بسم الله الرحمن الرحیم

جلسه ششم

خلاصه‌سازی برای مه‌داده

کران پایین

کران پایین برای تعداد اعداد F0 متفاوت

الگوریتم‌های ارائه شده (مثلا ++FM):

تصادفی

تقریبی

الگوریتم‌های ارائه شده (مثلا ++FM):

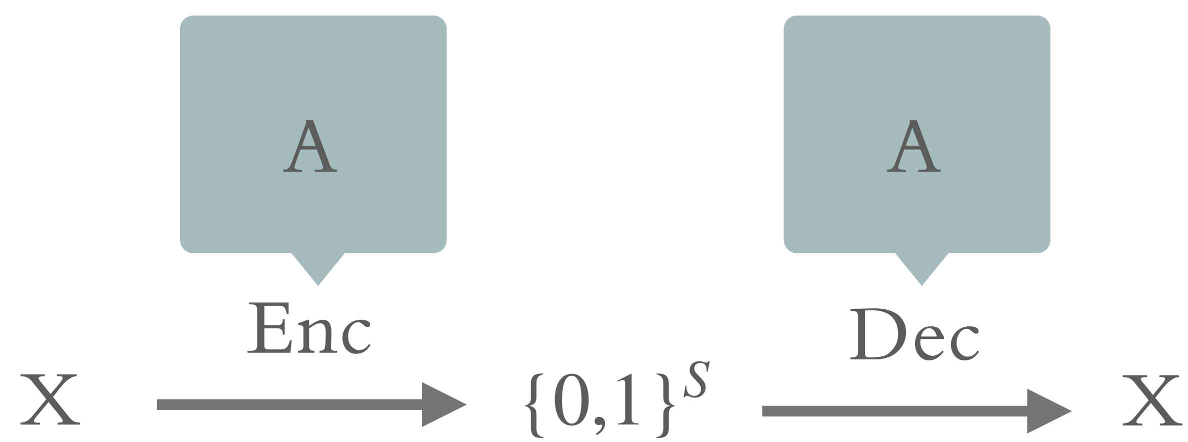
تصادفی

تقریبی

آیا هر دو لازمند؟

تکنیک برای تولید کران پایین

هدف: نمی‌تواند حافظه A کم باشد



پس: $S \geq \log |X|$

کران پایین برای حافظه الگوریتم دقیق قطعی

Theorem 3.1.1. *Suppose $\mathcal{A}$ is a deterministic streaming algorithm that computes the number of distinct elements exactly. Then $\mathcal{A}$ uses at least n bits of memory.*

کران پایین برای حافظه الگوریتم دقیق قطعی

Theorem 3.1.1. *Suppose $\mathcal{A}$ is a deterministic streaming algorithm that computes the number of distinct elements exactly. Then $\mathcal{A}$ uses at least n bits of memory.*

Enc(x): $x \in \{0,1\}^n$

۱ـ یک رشته با اعدادی که اندیسشان در ورودی هست (مرتب)

۲ـ رشته را به الگوریتم A بدهیم.

۳ـ خروجی (M): حافظه A

کران پایین برای حافظه الگوریتم دقیق قطعی

Theorem 3.1.1. *Suppose $\mathcal{A}$ is a deterministic streaming algorithm that computes the number of distinct elements exactly. Then $\mathcal{A}$ uses at least n bits of memory.*

Enc(x): $x \in \{0,1\}^n$

۱ـ یک رشته با اعدادی که اندیسشان در ورودی هست (مرتب)

۲ـ رشته را به الگوریتم A بدهیم.

۳ـ خروجی (M): حافظه A

Dec(M)

```
s ← A.query()  // support size of x, i.e. |{i : x_i ≠ 0}|
x ← (0, 0, ..., 0)
for i = 1...n:
    A.update(i) // append i to the stream
    r ← A.query() // will either be s or s+1
    if r = s:  // Encoder must have included i, so it wasn't a new distinct element
        x_i ← 1
    s ← r
return x
```

الگوریتم قطعی تقریبی ـ کران پایین

Theorem 3.1.6. *Suppose $\mathcal{A}$ is a deterministic streaming algorithm that always outputs a value $\tilde{t}$ when queried such that $t \leq \tilde{t} \leq 1.9t$, where t is the number of distinct elements. Then $\mathcal{A}$ uses at least cn bits of memory for some constant $c > 0$.*

