Оптимизация 2ого порядка в GB.

$$\sum_{i=1}^{n} L(y_i, a_{N-1}(x_i) + b_N(x_i)) \to \min_{b_N}$$

$$a_N(x_i) = \sum_{j=1}^{N} b_j(x_i)$$

$$a_N(x_i) = \sum_{j=1}^{N} \gamma_j b_j(x_i)$$

1) $$s_i^{(N)} = - \frac{\partial L(y_i, z)}{\partial z} \Big|_{z = a_{N-1}(x_i)}$$

2) $$b_N(x) = \arg\min_{b \in A} \sum_{i=1}^{n} (b(x_i) - s_i^{(N)})^2$$

3) $$\gamma_N = \arg\min_{\gamma \in \mathbb{R}} \sum_{i=1}^{n} L(y_i, a_{N-1}(x_i) + \gamma b_N(x_i))$$

$$\sum_{i=1}^{n} L(y_i, a_{N-1}(x_i) + b_i) \to \min_{b_i}$$

$$f(x+s) = f(x) + f'(x) \cdot s + \frac{1}{2} f''(x) \cdot s^2 + o(s^2)$$

$$f(x) = f(x_0) + f'(x_0) \underbrace{(x - x_0)}_{s}$$

$$s = x - x_0$$

$$x = s + x_0$$

$$f(x_0 + s) = f(x_0) + f'(x_0) \cdot s$$

$$\sum_{i=1}^{n}\left[L(y_i, a_{N-1}(x_i)) + \frac{\partial L(y_i, z)}{\partial z}\Big|_{z=a_{N-1}(x_i)} \cdot b_i + \right.$$

$$\left. + \frac{1}{2}\cdot \frac{\partial^2 L(y_i, z)}{\partial z^2}\Big|_{z=a_{N-1}(x_i)} \cdot b_i^2 \right] \to \min_{b_i}$$

$$g_i = \frac{\partial L(y_i, z)}{\partial z}\Big|_{z=a_{N-1}(x_i)}$$

$$h_i = \frac{\partial^2 L(y_i, z)}{\partial z^2}\Big|_{z=a_{N-1}(x_i)}$$

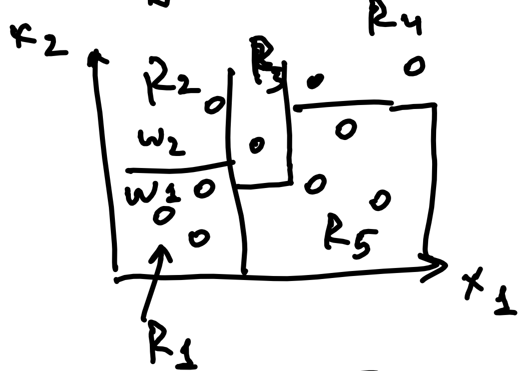
$$\boxed{\sum_{i=1}^{n}\left[g_i b_i + \frac{1}{2} h_i b_i^2 \right]}$$

$$\sum_{i=1}^{n}(b_i - s_i^{(N)})^2 = \sum_i (b_i^2 - 2 s_i^{(N)} b_i + \cancel{s_i^{2(N)}}) =$$

$$= 2\sum_{i=1}^{n}\left(\frac{1}{2}\cdot 1 \cdot b_i^2 + b_i \cdot \underbrace{(-s_i^{(N)})}_{g_i}\right)$$

$$L(y, x+s) \approx L(y, x) + L'_z(y, z)\Big|_{z=x} \cdot s + \ldots$$

$$\mathrm{tree}(x) = \sum_{t=1}^{T} w_t \cdot [x \in R_t]$$



$$Q = \sum_{t=1}^{T} \sum_{i \in R_t} \left[g_i \cdot w_t + \frac{1}{2} h_i \cdot w_t^2 \right] + \frac{\mu}{2} \sum_{t=1}^{T} w_t^2 + \lambda T$$

$$Q = \sum_{t=1}^{T} \left[\Big(\underbrace{\sum_{i \in R_t} g_i}_{G_t} \Big) \cdot w_t + \frac{1}{2} \Big(\underbrace{\sum_{i \in R_t} h_i}_{H_t} + \mu \Big) \cdot w_t^2 \right] + \lambda T$$

$$Q'_{w_t} = G_t + (H_t + \mu) \cdot w_t = 0$$

$$\boxed{w_t^* = - \frac{G_t}{H_t + \mu} = - \frac{\sum_{i \in R_t} g_i}{\sum_{i \in R_t} h_i + \mu}}$$

$$Q^* = \sum_{t=1}^{T} \left(- \frac{G_t^2}{H_t + \mu} + \frac{1}{2} \frac{G_t^2}{H_t + \mu} \right) + \lambda T =$$

