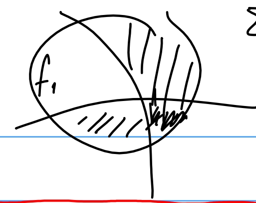
$h(\bar{x})$

$\approx$

$$F(\bar{x}) = \underbrace{f_1(\bar{x}) + f_2(\bar{x}) + \ldots + f_{n-1}(\bar{x})} + f_n(\bar{x})$$

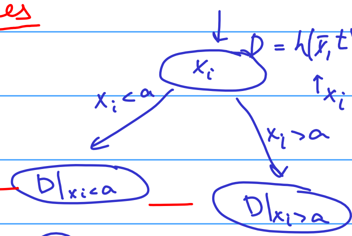
AdaBoost

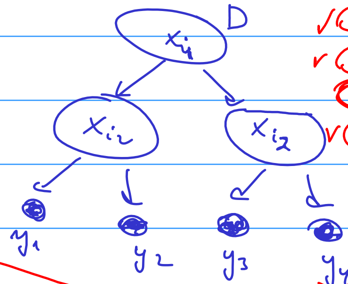


$\sum \alpha_i f_i(\bar{x})$

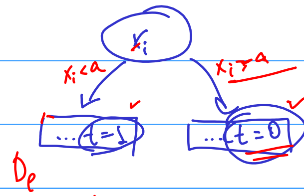
Decision trees

$D = \{(\bar{x}, t)\}$





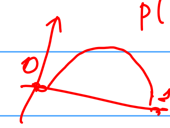
1. Как выбирать x_i?
2. Как делать расщепление?
3. Когда останавливаться?
4. Как предсказывать в листьях?



$\{(\bar{x}, t) \in D_2 \mid t = 0\}$

$D|_{x_i > a} = D_2 = D_{20} \cup D_{21}$, $D_{21} = \{\ldots\ t=1\}$

$p(1-p) = p_0 p_1$

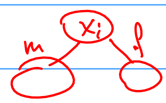


H_ℓ $\quad \underbrace{\frac{|D_\ell|}{|D|} H_\ell + \frac{|D_2|}{|D|} H_2} \to \min$

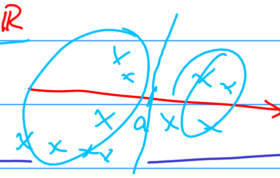
$$H_2 = -p_0 \log p_0 - p_1 \log p_1 = -\frac{|D_{20}|}{|D_2|} \cdot \log \frac{|D_{20}|}{|D_2|} - \frac{|D_{21}|}{|D_2|} \log \frac{|D_{21}|}{|D_2|} \to \min$$

Gini criterion

$x_i = \{m, f\}$



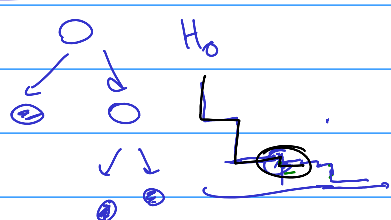
$x_i \in \mathbb{R}$

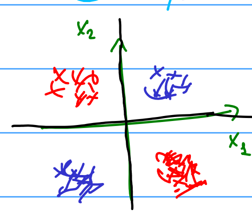


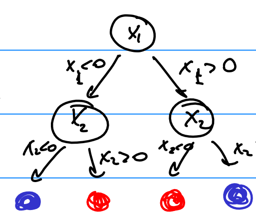
$L(T) + \lambda |T|$

$H(T) + \lambda |T|$

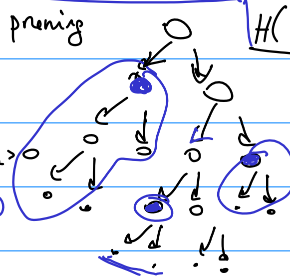
$\downarrow$ min



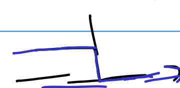




pruning



$J(f_n) = \sum [y_n \neq t_n] w_n$



$\ell(f_n)$

$$\frac{\partial \ell}{\partial f_n(\bar{x}_i)}$$

$$F_K(\bar{x}) = \sum_{k=1}^{K} f_k(\bar{x}) \qquad F_{k-1}(\bar{x}) = f_1 + f_2 + \ldots + f_{k-1}$$

