بسم الله الرحمن الرحیم

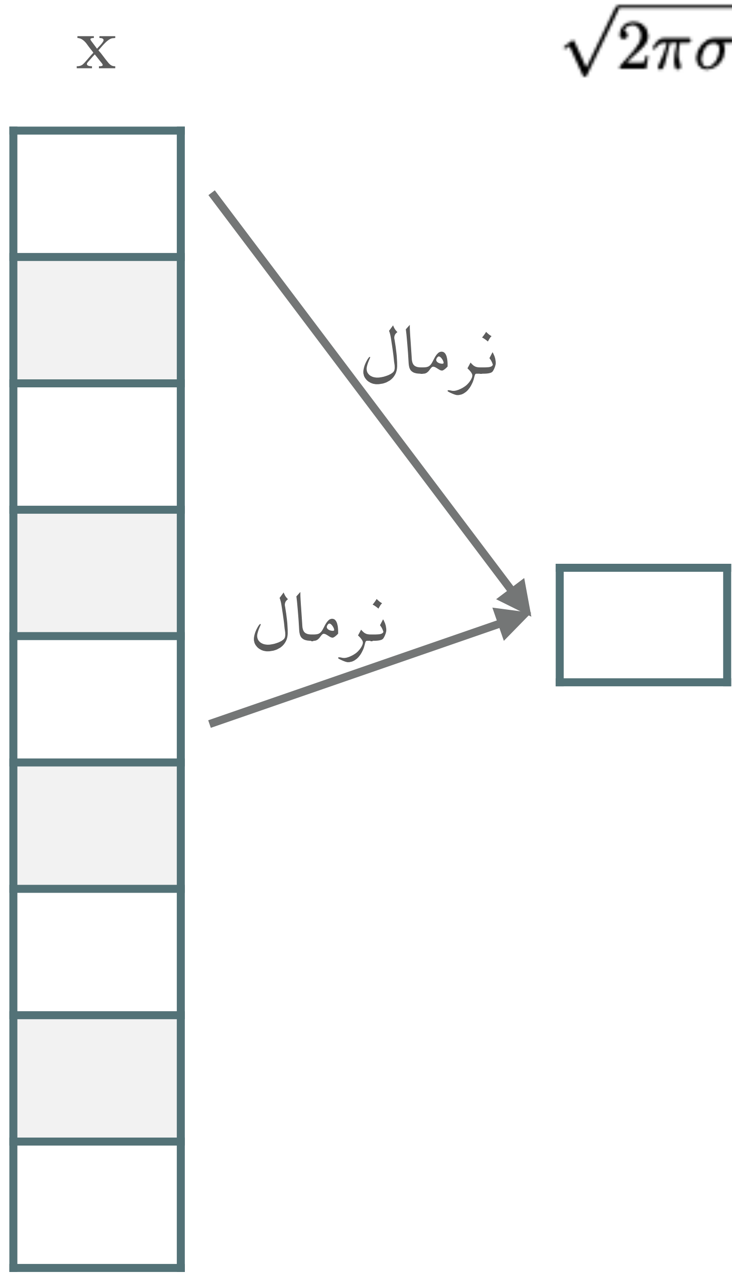
جلسه دهم

خلاصه‌سازی برای مه‌داده

خلاصه‌سازی JL

$$\frac{1}{\sqrt{2\pi\sigma^2}}\exp(-(x-\mu)^2/2\sigma^2)$$

- ماتریس R: n در k
  - Rij: نرمال (N (0, 1
- الگوریتم خلاصه‌سازی:
  - جواب: $\|Rx\|_2^2/k$
-



تحلیل الگوریتم JL:

- **امید:** $\mathbb{E}(\|Rx\|_2^2/k) = \|x\|_2^2$
- اثبات:

$$\mathbb{E}(\|Rx\|_2^2/k) = \frac{1}{k}\mathbb{E}(x^T R^T Rx) = \frac{1}{k}x^T\,\mathbb{E}(R^T R)x = \|x\|_2^2$$

- **واریانس:**
- حکم: $\mathbb{P}(|\|Rx\|_2^2 - k\|x\|_2^2| \geq \varepsilon k\|x\|_2^2) \leq \exp(-C\varepsilon^2 k)$
- حکم جدید: $\mathbb{P}(|\|Z\|_2^2 - k| \geq \varepsilon k) \leq \exp(-C\varepsilon^2 k)$

Rx = Z

حذف ||x||

$$\mathbb{P}(|\|Z\|_2^2 - k| \geq \varepsilon k) \leq \exp(-C\varepsilon^2 k)$$

$$\mathbb{P}(\|Z\|_2^2 \geq (1+\varepsilon)k) \leq \exp(-\varepsilon^2 \frac{k}{6} + O(k\varepsilon^3))$$

