LSML #5 Recap

· Lin Reg $x = [f_1, \dots f_d]$ $d > 10^6$

| Count-min Sketch

· Хотим оценивать частоты

X $\|X\| = n$

$\vec{a}(t) = (a_1(t), \dots, a_n(t))$

$\forall a_i(0) = 0$

· На шаге t: i_t

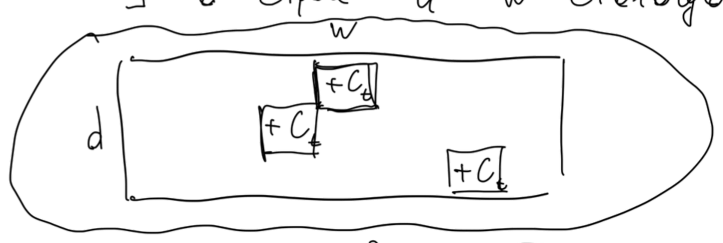
$$\begin{cases} a_{i_t}(t) = a_{i_t}(t-1) + c_t \\ a_{i'}(t) = a_{i'}(t-1) \quad \forall i' \neq i_t \end{cases}$$

$\exists$ d хешей по w значений

$h_i(x)$ $i \in [1, \dots, d]$ – хеши независимы

$i_t \to h_1^{i_t} \sim h_d^{i_t}$

$\exists$ d-строк и w столбцов



ошибка была $\pm\varepsilon$ с вероятностью δ

$w^* = \lceil \frac{e}{\varepsilon} \rceil$ $\quad d^* = \lceil \ln \frac{1}{\delta} \rceil$

| Теорема $a_i \leq \hat{a}_i$ $\quad \min_j count[j, h_j(i)] = \hat{a}_i$

$$P(\hat{a}_i > a_i + \varepsilon \cdot \|\vec{a}_n\|_1) \leq \delta$$

$$P\left(\frac{\hat{a}_i}{\|\vec{a}\|_1} > \frac{a_i}{\|\vec{a}\|_1} + \varepsilon\right) \leq \delta$$

$$\exists \; I_{ijk} = \begin{cases} 1 & \text{если } (i \neq k) \wedge (h_j(i) = h_j(k)) \\ 0 & \text{иначе} \end{cases}$$

$$\mathbb{E}(I_{ijk}) = P(h_j(i) = h_j(k)) = \frac{1}{w} \approx \frac{\varepsilon}{e}$$

$$\exists X_{ij} = \sum_{k=1}^{n} I_{ijk} \Rightarrow count[j, h_j(i)] = a_i + X_{ij}$$

$$\Downarrow$$

$$\min_j count \geq a_i$$

$$\mathbb{E}(X_{ij}) = \mathbb{E}\left(\sum_{k=1}^{n} I_{ijk} \cdot a_k\right) = \sum_{k=1}^{n} a_k \cdot \mathbb{E}(I_{ijk}) \leq \frac{\varepsilon}{e} \cdot \|a\|_1$$

$$P(\hat{a}_i > a_i + \varepsilon \cdot \|\vec{a}\|_1) = P(\forall_j \; count[j; h_j(i)] > a_i + \varepsilon \cdot \|\vec{a}\|_1) =$$

$$= P(\forall_j \; \cancel{a_i} + X_{ij} > \cancel{a_i} + \varepsilon \cdot \|\vec{a}\|_1) \leq P(\forall_j \; X_{ij} > e \cdot \mathbb{E}(X_{ij})) < e^{-d} \leq \delta$$

$$\boxed{P(X \geq t) \leq \frac{\mathbb{E}(X)}{t} \quad \forall t > 0}$$

$$d = \lceil \ln \frac{1}{\delta} \rceil$$

$-d \leq \ln \delta \qquad d \geq -\ln \delta$

$d \geq \ln \delta^{-1}$

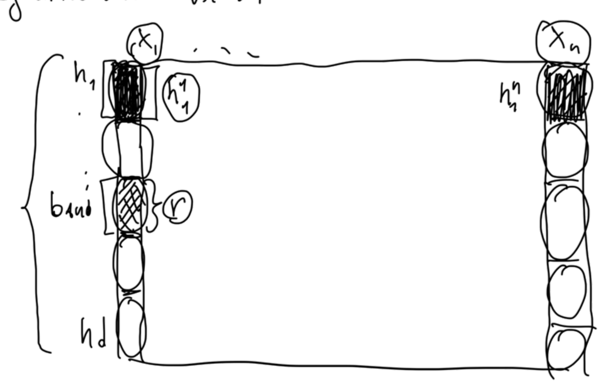
$d \geq \ln\left(\frac{1}{\delta}\right)$

$d^* = \lceil \ln\left(\frac{1}{\delta}\right) \rceil$

| LSH locally sensetive hashing

Идея: сделать хеш(x): в одном значение хеша попадали одинаковые эл-ты

	x_1		x_2
h_1	x	$\neq$	y
h_2	x	$=$	y
h_3	x	$\neq$	y



$b \cdot r = s$

$h_i = [1 \dots k]$

$\left(\frac{1}{k}\right)^r$

$1 - \left(\frac{1}{k}\right)^r$

$\left(1 - \left(\frac{1}{k}\right)^r\right)^b$

$1 - \left(1 - \left(\frac{1}{k}\right)^r\right)^b$

$b \approx 20/50$

$r \approx 5 - 10$

0.3588

1) map $h_1 \dots h_b$

| Online learning

X $\|X\| >> HDD$ Датасет кликов

$loss \xrightarrow[\theta]{} \min$ $\equiv$ offline learning

$$\text{Regret} = \sum_{i=1}^{T} l_i(w_i) - l_i(w^*)$$

$x_1 \to \hat{y}_1 \to y_1 \to w_2 \qquad x_2 \to \hat{y}_2 \to y_2 \to w_3$

| FTL = follow the leader

$$w_{t+1} = \arg\min_w \sum_{i=1}^{t} l_i(w)$$

– дорого и плохо

– веса улетают

$$w_{t+1} = \arg\min_w \sum_{i=1}^{t} l_i(w) + r_i(w)$$

$$r_i(w) = \lambda \cdot \|w - w_i\|_2^2$$

| Strong convexity ! $\quad f(y) \geq f(x) + \nabla f^T(x)(y - x) + m \cdot \|y - x\|_2^2$

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

① $\widehat{\text{Regret}} = \sum_{i=1}^{T} \cancel{\hat{l}_i(w_i)} - \hat{l}_i(w^*)$

② $\text{Regret} = \sum_{i=1}^{T} \cancel{l_i(w_i)} - l_i(w^*)$

$\boxed{1} \geq \boxed{2}$ (min, m_i)

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

$$\boxed{l_i(w^*)} \geq l_i(w_i) + \nabla l_i^T(w_i)(w^* - w_i) + m \cdot \|w^* - w_i\|_2^2$$

$m > 0$