_k G_{jk} =$$

$$= 2mT\Sigma_{k=1}^{n}\alpha_k - 2\Sigma_{k=1}^{n}\alpha_k\Sigma_{j=1}^{m}G_{jk}$$

$$\Rightarrow T_\star = \frac{\Sigma_{k=1}^{n}\alpha_k\Sigma_{j=1}^{m}G_{jk}}{m\Sigma_{k=1}^{n}\alpha_k}$$

### 1.2.2 Оптимальный подсчет функции потерь для всех порогов одновременно

Для каждого нетерминального узла необходимо решать задачу выбора признака и порога, по которым будет происходить его разбиение.

Если подсчитывать функцию потерь для каждой пары признак - порог, то итоговая асимптотика составит $N * M^2 * F$, где $M$ - кол-во объектов в узле, $N$ - кол-во компонент в функции потерь, $F$ - кол-во признаков.

Предложенный ниже метод подсчета функции потерь снижает асимптотику вычислений до $N * M * F$, а также позволяет вычислять значение функционала с помощью векторных инструкций одновременно для всех порогов, что существенно ускоряет вычисления.

Распишем значение функционала в случае константного оптимального предсказания:

$$\Phi(G) = \Sigma_{k=1}^{n}\alpha_k\Sigma_{j=1}^{m}(G_{jk} - T_\star)^2 =$$

$$= \Sigma_{k=1}^{n}\alpha_k\Sigma_{j=1}^{m}G_{jk}^2 - 2T_\star\Sigma_{k=1}^{n}\alpha_k\Sigma_{j=1}^{m}G_{jk} + mT_\star^2\Sigma_{k=1}^{n}\alpha_k =$$

$$= \Sigma_{k=1}^{n}\alpha_k\Sigma_{j=1}^{m}G_{jk}^2 - 2mT_\star^2\Sigma_{k=1}^{n}\alpha_k + mT_\star^2\Sigma_{k=1}^{n}\alpha_k =$$

$$= \Sigma_{k=1}^{n}\alpha_k\Sigma_{j=1}^{m}G_{jk}^2 - mT_\star^2\Sigma_{k=1}^{n}\alpha_k =$$

$$= \Sigma_{k=1}^{n}\alpha_k\Sigma_{j=1}^{m}G_{jk}^2 - \frac{1}{m\Sigma_{k=1}^{n}\alpha_k}(\Sigma_{k=1}^{n}\alpha_k\Sigma_{j=1}^{m}G_{jk})^2$$

После разбиения узла по порогу на две части, в каждой из них будет своё оптимальное предсказание, а значение функции потерь для исходного узла будет суммой значений функции потерь двух получившихся узлов.

$$\Phi_{split} = \Phi(G_{left}) + \Phi(G_{right}) =$$

$$= \Sigma_{k=1}^{n}\alpha_k\Sigma_{j=1}^{m}G_{jk}^2 - \frac{1}{|L|\Sigma_{k=1}^{n}\alpha_k}(\Sigma_{k=1}^{n}\alpha_k\Sigma_{j\in L}G_{jk})^2 - \frac{1}{|R|\Sigma_{k=1}^{n}\alpha_k}(\Sigma_{k=1}^{n}\alpha_k\Sigma_{j\in R}G_{jk})^2$$

где |L| и |R| - кол-ва объектов в левом и правом узлах после разбиения соответственно.

Отбросив часть, независящую от разбиения ( $\Sigma_{k=1}^{n}\alpha_k\Sigma_{j=1}^{m}G_{jk}^2$ ), и домножив на отрицательную константу ( $-\Sigma_{k=1}^{n}\alpha_k$ ) получим функционал, который необходимо максимизировать:

$$\frac{1}{|L|}(\Sigma_{k=1}^{n}\alpha_k\Sigma_{j\in L}G_{jk})^2 + \frac{1}{|R|}(\Sigma_{k=1}^{n}\alpha_k\Sigma_{j\in R}G_{jk})^2$$