$1 + \epsilon \geq (1 + \epsilon')^2$ $\epsilon' := \epsilon/3$

$$P[Y > k(1+\epsilon)] \leq P[Y > k(1+\epsilon')^2]$$

$\leq$

$$\mathbb{P}(|\|Z\|_2^2 - k| \geq \varepsilon k) \leq \exp(-C\varepsilon^2 k)$$

$$\mathbb{P}(\|Z\|_2^2 \geq (1+\varepsilon)k) \leq \exp(-\varepsilon^2 \frac{k}{6} + O(k\varepsilon^3))$$

$Y = \|Z\|_2^2$

$1 + \epsilon \geq (1 + \epsilon')^2$

$\epsilon' := \epsilon/3$

$$P[Y > k(1+\epsilon)] \leq P[Y > k(1+\epsilon')^2]$$

$\leq$

$$\mathbb{P}(|\|Z\|_2^2 - k| \geq \varepsilon k) \leq \exp(-C\varepsilon^2 k)$$

$$\mathbb{P}(\|Z\|_2^2 \geq (1+\varepsilon)k) \leq \exp(-\varepsilon^2 \frac{k}{6} + O(k\varepsilon^3))$$

$Y = \|Z\|_2^2$

$1 + \epsilon \geq (1 + \epsilon')^2$

$\epsilon' := \epsilon/3$

$$P[Y > k(1+\epsilon)] \leq P[Y > k(1+\epsilon')^2]$$

$$\mathbb{P}(Y > \alpha) = \mathbb{P}(\exp(sY) > \exp(s\alpha))$$

$\leq$

$$\mathbb{P}(|\|Z\|_2^2 - k| \geq \varepsilon k) \leq \exp(-C\varepsilon^2 k)$$

$$\mathbb{P}(\|Z\|_2^2 \geq (1+\varepsilon)k) \leq \exp(-\varepsilon^2 \frac{k}{6} + O(k\varepsilon^3))$$

$Y = \|Z\|_2^2$ $1 + \epsilon \geq (1+\epsilon')^2$ $\epsilon' := \epsilon/3$

$$P[Y > k(1+\epsilon)] \leq P[Y > k(1+\epsilon')^2]$$

$$\mathbb{P}(Y > \alpha) = \mathbb{P}(\exp(sY) > \exp(s\alpha))$$

$\alpha := k(1+\epsilon')^2$

$\leq$

$$\mathbb{P}(|\|Z\|_2^2 - k| \geq \varepsilon k) \leq \exp(-C\varepsilon^2 k)$$

$$\mathbb{P}(\|Z\|_2^2 \geq (1+\varepsilon)k) \leq \exp(-\varepsilon^2 \frac{k}{6} + O(k\varepsilon^3))$$

$Y = \|Z\|_2^2$ $1 + \epsilon \geq (1+\epsilon')^2$ $\epsilon' := \epsilon/3$

$$P[Y > k(1+\epsilon)] \leq P[Y > k(1+\epsilon')^2]$$

$$\mathbb{P}(Y > \alpha) = \mathbb{P}(\exp(sY) > \exp(s\alpha)) \leq \exp(-s\alpha)\,\mathbb{E}(\exp(sY))$$

$\alpha := k(1+\epsilon')^2$

$\leq$

$$\mathbb{P}(|\|Z\|_2^2 - k| \geq \varepsilon k) \leq \exp(-C\varepsilon^2 k)$$

$$\mathbb{P}(\|Z\|_2^2 \geq (1+\varepsilon)k) \leq \exp(-\varepsilon^2 \frac{k}{6} + O(k\varepsilon^3))$$

$Y = \|Z\|_2^2$

$1 + \epsilon \geq (1+\epsilon')^2$

$\epsilon' := \epsilon/3$

$$P[Y > k(1+\epsilon)] \leq P[Y > k(1+\epsilon')^2]$$

$$\mathbb{P}(Y > \alpha) = \mathbb{P}(\exp(sY) > \exp(s\alpha)) \leq \exp(-s\alpha)\,\mathbb{E}(\exp(sY))$$

$\alpha := k(1+\epsilon')^2$

ماركوف

$\leq$

$$\mathbb{P}(|\|Z\|_2^2 - k| \geq \varepsilon k) \leq \exp(-C\varepsilon^2 k)$$

$$\mathbb{P}(\|Z\|_2^2 \geq (1+\varepsilon)k) \leq \exp(-\varepsilon^2 \frac{k}{6} + O(k\varepsilon^3))$$

$Y = \|Z\|_2^2$

$1 + \epsilon \geq (1+\epsilon')^2$

$\epsilon' := \epsilon/3$

$$P[Y > k(1+\epsilon)] \leq P[Y > k(1+\epsilon')^2]$$

$$\mathbb{P}(Y > \alpha) = \mathbb{P}(\exp(sY) > \exp(s\alpha)) \leq \exp(-s\alpha)\,\mathbb{E}(\exp(sY))$$

