LSML #6 Linear models

① $R_{egret} = \sum_i^T l_i(w_i) - l_i(w^*)$

② $\hat{R}_{egret} = \sum_i^T \hat{l}_i(w_i) - \hat{l}_i(w^*)$

$\hat{l}_i(w) = l_i(w_i) + \nabla l_i^T(w_i)(w - w_i)$

FTL → FTRL

$w_{t+1} = \arg\min_w \sum_i^T l_i(w) + r(w)$

$w_{t+1} = \arg\min_w \left(\sum_i^t \hat{l}_i(w) + r(w)\right)$

$l_i(w_i) + \nabla l_i^T(w_i)(w - w_i) + \lambda \cdot \|w - w_i\|_2^2 \xrightarrow[w]{} \min$

$\nabla l(w_i) + 2\lambda(w - w_i) = 0$

$2\lambda(w - w_i) = -\nabla l(w_i)$

$w - w_i = \frac{1}{2\lambda} \cdot (-1) \cdot \nabla l(w_i)$

$w_{i+1} = w_i - \frac{1}{2\lambda} \cdot \nabla l(w_i)$

FTRL-proximal

$x \in \mathbb{R}^n$ n - большое

$w_{t+1} = \arg\min_w \left(\sum_i^t l_i(w) + r(w)\right)$

$w_{t+1} = \arg\min_w \left(\sum_i^t \nabla l_i^T(w_i) \cdot w + \sum_s^t \sigma_s \|w - w_s\|_2^2 + \lambda \|w\|_1\right)$

$\nabla l_i(w_i) = g_i \qquad g_{1:t} = \sum_{i=1}^t g_i \qquad \sum_{s=1}^t \sigma_s = \frac{1}{\eta_t}$

$w_{t+1} = \arg\min_w \left(g_{1:t}^T \cdot w + \frac{1}{2} \sum_{s=1}^t \sigma_s \|w - w_s\|_2^2 + \lambda \|w\|_1\right)$

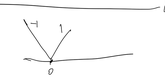
$g_{1:t} \cdot w + \frac{1}{2} \cdot \sum_{s=1}^t \sigma_s \cdot \|w\|_2^2 + \cancel{\frac{1}{2} \sum_{s=1}^t \sigma_s \cdot \|w_s\|_2^2} - \sum_{s=1}^t \sigma_s \cdot w w_s + \lambda \|w\|_1$

$\left(g_{1:t} - \sum_{s=1}^t \sigma_s \cdot w_s\right) \cdot w + \left(\frac{1}{2\eta_t} \cdot \|w\|_2^2\right) + \lambda \cdot \|w\|_1 \to \min$

$\left(g_{1:t} - \sum_{s=1}^t \sigma_s \cdot w_s\right) + \frac{1}{\eta_t} \cdot w + \lambda \frac{\partial}{\partial w} \|w\|_1 = 0$

$$\boxed{Z_t \Rightarrow \boxed{Z_t = Z_{t-1} + g_t + \left(\frac{1}{\eta_t} - \frac{1}{\eta_{t-1}}\right) w_t}}$$

$Z_t + \frac{1}{\eta_t} w + \lambda \cdot \boxed{\frac{\partial}{\partial w} \|w\|_1} = 0$

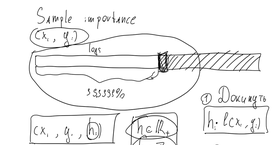


$$\boxed{w_{t+1} = -\eta_t \cdot (Z_t \mp \lambda)}$$

$\|Z_t\| < \lambda \Rightarrow w_{t+1} = 0$

$\|Z_t\| \geq \lambda \Rightarrow w_{t+1} = -\eta_t (Z_t - \lambda \cdot \mathrm{sign}(w))$

$w \qquad Z_t$



(x_i, y_i, h_i) $\quad h \in \mathbb{R}_+$, $h \in \mathbb{Z}_+$

① Домножить вес в лосс $\boxed{h_i \cdot l(x_i, y_i)}$

② Домножить l или градиент $\boxed{w_{i+1} = w_i - \lambda h \nabla_w l}$

s(h)

$\exists\, p = w^T x \quad \nabla_w l = \frac{\partial}{\partial x} l(p, y) \ominus$

$\ominus \frac{\partial}{\partial p} l(x, y) \cdot \frac{\partial p}{\partial w} = \boxed{\frac{\partial l(x,y)}{\partial p}} \cdot x$