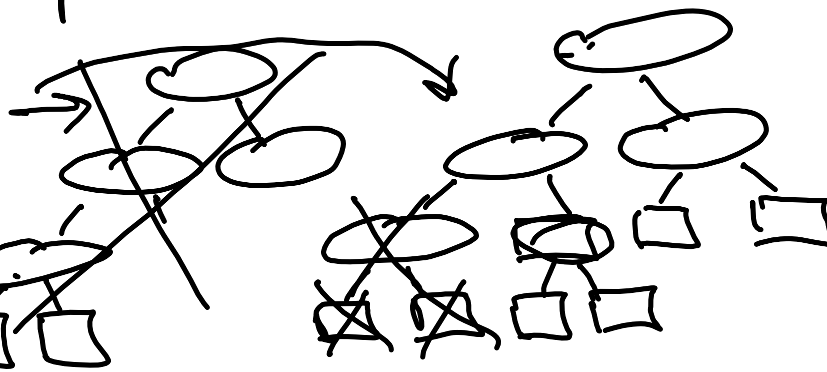
$$\boxed{Q^* = -\frac{1}{2}\sum_{t=1}^{T}\frac{G_t^2}{H_t+\mu} + \lambda T}$$

$$H(R) = -\frac{1}{2}\,\frac{\left(\sum_{i\in R} g_i\right)^2}{\sum_{i\in R} h_i + \mu} + \lambda$$

$$H(R) - H(R_\ell) - H(R_r)$$

$$D = \left[-\frac{1}{2}\cdot\frac{G_R^2}{H_R+\mu} + \frac{1}{2}\cdot\frac{G_{R_\ell}^2}{H_{R_\ell}+\mu} + \frac{1}{2}\,\frac{G_{R_r}^2}{H_{R_r}+\mu}\right] - \lambda \to \max$$