$\alpha := k(1+\epsilon')^2$

ماركوف

$Y = \sum_i Z_i^2$

$\leq$

$$\mathbb{P}(|\|Z\|_2^2 - k| \geq \varepsilon k) \leq \exp(-C\varepsilon^2 k)$$

$$\mathbb{P}(\|Z\|_2^2 \geq (1+\varepsilon)k) \leq \exp(-\varepsilon^2 \frac{k}{6} + O(k\varepsilon^3))$$

$Y = \|Z\|_2^2$

$1 + \epsilon \geq (1+\epsilon')^2$

$\epsilon' := \epsilon/3$

$$P[Y > k(1+\epsilon)] \leq P[Y > k(1+\epsilon')^2]$$

$$\mathbb{P}(Y > \alpha) = \mathbb{P}(\exp(sY) > \exp(s\alpha)) \leq \exp(-s\alpha)\,\mathbb{E}(\exp(sY))$$

$\alpha := k(1+\epsilon')^2$

ماركوف

$= \prod_i E(\exp(sZ_i^2))$

$Y = \sum_i Z_i^2$

$\leq$

$$\mathbb{P}(|\|Z\|_2^2 - k| \geq \varepsilon k) \leq \exp(-C\varepsilon^2 k)$$

$$\mathbb{P}(\|Z\|_2^2 \geq (1+\varepsilon)k) \leq \exp(-\varepsilon^2 \frac{k}{6} + O(k\varepsilon^3))$$

$Y = \|Z\|_2^2$

$1 + \epsilon \geq (1+\epsilon')^2$

$\epsilon' := \epsilon/3$

$$P[Y > k(1+\epsilon)] \leq P[Y > k(1+\epsilon')^2]$$

$$\mathbb{P}(Y > \alpha) = \mathbb{P}(\exp(sY) > \exp(s\alpha)) \leq \exp(-s\alpha)\,\mathbb{E}(\exp(sY))$$

$\alpha := k(1+\epsilon')^2$

مارکوف

$= \prod_i E(\exp(sZ_i^2))$

$Y = \sum_i Z_i^2$

Zi: نرمال

$$\mathbb{E}(Z_i) = \sum \mathbb{E}(r_{ij}x_j) = 0$$

$$\mathrm{var}(Z_i) = \mathbb{E}(Z_i^2) = \sum_j \mathbb{E}(r_{ij}^2 x_j^2) = \|x\|_2^2 = 1$$

$\leq$

$$\mathbb{P}(|\|Z\|_2^2 - k| \geq \varepsilon k) \leq \exp(-C\varepsilon^2 k)$$

$$\mathbb{P}(\|Z\|_2^2 \geq (1+\varepsilon)k) \leq \exp(-\varepsilon^2 \frac{k}{6} + O(k\varepsilon^3))$$

$Y = \|Z\|_2^2$

$1 + \epsilon \geq (1+\epsilon')^2$

$\epsilon' := \epsilon/3$

$$P[Y > k(1+\epsilon)] \leq P[Y > k(1+\epsilon')^2]$$

$$\mathbb{P}(Y > \alpha) = \mathbb{P}(\exp(sY) > \exp(s\alpha)) \leq \exp(-s\alpha)\,\mathbb{E}(\exp(sY))$$

$\alpha := k(1+\epsilon')^2$

ماركوف

$= \prod_i E(\exp(sZ_i^2))$

$Y = \sum_i Z_i^2$

Zi: نرمال

$\mathbb{E}(Z_i) = \sum \mathbb{E}(r_{ij}x_j) = 0$

$\mathrm{var}(Z_i) = \mathbb{E}(Z_i^2) = \sum \mathbb{E}(r_{ij}^2 x_j^2) = \|x\|_2^2 = 1$

$$\mathbb{E}(\exp(sZ_i^2)) = \frac{1}{\sqrt{2\pi}} \int \exp(st^2)\exp(-t^2/2)dt = \frac{1}{\sqrt{1-2s}}$$

$\leq$

$$\mathbb{P}(|\|Z\|_2^2 - k| \geq \varepsilon k) \leq \exp(-C\varepsilon^2 k)$$

$$\mathbb{P}(\|Z\|_2^2 \geq (1+\varepsilon)k) \leq \exp(-\varepsilon^2 \frac{k}{6} + O(k\varepsilon^3))$$

$\boxed{Y = \|Z\|_2^2}$ $1+\epsilon \geq (1+\epsilon')^2$ $\epsilon' := \epsilon/3$

$$P[Y > k(1+\epsilon)] \leq P[Y > k(1+\epsilon')^2]$$

$$\mathbb{P}(Y > \alpha) = \mathbb{P}(\exp(sY) > \exp(s\alpha)) \leq \exp(-s\alpha)\,\mathbb{E}(\exp(sY))$$