$F_n(\bar{x}) \approx y$

$\boxed{f_n(\bar{x}) \approx y - F_{n-1}(\bar{x})}$

$$\boxed{F_n(\bar{x})} = f_1 + \ldots + f_{n-1} + f_n(\bar{x}) = \underbrace{F_{n-1}(\bar{x})}_{\text{fixed}} + f_n(\bar{x}) = \hat{y}_i(\bar{x})$$

$$L(F_n) = \sum_{i=1}^{N} \ell(y_i, \hat{y}(\bar{x}_i)) + \sum_{k=1}^{n} \Omega(f_k) \xrightarrow[f_n]{} \min$$

$= \ell n(\hat{y}_{n-1} + f_n)$

$y_n \ln \hat{y}_n + (1-y_n) \ln(1-\hat{y}_n)$

$\boxed{\ell(y_i, \hat{y}(\bar{x}_i)) = (y_i - \hat{y}(\bar{x}_i))^2}$

$$L(F_n) = \sum_{i=1}^{N} \underbrace{(y_i - \hat{y}(\bar{x}_i))^2} + \underbrace{\sum_{k=1}^{n} \Omega(f_k)}_{\Omega(f_n) + \text{const}}$$

$$\ell(y_i, \overbrace{\hat{y}_{n-1}(\bar{x}_i)}^{\text{const}} + f_n(\bar{x}_i)) = \ell(y_i, \hat{y}_{n-1}(\bar{x}_i)) + \underbrace{\frac{\partial \ell}{\partial f_n(\bar{x}_i)}\Big|_{(y_i, \hat{y}_{n-1})}}_{g_i} \cdot f_n(x_i) + \frac{1}{2} \underbrace{\frac{\partial^2 \ell}{\partial f_n(x_i)^2}\Big|_{(y_i, \hat{y}_{n-1})}}_{h_i} \cdot f_n^2(x_i) + \ldots$$

$\ln(1 - \hat{y}_{n-1} - f_n)$

$$\frac{y_n}{\hat{y}_{n-1} + f_n} - \frac{1 - y_n}{1 - \hat{y}_{n-1} - f_n} = g_i$$

h_i

$$\big((y_i - F_{n-1}(\bar{x}_i)) - f_n(\bar{x}_i)\big)^2 = (y_i - \cancel{F_{n-1}(\bar{x}_i)})^2 - 2(y_i - F_{n-1}(\bar{x}_i)) f_n(\bar{x}_i) + f_n^2(\bar{x}_i)$$

const

$$L(F_n) = \text{Const} + \sum_{i=1}^{N} \left(f_n^2(\bar{x}_i) - 2(y_i - F_{n-1}(\bar{x}_i)) f_n(\bar{x}_i) \right) + \Omega(f_n) \to \min$$

$$\boxed{L(f_n) = \sum_{i=1}^{N} \left(g_i \cdot f_n(\bar{x}_i) + \frac{1}{2} h_i \cdot f_n^2(\bar{x}_i) \right) + \Omega(f_n)}$$

$$L(f_n) = \sum_{i=1}^{N} \left(g_i f_n(\bar{x}_i) + \frac{1}{2} h_i f_n^2(\bar{x}_i) \right) + \Omega(f_n)$$

$h_i = 2, \quad g_i = -2(y_i - F_{n-1}(\bar{x}_i))$

$$\hat{y}^{(t)}(\bar{x}_i) = \hat{y}^{(t-1)}(\bar{x}_i) + f_t(\bar{x}_i) \qquad g_i,\ h_i = \ldots\ \hat{y}^{(t-1)}(x_i),\ y_i,\ \bar{x}_i \ldots$$

$$L(f_t) = \sum_{i=1}^{N}\left(g_i \cdot f_t(\bar{x}_i) + \frac{1}{2} h_i \cdot f_t^2(\bar{x}_i)\right) + \Omega(f_t) \longrightarrow \min$$

$$\sum_{j=1}^{T} \sum_{i: q(\bar{x}_i)=j}$$

T - размер дерева

$w_1, w_2, \ldots, w_T$ - знач. в листьях

$q: \bar{x} \to \{1, \ldots, T\}$

$f_t(\bar{x})$

$w_1 \ldots w_T$

$$= \sum_{i=1}^{N}\left(g_i \cdot w_{q(\bar{x}_i)} + \frac{1}{2} h_i w^2_{q(\bar{x}_i)}\right) + \lambda T + \frac{1}{2}\gamma \sum_{j=1}^{T} w_j^2 =$$

$$= \sum_{j=1}^{T} \sum_{i:q(\bar{x}_i)=j}\left[g_i w_j + \frac{1}{2} h_i w_j^2\right] + \lambda T + \frac{1}{2}\gamma \sum w_j^2$$

$$= \sum_{j=1}^{T}\left[\left(\sum_{i:q(\bar{x}_i)=j} g_i\right) \cdot w_j + \frac{1}{2}\left(\sum_{i:q(\bar{x}_i)=j} h_i + \gamma\right) \cdot w_j^2\right] + \lambda T = \sum_{j=1}^{T}\left[G_j \cdot w_j + \frac{1}{2}(H_j + \gamma) w_j^2\right] + \lambda T$$