کدهای اصلاح خطا

اندازه بلوک

کد اصلاح خطا: $C \subset [q]^{l}$

الفبا

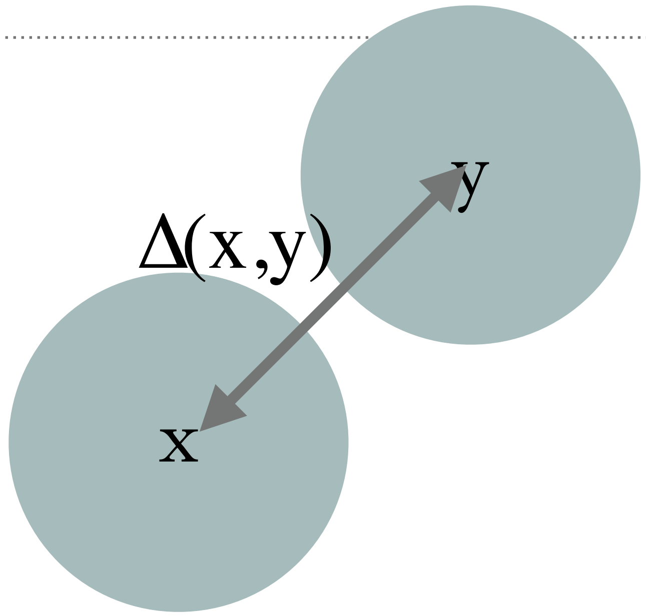
Hamming distance: $\Delta(x,y) := |\{i : x_i \neq y_i\}|$

relative Hamming distance : $\delta(x, y) = \Delta(x, y)/l$.

distance of the code : $\min_{c,c' \in C} \Delta(c, c')$

relative distance : $\min_{c,c' \in C} \delta(c, c')$

کدهای اصلاح خطا



اندازه بلوک

کد اصلاح خطا: $C \subset [q]^{l}$

الفبا

Hamming distance: $\Delta(x,y) := |\{i : x_i \neq y_i\}|$

relative Hamming distance : $\delta(x, y) = \Delta(x, y)/l$.

distance of the code : $\min_{c,c' \in C} \Delta(c, c')$

relative distance : $\min_{c,c' \in C} \delta(c, c')$

چرا مهم است؟

Theorem 3.1.4. *For any integers $q, n > 1$, there exists a code $\mathcal{C}$ with $|\mathcal{C}| = n$ and block length $\ell = O(q \log n)$ with relative distance $1 - 6/q$.*

Theorem 3.1.4. *For any integers $q, n > 1$, there exists a code $\mathcal{C}$ with $|\mathcal{C}| = n$ and block length $\ell = O(q \log n)$ with relative distance $1 - 6/q$.*

C: شامل n تا ci هر کدام با توزیع یکنواخت از $[q]^l$

Theorem 3.1.4. *For any integers $q, n > 1$, there exists a code $\mathcal{C}$ with $|\mathcal{C}| = n$ and block length $\ell = O(q \log n)$ with relative distance $1 - 6/q$.*

C: شامل n تا ci هر کدام با توزیع یکنواخت از $[q]^l$

$Y_{i,j}$: تعداد شباهت ci و cj

$E[Y_{i,j}] = 1/q$

Theorem 3.1.4. *For any integers $q, n > 1$, there exists a code $\mathcal{C}$ with $|\mathcal{C}| = n$ and block length $\ell = O(q \log n)$ with relative distance $1 - 6/q$.*

C: شامل n تا ci هرکدام با توزیع یکنواخت از $[q]^l$

$Y_{i,j}$: تعداد شباهت ci و cj

$E[Y_{i,j}] = l/q$

$$P[Y_{i,j} > 6l/q] < 2exp(-25/3 \cdot l/q)$$

Theorem 3.1.4. *For any integers $q, n > 1$, there exists a code $\mathcal{C}$ with $|\mathcal{C}| = n$ and block length $\ell = O(q \log n)$ with relative distance $1 - 6/q$.*

C: شامل n تا ci هرکدام با توزیع یکنواخت از $[q]^l$

$Y_{i,j}$: تعداد شباهت ci و cj

$E[Y_{i,j}] = 1/q$

$$P[Y_{i,j} > 6l/q] < 2exp(-25/3 \cdot l/q) < \frac{1}{n^2}$$

با انتخاب l بزرگ

Theorem 3.1.4. *For any integers $q, n > 1$, there exists a code $\mathcal{C}$ with $|\mathcal{C}| = n$ and block length $\ell = O(q \log n)$ with relative distance $1 - 6/q$.*