$\boxed{\alpha := k(1+\epsilon')^2}$

مارکوف

$$= \boxed{\prod_i E(\exp(sZ_i^2))}$$

$Y = \sum_i Z_i^2$

Zi: نرمال

$$\mathbb{E}(Z_i) = \sum \mathbb{E}(r_{ij}x_j) = 0$$

$$\mathrm{var}(Z_i) = \mathbb{E}(Z_i^2) = \sum \mathbb{E}(r_{ij}^2 x_j^2) = \|x\|_2^2 = 1$$

$$\boxed{\mathbb{E}(\exp(sZ_i^2)) = \frac{1}{\sqrt{2\pi}} \int \exp(st^2)\exp(-t^2/2)dt = \frac{1}{\sqrt{1-2s}}}$$

$$\leq \exp(-s\alpha)(1-2s)^{-k/2}$$

نیمه اول اثبات: $\mathbb{P}(\|Z\|_2^2 \geq (1+\varepsilon)k) \leq \exp(-\varepsilon^2 \frac{k}{6} + O(k\varepsilon^3))$

$$\mathbb{P}(Y \geq \alpha) = \exp(-s\alpha)(1-2s)^{-k/2}$$

$s := \frac{1}{2} - \frac{k}{2\alpha}$

$1 - 2s = \frac{k}{\alpha}$

$\ln(k/\alpha) = \ln(1/(1+\epsilon')^2) = -2\ln(1+\epsilon')$

$\ln(1+x) = x - \frac{x^2}{2} + O(x^3)$

نیمه اول اثبات:

$$\mathbb{P}(\|Z\|_2^2 \geq (1+\varepsilon)k) \leq \exp(-\varepsilon^2 \frac{k}{6} + O(k\varepsilon^3))$$

$$\mathbb{P}(Y \geq \alpha) = \exp(-s\alpha)(1-2s)^{-k/2}$$

$$\boxed{\alpha := k(1+\epsilon')^2}$$

$$s := \frac{1}{2} - \frac{k}{2\alpha}$$

$$1 - 2s = \frac{k}{\alpha}$$

$$\ln(k/\alpha) = \ln(1/(1+\epsilon')^2) = -2\ln(1+\epsilon')$$

$$\ln(1+x) = x - \frac{x^2}{2} + O(x^3)$$

نیمه اول اثبات: $\mathbb{P}(\|Z\|_2^2 \geq (1+\varepsilon)k) \leq \exp(-\varepsilon^2 \frac{k}{6} + O(k\varepsilon^3))$

$$\mathbb{P}(Y \geq \alpha) = \exp(-s\alpha)(1-2s)^{-k/2}$$

$$\alpha := k(1+\epsilon')^2$$

$$s := \frac{1}{2} - \frac{k}{2\alpha}$$

$$1 - 2s = \frac{k}{\alpha}$$

$$= exp\left(-(\frac{1}{2} - \frac{k}{2\alpha})\alpha\right)\left(\frac{k}{\alpha}\right)^{-k/2}$$

$$\ln(k/\alpha) = \ln(1/(1+\epsilon')^2) = -2\ln(1+\epsilon')$$

$$\ln(1+x) = x - \frac{x^2}{2} + O(x^3)$$

نیمه اول اثبات: $\mathbb{P}(\|Z\|_2^2 \geq (1+\varepsilon)k) \leq \exp(-\varepsilon^2 \frac{k}{6} + O(k\varepsilon^3))$

$$\mathbb{P}(Y \geq \alpha) = \exp(-s\alpha)(1-2s)^{-k/2}$$

$$\alpha := k(1+\epsilon')^2$$

$$s := \frac{1}{2} - \frac{k}{2\alpha}$$

$$1 - 2s = \frac{k}{\alpha}$$

$$= exp\left(-(\frac{1}{2} - \frac{k}{2\alpha})\alpha\right)\left(\frac{k}{\alpha}\right)^{-k/2}$$

$$= exp\left(-\frac{1}{2}(\alpha - k) - \frac{k}{2}\ln(k/\alpha)\right)$$

$$\ln(k/\alpha) = \ln(1/(1+\epsilon')^2) = -2\ln(1+\epsilon')$$

$$\ln(1+x) = x - \frac{x^2}{2} + O(x^3)$$

نیمه اول اثبات: $\mathbb{P}(\|Z\|_2^2 \geq (1+\varepsilon)k) \leq \exp(-\varepsilon^2 \frac{k}{6} + O(k\varepsilon^3))$

$$\mathbb{P}(Y \geq \alpha) = \exp(-s\alpha)(1-2s)^{-k/2}$$

$$\boxed{\alpha := k(1+\epsilon')^2}$$

$$s := \frac{1}{2} - \frac{k}{2\alpha}$$

$$1 - 2s = \frac{k}{\alpha}$$

$$= exp\left(-(\frac{1}{2} - \frac{k}{2\alpha})\alpha\right)\left(\frac{k}{\alpha}\right)^{-k/2}$$