Лемма $\exists\, h \in \mathbb{Z}_+$

$w_{t+1} = w_t - s(h_{t+1}) \cdot x_{t+1}$

$\forall\, s(h+1) = s(h) + \lambda \cdot \frac{\partial l}{\partial p}\Big|_{p = (w_t - s(h) x_i)^T x_t}$

$s(0) = 0$

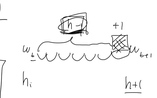
Базу

$w_2 = w_1 - s(1) \cdot x = w_1 - \left(\cancel{s(0)} + \lambda \cdot \frac{\partial l}{\partial p}\Big|_{w_1^T x - \cancel{s(0) x_1}}\right) x$

$w_2 = w_1 - \boxed{\lambda \cdot \frac{\partial l}{\partial p}\Big|_{w_1^T x}} x$

s(h)

Индукция



$\boxed{\forall\, w_t = s(h) \cdot x \;\checkmark}$

$w_{t+1} = w_t - s(h+1) \cdot x = w_t - \left(\underbrace{s(h)}_{\text{const}} + \underbrace{\lambda}_{\text{const}} \underbrace{\boxed{\frac{\partial l}{\partial p}\Big|_{(w_t - s(h) \cdot x_t)^T x_t}}}_{s(h+1)}\right) x \ominus$

$\ominus\, w_t = (s(h) + \lambda \cdot x_t) x$

$s(h+1) \sim s(h)$

$\boxed{s(h) = \frac{w^T x - y}{x^T x} \cdot \left(1 - (1 - \lambda x^T x)^h\right)}$ $\quad \boxed{h \in \mathbb{Z}_+}$ $\quad h \in \mathbb{R}_+$

$3 ; \frac{1}{2} ; \frac{1}{3} \quad | \quad 30 ; 5 ; 2$

$\nabla_w l(x, y)$

s(h) λ

$\boxed{s(h) \cdot \frac{\lambda}{h}}$

$\boxed{h} \to \boxed{p} \in \mathbb{R}_+$

$\frac{w^T x - y}{x^T x} \cdot \left(1 - \left(1 - \frac{\lambda}{h} x^T x\right)^h\right)$

$h \to \infty$

$$\boxed{\frac{w^T x - y}{x^T x} \cdot \left(1 - \exp(-\lambda h x^T x)\right)}$$

IDT

A_1	A_1	A_1	A_2	A_2
y_1	y_2	$\hat{y}_1$	$\hat{y}_2$	$\hat{y}_2$

$\sum_N (y_i - c)^2 - \sum_{N_1} (y_i - c_1)^2 - \sum_{N_2} (y_i - c_2)^2 \to \max$

$\sum_N (y_i - c)^2 = \sum_N y_i^2 - 2 \boxed{\sum_N y_i} c + \sum_{\boxed{N}} \boxed{c^2} = \sum_N y_i^2 - 2 \cdot c^2 N + c^2 N = \boxed{\sum_N y_i^2 - c^2 N}$

$c^* = \sum_N \frac{y_i}{N} \Rightarrow \boxed{\sum_N y_i = c^* \cdot N} = Y \qquad \sum_N y_i^2 - c^2 N - \sum_{N_1} y_i^2 + c_1^2 N_1 - \sum_{N_2} y_i^2 + c_2^2 N_2 \to \max$

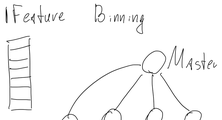
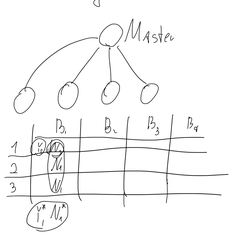
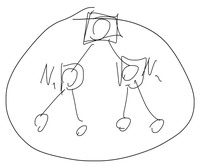
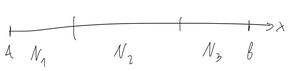
$\boxed{\cancel{\sum_N y_i^2} + c_1^2 N_1 + c_2^2 N_2} \to \max$

$\exists\, \cancel{\sum_N y_i} \quad Y_1 = \sum_{N_1} y_i \quad Y_2 = \sum_{N_2} y_i$

$Y_1^2 \qquad Y_2^2$

$\frac{c_1^2 N_1 \cdot N_1}{N_1} + \frac{c_2^2 N_2 \cdot N_2}{N_2} \Rightarrow \frac{Y_1^2}{N_1} + \frac{Y_2^2}{N_2} \to \max$

$\boxed{-\frac{Y_1^2}{N_1} - \frac{Y_2^2}{N_2} \to \min}$

$\sum I$

Feature Binning