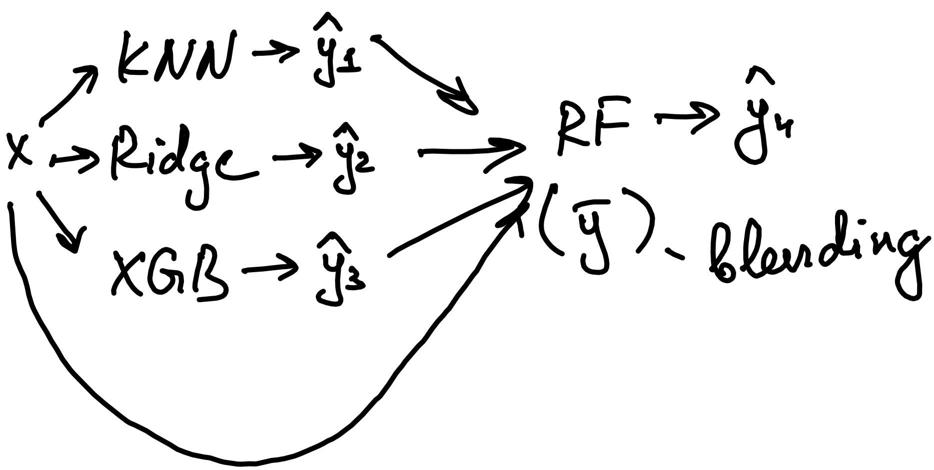
ODT →



$$a_N(x) = \eta \sum_{j=1}^{N} b_j(x) \quad , \eta \in (0, 1]$$

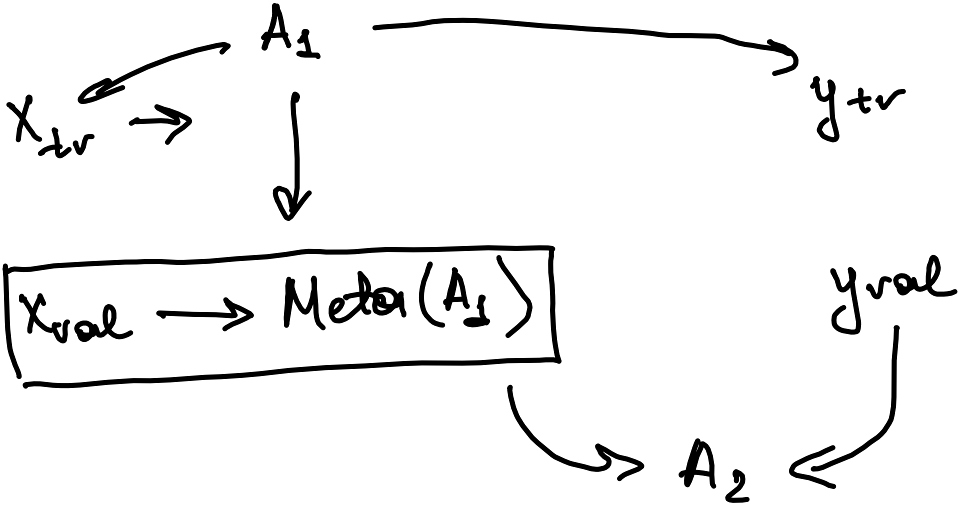
$\boxed{\text{Stacking}}$

$$a(x) = d\big(b_1(x), b_2(x), \dots, b_k(x)\big)$$



$X, y \quad A_1 \quad A_2 \quad X_{test}$

$X_{tr} \quad X_{val} \quad y_{tr} \quad y_{val}$



$$[X_{test} \rightarrow Meta(A_1, X_{test})]$$

$$Meta(A_1) = A_1.predict(X_{val})$$

$$A_1.fit(X_{tr}, y_{tr})$$

$$Meta(A_1, X_{val}) = A.predict(X_{val})$$

$$A_2.fit([X_{val}, Meta(A_1, X_{val})], y_{val})$$

$$A_2.predict([X_{test}, \underbrace{A_1.predict(X_{test})}_{Meta(A_1,\ X_{test})}])$$

Семинар

1. EM $\quad \log P(X|\theta) = \int q(z) \log P(X|\theta)\, dz$

$$\log P(X|\theta) = KL\big(q(z) \,||\, p(z|X,\theta)\big) + \underbrace{\int q(z) \cdot \log \frac{P(X,z|\theta)}{q(z)}\, dz}_{\mathcal{L}(q,\theta) = ELBO.}$$

E-шаг:

$$\mathcal{L} \to \max_q$$

$$q(z) = p(z|X, \theta^{old})$$

M-шаг:

$$\theta^{new} = \arg\max_\theta \mathbb{E}_q \log P(X,z|\theta)$$

② . PCA

$$\begin{cases} a_k^T S a_k \to \max\limits_{a_k} \\ \|a_k\|^2 = 1 \end{cases}$$

③. $$Q = \sum_{i=1}^{n} \log(1 + \exp(-y_i \langle w, x_i \rangle)) + \lambda \sum_{k=1}^{d} w_k^2$$

$$Q = \frac{1}{n} \sum_{i=1}^{n} [y_i \neq a(x_i)] = \frac{1}{n} \sum_{i=1}^{n} [\underbrace{\overbrace{a(x_i)}^{\text{sign}(\langle w, x \rangle)} y_i}_{M_i} < 0]$$

$\log(1 + e^{-M}) \to$ logloss.

④. Hard-Margin.

$$\begin{cases} \frac{1}{2} \|w\|^2 \to \min\limits_{w, b} \\ y_i(\langle w, x_i \rangle + b) \geq 1 \end{cases}$$

Soft-Margin

$$\begin{cases} \frac{1}{2} \|w\|^2 + C \cdot \sum\limits_{i=1}^{n} \xi_i \to \min\limits_{w, b, \xi_i} \\ y_i(\langle w, x_i \rangle + b) \geq 1 - \xi_i \\ \xi_i \geq 0 \end{cases}$$

Hinge-Loss : $\max(0, 1 - M)$

$$\xi_i \geq 1 - y_i(\langle w, x_i\rangle + b)$$

$$\xi_i \geq 0$$

$$\frac{1}{2}\|w\|^2 + C \cdot \sum_{i=1}^{n} \max(0, 1 - y_i \underbrace{(\langle w, x_i\rangle + b)}_{M_i})$$

$$\mathbb{P} = \prod_{t=1}^{T} \prod_{\substack{-d \le j \le d \\ j \neq 0}} \mathbb{P}(w_{t+j} | w_t) \to \max$$

$$-\log \mathbb{P} = -\frac{1}{T} \sum_t \sum_j \log \mathbb{P}(w_{t+j} | w_t) \to \min$$

$$\begin{tikzcd}
 & \text{LinReg} \arrow[r] & \hat{y}_1 \arrow[r] & \text{LinReg} \\
X \arrow[ur] \arrow[r] & \text{RF} \arrow[r] & \hat{y}_2 \arrow[ur] &
\end{tikzcd}$$

$$\begin{tikzcd}
\boxed{X_{tr}} \arrow[r] & A_1 & \boxed{y_{tr}} \arrow[l] & \\
\boxed{X_{val}} \arrow[r] & \boxed{A_1.\text{predict}(X_{val})} \arrow[r] & A_2 & y_{val} \arrow[l]
\end{tikzcd}$$

$\boxed{X_{test}}$ $A_2.\text{predict}(A_1.\text{predict}(X_{test}))$