$$= exp\left(-\frac{1}{2}(\alpha - k) - \frac{k}{2}\ln(k/\alpha)\right)$$

$$\ln(k/\alpha) = \ln(1/(1+\epsilon')^2) = -2\ln(1+\epsilon')$$

$$= exp\left(-\frac{1}{2}(k(1+\epsilon')^2 - k) + k\ln(1+\epsilon')\right)$$

$$\ln(1+x) = x - \frac{x^2}{2} + O(x^3)$$

نیمه اول اثبات:

$$\mathbb{P}(\|Z\|_2^2 \geq (1+\varepsilon)k) \leq \exp(-\varepsilon^2 \frac{k}{6} + O(k\varepsilon^3))$$

$$\mathbb{P}(Y \geq \alpha) = \exp(-s\alpha)(1-2s)^{-k/2}$$

$$\alpha := k(1+\epsilon')^2$$

$$s := \frac{1}{2} - \frac{k}{2\alpha}$$

$$1 - 2s = \frac{k}{\alpha}$$

$$= exp\left(-(\frac{1}{2} - \frac{k}{2\alpha})\alpha\right)\left(\frac{k}{\alpha}\right)^{-k/2}$$

$$= exp\left(-\frac{1}{2}(\alpha - k) - \frac{k}{2}\ln(k/\alpha)\right)$$

$$\ln(k/\alpha) = \ln(1/(1+\epsilon')^2) = -2\ln(1+\epsilon')$$

$$= exp\left(-\frac{1}{2}(k(1+\epsilon')^2 - k) + k\ln(1+\epsilon')\right)$$

$$\ln(1+x) = x - \frac{x^2}{2} + O(x^3)$$

$$= exp\left(-\frac{k}{2}(2\epsilon' + \epsilon'^2) + k(\epsilon' - \epsilon'^2/2 + O(\epsilon'^3))\right)$$

نیمه اول اثبات: $\mathbb{P}(\|Z\|_2^2 \geq (1+\varepsilon)k) \leq \exp(-\varepsilon^2 \frac{k}{6} + O(k\varepsilon^3))$

$$\mathbb{P}(Y \geq \alpha) = \exp(-s\alpha)(1-2s)^{-k/2}$$

$\alpha := k(1+\epsilon')^2$

$s := \frac{1}{2} - \frac{k}{2\alpha}$

$1 - 2s = \frac{k}{\alpha}$

$$= exp\left(-(\frac{1}{2} - \frac{k}{2\alpha})\alpha\right)\left(\frac{k}{\alpha}\right)^{-k/2}$$

$$= exp\left(-\frac{1}{2}(\alpha - k) - \frac{k}{2}\ln(k/\alpha)\right)$$

$\ln(k/\alpha) = \ln(1/(1+\epsilon')^2) = -2\ln(1+\epsilon')$

$$= exp\left(-\frac{1}{2}(k(1+\epsilon')^2 - k) + k\ln(1+\epsilon')\right)$$

$\ln(1+x) = x - \frac{x^2}{2} + O(x^3)$

$$= exp\left(-\frac{k}{2}(2\epsilon' + \epsilon'^2) + k(\epsilon' - \epsilon'^2/2 + O(\epsilon'^3))\right)$$

$$= exp\left(-k\frac{3}{2}\epsilon'^2 + kO(\epsilon'^3))\right)$$

نیمه اول اثبات: $\mathbb{P}(\|Z\|_2^2 \geq (1+\varepsilon)k) \leq \exp(-\varepsilon^2 \frac{k}{6} + O(k\varepsilon^3))$

$$\mathbb{P}(Y \geq \alpha) = \exp(-s\alpha)(1-2s)^{-k/2}$$

$\alpha := k(1+\epsilon')^2$

$s := \frac{1}{2} - \frac{k}{2\alpha}$

$1 - 2s = \frac{k}{\alpha}$

$$= exp\left(-(\frac{1}{2} - \frac{k}{2\alpha})\alpha\right)\left(\frac{k}{\alpha}\right)^{-k/2}$$

$$= exp\left(-\frac{1}{2}(\alpha - k) - \frac{k}{2}\ln(k/\alpha)\right)$$

$\ln(k/\alpha) = \ln(1/(1+\epsilon')^2) = -2\ln(1+\epsilon')$

$$= exp\left(-\frac{1}{2}(k(1+\epsilon')^2 - k) + k\ln(1+\epsilon')\right)$$

$\ln(1+x) = x - \frac{x^2}{2} + O(x^3)$

$$= exp\left(-\frac{k}{2}(2\epsilon' + \epsilon'^2) + k(\epsilon' - \epsilon'^2/2 + O(\epsilon'^3))\right)$$

$$= exp\left(-k\frac{3}{2}\epsilon'^2 + kO(\epsilon'^3))\right) = exp\left(-\frac{k}{6}\epsilon^2 + O(k\epsilon^3)\right)$$