C: شامل n تا ci هر کدام با توزیع یکنواخت از $[q]^l$

$Y_{i,j}$: تعداد شباهت ci و cj

$E[Y_{i,j}]$ = 1/q

$$P[Y_{i,j} > 6l/q] < 2exp(-25/3 \cdot l/q) < \frac{1}{n^2}$$

با انتخاب l بزرگ

احتمال بد بودن < ۱

Corollary 3.1.5. *For any integer $n > 0$ and any integers $\ell, q > 1$ such that $n = q\ell$, there exists a subset $\mathcal{B}_{q,\ell}$ of $\{0,1\}^n$ satisfying the following properties:*

1. *Every $c \in \mathcal{B}_{q,\ell}$ has support size ℓ, i.e. $|\{i : c_i \neq 0\}| = \ell$.*
2. *For $c \neq c' \in \mathcal{B}_{q,\ell}$, $|\{i : c_i = c'_i\}| \leq 6\ell/q$.*
3. *$|\mathcal{B}_{q,\ell}| = \exp(\Omega(\ell/q))$.*

ci=1 و

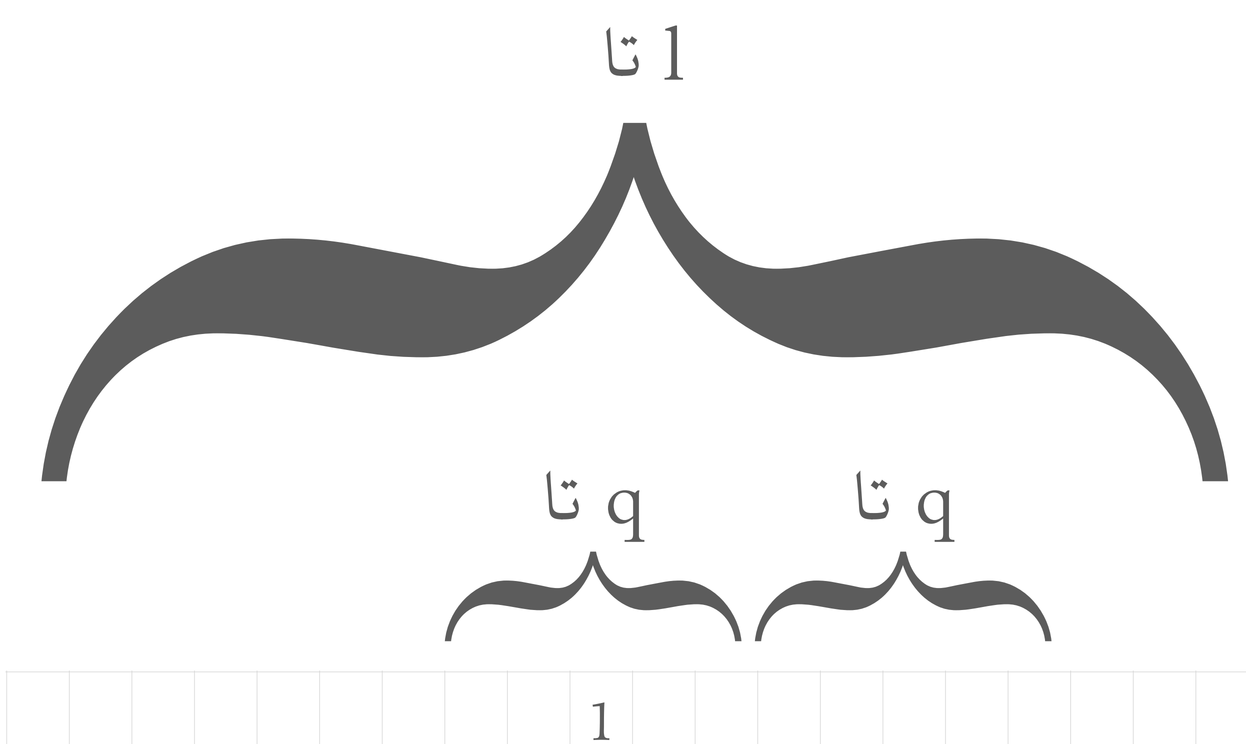
$$P[Y_{i,j} > 6l/q] < 2exp(-25/3 \cdot l/q)$$