" G_j $\quad H_j$

$$G_j = \sum_{i:q(x_i)=j} g_i$$

fix q:

$$L(w_1 \ldots w_T) = \sum_{j=1}^{T}\left[G_j w_j + \frac{1}{2}(H_j + \gamma) w_j^2\right] \to \min \qquad \boxed{w_j^* = -\frac{G_j}{H_j + \gamma}}$$

$\to \min$

(t): $\forall$ структура q $\longrightarrow w_j^* = -\frac{G_j}{H_j + \gamma} \longrightarrow \tilde{L}(q) = L(w_1^* \ldots w_T^*) = \boxed{\sum_{j=1}^{T} \frac{-G_j^2}{2(H_j + \gamma)} + \lambda T}$

$G_1 H_1$, $G_2 H_2$, w_j, (G_j, H_j), (G_j^L, H_j^L), (G_j^R, H_j^R)

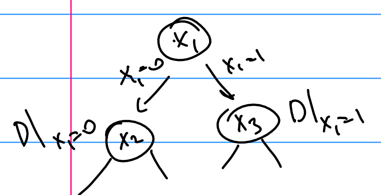
$$L_{before} = -\frac{1}{2}\sum_{j=1}^{T} \frac{G_j^2}{H_j + \gamma} + \lambda T$$

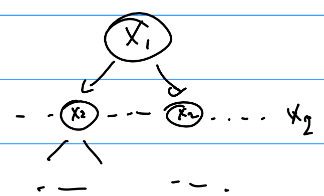
$$L_{after} = -\frac{1}{2} \cancel{\sum_{j' \neq j} \frac{G_{j'}^2}{H_{j'} + \gamma}} - \frac{1}{2}\left[\frac{G_j^{L\,2}}{H_j^L + \gamma} + \frac{G_j^{R\,2}}{H_j^R + \gamma}\right] + \lambda(T+1)$$

$$\boxed{f_t(\bar{x}_i) \approx y_i - \hat{y}^{(t-1)}(\bar{x}_i)}$$

$$\boxed{Gain = -\frac{1}{2}\left[\frac{G_j^2}{H_j^2 + \gamma} - \frac{G_j^{L\,2}}{H_j^L + \gamma} - \frac{G_j^{R\,2}}{H_j^R + \gamma}\right] - \lambda}$$

xgboost





gender: $[m, f]$ $\longrightarrow$ gender: m, f

city: [Kyiv, NY, ...] $\quad$ 10000 $\quad$ city

TBS - target-based statistics $\quad$ $y \in \mathbb{R}$ - income

city $\in \mathbb{R}$

$p(\hat{x}_i | y_i)$ $\qquad$ $\hat{x}_i = y_i$ $\qquad$ $(\ldots\ x_i\ \ldots)$ $\quad$ y_i

$(\ldots\ \hat{x}_i\ \ldots)$ $\quad$ y_i

$$\hat{x}_i = \frac{\sum_j x_j + aP}{N_{y_k} + a} \qquad \hat{x}_i = \frac{y_i + aP}{a}$$

$$\hat{x}_k = \frac{\sum_{j \in D_k} y_j [x_j = x_k] + aP}{\sum_{j \in D_k}[x_j = x_k] + a}$$

$k \notin D_k$: $\quad$ - $D = D_0 \cup D_1$ $\qquad$ $\bar{x}^1, \bar{x}^2, \bar{x}^3, \ldots, \bar{x}^{k-m}, \ldots, \bar{x}^{k-1}, \bar{x}^k, \bar{x}^{k+1}$ $\quad$ m

- $D_k = D \setminus \{k\}$ $\qquad$ $D_k \subseteq \{1, 2, \ldots, k-1\}$

$\sigma: \{1..N\} \to \{1..N\}$ $\qquad$ $\bar{x}^{\sigma(1)}, \bar{x}^{\sigma(2)}, \ldots$ $\quad$ $\bar{x}^{\sigma(k)}$ $\qquad$ $D_{\sigma(k)} = \{\sigma(1), \ldots, \sigma(k-1)\}$