خلاصه‌سازی JL: جمع‌بندی

$$k = O(\frac{1}{\varepsilon^2}\log\frac{1}{\delta})$$

$$\exp(-C\varepsilon^2 k) = \delta$$

خلاصه‌سازی JL: جمع‌بندی

تقریب نرم ۲ (روش JL)
حافظه: $\epsilon^{-2} \log 1/\delta$ تا عدد

$$k = O(\frac{1}{\varepsilon^2} \log \frac{1}{\delta})$$

$$\exp(-C\varepsilon^2 k) = \delta$$

خلاصه‌سازی مشابه، جمع‌بندی متفاوت

- ماتریس R: n در k
  - Rij: نرمال N (0, 1)
- Rx = [Z1 · · · Zk]
  - Zi: توزیع نرمال، میانگین: ۰، واریانس: $||x||_2^2 = \sum_i x_i^2$
  - یا Zi: G_i $||x||_2$ که Gi نرمال ۰ و ۱

خلاصه‌سازی مشابه، جمع‌بندی متفاوت

- ماتریس R: n در k
  - Rij: نرمال (N (0, 1
- Rx = [Z1 · · · Zk]
  - Zi: توزیع نرمال، میانگین: ۰، واریانس: $\sum_i x_i^2 = ||x||_2^2$
  - یا Zi: $||x||_2 G_i$ که Gi نرمال ۰ و ۱

- روش JL: میانگین Zi^2ها
- روش ما: میانه Ziها؟

$$Y = \frac{\mathrm{median}[|Z_1|, \cdots, |Z_k|]}{\mathrm{median}[|G|]}$$

Lemma 1: Let $U_1 \cdots U_k$ be i.i.d. real random variables chosen from any distribution having continuous c.d.f. F and median M, i.e., $F(t) = Pr[U_i < t]$ and $F(M) = 1/2$. Define $U = \text{median}[U_1, \cdots, U_k]$. Then, for some absolute constant $C > 0$,

$$Pr[F(U) \in (\frac{1}{2} - \epsilon, \frac{1}{2} + \epsilon)] \geq 1 - e^{-C\epsilon^2 k}$$

Lemma 1: Let $U_1 \cdots U_k$ be i.i.d. real random variables chosen from any distribution having continuous c.d.f. F and median M, i.e., $F(t) = Pr[U_i < t]$ and $F(M) = 1/2$. Define $U = \text{median}[U_1, \cdots, U_k]$. Then, for some absolute constant $C > 0$,

$$Pr[F(U) \in (\frac{1}{2} - \epsilon, \frac{1}{2} + \epsilon)] \geq 1 - e^{-C\epsilon^2 k}$$

$E_i : F(U_i) < \frac{1}{2} - \epsilon$

Lemma 1: Let $U_1 \cdots U_k$ be i.i.d. real random variables chosen from any distribution having continuous c.d.f. F and median M, i.e., $F(t) = Pr[U_i < t]$ and $F(M) = 1/2$. Define $U = \text{median}[U_1, \cdots, U_k]$. Then, for some absolute constant $C > 0$,

$$Pr[F(U) \in (\frac{1}{2} - \epsilon, \frac{1}{2} + \epsilon)] \geq 1 - e^{-C\epsilon^2 k}$$

$$E_i : F(U_i) < \tfrac{1}{2} - \epsilon$$

$$p = Pr[E_i] = \tfrac{1}{2} - \epsilon$$

Lemma 1: Let $U_1 \cdots U_k$ be i.i.d. real random variables chosen from any distribution having continuous c.d.f. F and median M, i.e., $F(t) = Pr[U_i < t]$ and $F(M) = 1/2$. Define $U =$ median$[U_1, \cdots, U_k]$. Then, for some absolute constant $C > 0$,

$$Pr[F(U) \in (\frac{1}{2} - \epsilon, \frac{1}{2} + \epsilon)] \geq 1 - e^{-C\epsilon^2 k}$$

$$E_i : F(U_i) < \tfrac{1}{2} - \epsilon$$

$$p = Pr[E_i] = \tfrac{1}{2} - \epsilon$$

$F(U) < \tfrac{1}{2} - \epsilon$ ==> k/2 بار Eiها رخداده

Lemma 1: Let $U_1 \cdots U_k$ be i.i.d. real random variables chosen from any distribution having continuous c.d.f. F and median M, i.e., $F(t) = Pr[U_i < t]$ and $F(M) = 1/2$. Define $U = \text{median}[U_1, \cdots, U_k]$. Then, for some absolute constant $C > 0$,

$$Pr[F(U) \in (\frac{1}{2} - \epsilon, \frac{1}{2} + \epsilon)] \geq 1 - e^{-C\epsilon^2 k}$$