Corollary 3.1.5. *For any integer $n > 0$ and any integers $\ell, q > 1$ such that $n = q\ell$, there exists a subset $\mathcal{B}_{q,\ell}$ of $\{0,1\}^n$ satisfying the following properties:*

1. *Every $c \in \mathcal{B}_{q,\ell}$ has support size ℓ, i.e. $|\{i : c_i \neq 0\}| = \ell$.*
2. *For $c \neq c' \in \mathcal{B}_{q,\ell}$, $|\{i : c_i = c'_i\}| \leq 6\ell/q$.*
3. *$|\mathcal{B}_{q,\ell}| = \exp(\Omega(\ell/q))$.*

و ci=1

$$P[Y_{i,j} > 6l/q] < 2exp(-25/3 \cdot l/q)$$



Theorem 3.1.6. *Suppose $\mathcal{A}$ is a deterministic streaming algorithm that always outputs a value $\tilde{t}$ when queried such that $t \le \tilde{t} \le 1.9t$, where t is the number of distinct elements. Then $\mathcal{A}$ uses at least cn bits of memory for some constant $c > 0$.*

Theorem 3.1.6. *Suppose $\mathcal{A}$ is a deterministic streaming algorithm that always outputs a value $\tilde{t}$ when queried such that $t \le \tilde{t} \le 1.9t$, where t is the number of distinct elements. Then $\mathcal{A}$ uses at least cn bits of memory for some constant $c > 0$.*

$\mathrm{Enc} : B_{q,l} \to \{0, 1\}^S$

$q = 100$ and $l = n/q$

$\longrightarrow$

$S \geq \log |B_{q,l}| = \Omega(n)$

Theorem 3.1.6. *Suppose $\mathcal{A}$ is a deterministic streaming algorithm that always outputs a value $\tilde{t}$ when queried such that $t \leq \tilde{t} \leq 1.9t$, where t is the number of distinct elements. Then $\mathcal{A}$ uses at least cn bits of memory for some constant $c > 0$.*

$\mathrm{Enc} : B_{q,l} \rightarrow \{0, 1\}^S$

$q = 100$ and $l = n/q$

$\longrightarrow$ $S \geq \log |B_{q,l}| = \Omega(n)$

$\mathrm{Enc} : x \rightarrow \{i \mid x[i]=1\} \rightarrow A \rightarrow$ حافظه A

Theorem 3.1.6. *Suppose $\mathcal{A}$ is a deterministic streaming algorithm that always outputs a value $\tilde{t}$ when queried such that $t \leq \tilde{t} \leq 1.9t$, where t is the number of distinct elements. Then $\mathcal{A}$ uses at least cn bits of memory for some constant $c > 0$.*

$\text{Enc} : B_{q,l} \to \{0, 1\}^S$

$q = 100$ and $l = n/q$

$\longrightarrow$ $S \geq \log |B_{q,l}| = \Omega(n)$

$\text{Enc} : x \to \{i \mid x[i]=1\} \to A \to$ حافظه A

Dec:

for $c \in \mathcal{B}_{q,t}$:
    $\mathcal{A}$.init(M) `// initialize` $\mathcal{A}$`'s memory to` M
    for $i = 1, 2, \ldots, n$:
        if $c_i = 1$:
            $\mathcal{A}$.update(i)
    if $\mathcal{A}$.query() $\leq 1.9\ell$
        return c

Theorem 3.1.6. *Suppose $\mathcal{A}$ is a deterministic streaming algorithm that always outputs a value $\tilde{t}$ when queried such that $t \leq \tilde{t} \leq 1.9t$, where t is the number of distinct elements. Then $\mathcal{A}$ uses at least cn bits of memory for some constant $c > 0$.*

$\mathrm{Enc} : B_{q,l} \rightarrow \{0, 1\}^S$

$q = 100$ and $l = n/q$ $\longrightarrow$ $S \geq \log |B_{q,l}| = \Omega(n)$

$\mathrm{Enc} : x \rightarrow \{i \mid x[i]=1\} \rightarrow A \rightarrow$ حافظه A

Dec:

```
for c ∈ B_{q,t}:
    A.init(M) // initialize A's memory to M
    for i = 1, 2, ..., n:
        if c_i = 1:
            A.update(i)
    if A.query() ≤ 1.9ℓ
        return c
```

اگر c=x:

وگرنه:

$|c \cup x| = |c| + |x| - |c \cap x| > 2l - 6l/q = 1.94 \times l$

الگوریتم تصادفی دقیق ـ کران پایین

Theorem 3.1.7. *Suppose $\mathcal{A}$ is a randomized streaming algorithm that outputs the exact number of distinct elements with success probability at least $2/3$ for the last query in any fixed sequence of stream updates and queries. Then $\mathcal{A}$ uses at least cn bits of memory for some constant $c > 0$.*

کم کردن احتمال خطا

احتمال خطا: 10^{-6}

حافظه: $O(S)$

$A \rightarrow A'$

احتمال درستی: ۲/۳

حافظه: S

۱

۲

$\in B_{q,l}$

همان قبلی

$x \longrightarrow \text{Enc} \longrightarrow \text{Dec} \longrightarrow x'$

منبع تصادفی

الگوریتم تصادفی A: A(ورودی , r)

۲

$\in B_{q,l}$

همان قبلی

$x \longrightarrow \mathrm{Enc} \longrightarrow \mathrm{Dec} \longrightarrow x'$

Yx: تعداد بیت‌های متفاوت x و x’

$E\, Y_X \le n/10^6$

منبع تصادفی

الگوریتم تصادفی A: $A($ورودی$, r)$

۲

$\in B_{q,l}$

همان قبلی

$x \longrightarrow Enc \longrightarrow Dec \longrightarrow x'$

Yx: تعداد بیت‌های متفاوت x و x’

$E\, Y_X \le n/10^6$

حالت بد: $Y_x > ۲l/q = ۲n/q^2$

Zx: $Y_x > 2n/10^6$

منبع تصادفی

الگوریتم تصادفی A: A(ورودی , r)

۲

$\in B_{q,l}$

همان قبلی

$x \longrightarrow \mathrm{Enc} \longrightarrow \mathrm{Dec} \longrightarrow x'$

Yx: تعداد بیت‌های متفاوت x و x'

$E\,Y_x \le n/10^6$

حالت بد: $Y_x > ۲l/q = ۲n/q^2$

Zx: $Y_x > 2n/10^6$

$P(Z_x = 1) < 1/2$ $\xleftarrow{\text{پس}}$

منبع تصادفی

الگوریتم تصادفی A: A(ورودی , r)

۲

$\in B_{q,l}$

همان قبلی

x → Enc → Dec → x’

Yx: تعداد بیت‌های متفاوت x و x’

E Yx ≤ n/10^6

حالت بد: Yx > ۲l/q = ۲n/q^2

Zx: $Y_x > 2n/10^6$

پس ←

P(Zx = 1) < 1/2

پس ←

$$E\sum_{x\in B_{q,l}} Z_x < \frac{1}{2}|B_{q,l}|$$

منبع تصادفی

A: الگوریتم تصادفی A(ورودی , r)

۲

$\in B_{q,l}$

همان قبلی

$x \longrightarrow \mathrm{Enc} \longrightarrow \mathrm{Dec} \longrightarrow x'$

Yx: تعداد بیت‌های متفاوت x و x'

$E\,Y_x \le n/10^6$

حالت بد: $Y_x > ۲l/q = ۲n/q^2$

Zx: $Y_x > 2n/10^6$

پس ←

$P(Z_x = 1) < 1/2$

پس ←

$$E\sum_{x\in B_{q,l}} Z_x < \frac{1}{2}|B_{q,l}|$$

پس ← حداقل نیمی از Bq,l درست پاسخ می‌دهند.

منبع تصادفی

A: الگوریتم تصادفی A(ورودی , r)

۲

$\in B_{q,l}$

همان قبلی

$x \longrightarrow \mathrm{Enc} \longrightarrow \mathrm{Dec} \longrightarrow x'$

Yx: تعداد بیت‌های متفاوت x و x’

$E\, Y_X \le n/10^6$

حالت بد: $Yx > ۲l/q = ۲n/q^2$

$Y_x > 2n/10^6$:Zx

پس ⟵ $P(Z_X = 1) < 1/2$

پس ⟵ $E \sum_{x \in B_{q,l}} Z_x < \frac{1}{2} |B_{q,l}|$

پس ⟵ حداقل نیمی از Bq,l درست پاسخ می‌دهند.

منبع تصادفی

قطعی‌سازی ⟵ الگوریتم تصادفی A: A(ورودی , r)