$$E_i : F(U_i) < \frac{1}{2} - \epsilon$$

$$p = Pr[E_i] = \frac{1}{2} - \epsilon$$

$F(U) < \frac{1}{2} - \epsilon$ ==> k/2 بار Eiها رخداده

$$Pr[F(u) < \frac{1}{2} - \epsilon] \leq e^{-C\epsilon^2 k}$$

Lemma 1: Let $U_1 \cdots U_k$ be i.i.d. real random variables chosen from any distribution having continuous c.d.f. F and median M, i.e., $F(t) = Pr[U_i < t]$ and $F(M) = 1/2$. Define $U = \text{median}[U_1, \cdots, U_k]$. Then, for some absolute constant $C > 0$,

$$Pr[F(U) \in (\frac{1}{2} - \epsilon, \frac{1}{2} + \epsilon)] \geq 1 - e^{-C\epsilon^2 k}$$

$$E_i : F(U_i) < \tfrac{1}{2} - \epsilon$$

$$p = Pr[E_i] = \tfrac{1}{2} - \epsilon$$

$F(U) < \frac{1}{2} - \epsilon$ ==> k/2 بار Eiها رخداده

$$Pr[F(u) < \tfrac{1}{2} - \epsilon] \leq e^{-C\epsilon^2 k}$$

هوفدینگ:

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i > (p+\epsilon)n\right) < e^{-2\epsilon^2 n}$$

Lemma 2: Let F be the CDF of a random variable $|G|$, where G drawn from $N(0,1)$. There exists a $C' > 0$ such that if for some z we have

$$F(z) \in (\frac{1}{2} - \epsilon, \frac{1}{2} + \epsilon),$$

برای ϵ کوچک

then

$$z = \text{median}(g) \pm C'\epsilon$$

Theorem: If we use median estimator

$$Y = \frac{\text{median}[|Z_1|, \cdots, |Z_k|]}{\text{median}[|g|]}$$

where $Z_j = \sum_i r_{ij} x_i$, r_{ij} is chosen i.i.d. from $N(0,1)$, then we have

$$Y = ||x||_2 \frac{[\text{median}(g) \pm C'\epsilon]}{\text{median}[|g|]} = ||x||_2(1 \pm C''\epsilon)$$

with probability

$$1 - e^{-C\epsilon^2 k}.$$

Theorem: If we use median estimator

$$Y = \frac{\text{median}[|Z_1|, \cdots, |Z_k|]}{\text{median}[|g|]}$$

where $Z_j = \sum_i r_{ij} x_i$, r_{ij} is chosen i.i.d. from $N(0,1)$, then we have

$$Y = ||x||_2 \frac{[\text{median}(g) \pm C'\epsilon]}{\text{median}[|g|]} = ||x||_2 (1 \pm C''\epsilon)$$

with probability

$$1 - e^{-C\epsilon^2 k}.$$

$$\text{median}[|Z_1|, \cdots, |Z_k|] = U \cdot ||x||_2^2$$

Theorem: If we use median estimator

$$Y = \frac{\text{median}[|Z_1|, \cdots, |Z_k|]}{\text{median}[|g|]}$$

where $Z_j = \sum_i r_{ij} x_i$, r_{ij} is chosen i.i.d. from $N(0,1)$, then we have

$$Y = ||x||_2 \frac{[\text{median}(g) \pm C'\epsilon]}{\text{median}[|g|]} = ||x||_2(1 \pm C''\epsilon)$$

with probability

$$1 - e^{-C\epsilon^2 k}.$$

$$\text{median}[|Z_1|, \cdots, |Z_k|] = U \cdot ||x||^2_2$$

$$Pr[F(U) \in (\frac{1}{2} - \epsilon, \frac{1}{2} + \epsilon)] \geq 1 - e^{-C\epsilon^2 k}$$

Theorem: If we use median estimator

$$Y = \frac{\text{median}[|Z_1|, \cdots, |Z_k|]}{\text{median}[|g|]}$$

where $Z_j = \sum_i r_{ij} x_i$, r_{ij} is chosen i.i.d. from $N(0,1)$, then we have

$$Y = ||x||_2 \frac{[\text{median}(g) \pm C'\epsilon]}{\text{median}[|g|]} = ||x||_2(1 \pm C''\epsilon)$$

with probability

$$1 - e^{-C\epsilon^2 k}.$$

$$\text{median}[|Z_1|, \cdots, |Z_k|] = U \cdot ||x||_2^2$$

$$Pr[F(U) \in (\frac{1}{2} - \epsilon, \frac{1}{2} + \epsilon)] \geq 1 - e^{-C\epsilon^2 k}$$

$$U = \text{median}(g) \pm C'\epsilon$$

تعمیم به LP برای P<2

- برای P=2:
  - U نرمال —> $x_1U_1 + \cdots + x_mU_m$ نرمال

$$(|x_1|^p + \cdots + |x_m|^p)^{\frac{1}{p}} U$$

- برای P=1:
  - توزیع کوشی
- برای P=½: توزیع لوی

تعمیم به LP برای P>2

- برای P=2:
  - U نرمال —> $x_m U_m + \cdots + x_1 U_1$ نرمال
$$(|x_1|^p + \cdots + |x_m|^p)^{\frac{1}{p}} U$$
- برای P=1:
  - توزیع کوشی
- برای P=½: توزیع لوی

روش ۱: دو مرحله‌ای

Pass 1: Pick a stream element $i = i_j$ uniformly at random

Pass 2: Compute x_i

Return $Y = nx_i^{k-1}$

روش ۱: دو مرحله‌ای

Pass 1: Pick a stream element $i = i_j$ uniformly at random

Pass 2: Compute x_i

Return $Y = nx_i^{k-1}$

$$E[Y] = \sum_i \frac{x_i}{n} n x_i^{k-1} = \sum_i x_i^k = F_k$$

روش ۱: دو مرحله‌ای

Pass 1: Pick a stream element $i = i_j$ uniformly at random

Pass 2: Compute x_i

Return $Y = n x_i^{k-1}$

$$E[Y] = \sum_i \frac{x_i}{n} n x_i^{k-1} = \sum_i x_i^k = F_k$$

$$E[Y^2] = \sum_i \frac{x_i}{n} n^2 x_i^{2k-2} = n \sum_i x_i^{2k-1} = n F_{2k-1}$$

روش ۱: دو مرحله‌ای

Pass 1: Pick a stream element $i = i_j$ uniformly at random

Pass 2: Compute x_i

Return $Y = n x_i^{k-1}$

$$E[Y] = \sum_i \frac{x_i}{n} n x_i^{k-1} = \sum_i x_i^k = F_k$$

$$E[Y^2] = \sum_i \frac{x_i}{n} n^2 x_i^{2k-2} = n \sum_i x_i^{2k-1} = n F_{2k-1}$$

چبیشف

$$\mathbb{P}(|X - \mathbb{E}\, X| > \lambda) < \frac{\mathbb{E}(X - \mathbb{E}\, X)^2}{\lambda^2}$$

P[|X-EX| > ϵ EX] < Var[X] / (EX)^2 ϵ^2

روش ۱: دو مرحله‌ای

Pass 1: Pick a stream element $i = i_j$ uniformly at random

Pass 2: Compute x_i

Return $Y = nx_i^{k-1}$

$$E[Y] = \sum_i \frac{x_i}{n} n x_i^{k-1} = \sum_i x_i^k = F_k$$

$$E[Y^2] = \sum_i \frac{x_i}{n} n^2 x_i^{2k-2} = n \sum_i x_i^{2k-1} = nF_{2k-1}$$

چبیشف

$$\mathbb{P}(|X - \mathbb{E}\,X| > \lambda) < \frac{\mathbb{E}(X - \mathbb{E}\,X)^2}{\lambda^2}$$

$P[\ |X-EX| > \epsilon\, EX] < Var[X] / (EX)^2 \epsilon^2$

با k بار
تکرار

$P[\ |X-EX| > \epsilon\, EX] < 1/k \,.\, Var[X] / (EX)^2 \epsilon^2$

روش ۱: دو مرحله‌ای

Pass 1: Pick a stream element $i = i_j$ uniformly at random

Pass 2: Compute x_i

Return $Y = n x_i^{k-1}$

$$E[Y] = \sum_i \frac{x_i}{n} n x_i^{k-1} = \sum_i x_i^k = F_k$$

$$E[Y^2] = \sum_i \frac{x_i}{n} n^2 x_i^{2k-2} = n \sum_i x_i^{2k-1} = n F_{2k-1} \leq m^{1-\frac{1}{k}} (F_k)^2.$$

چبیشف

$$\mathbb{P}(|X - \mathbb{E}\,X| > \lambda) < \frac{\mathbb{E}(X - \mathbb{E}\,X)^2}{\lambda^2}$$

P[|X–EX| > ϵ EX] < Var[X] / (EX)^2 ϵ^2

k= $O(\frac{m^{1-\frac{1}{k}}}{\epsilon^2})$

با k بار تکرار

P[|X–EX| > ϵ EX] < 1/k . Var[X] / (EX)^2 ϵ^2

روش ۱: دو مرحله‌ای

Pass 1: Pick a stream element $i = i_j$ uniformly at random

Pass 2: Compute x_i

Return $Y = nx_i^{k-1}$

$$E[Y] = \sum_i \frac{x_i}{n} n x_i^{k-1} = \sum_i x_i^k = F_k$$

قسمت
باقی‌مانده

$$E[Y^2] = \sum_i \frac{x_i}{n} n^2 x_i^{2k-2} = n \sum_i x_i^{2k-1} = nF_{2k-1} \leq m^{1-\frac{1}{k}} (F_k)^2$$

چبیشف

$$\mathbb{P}(|X - \mathbb{E}\,X| > \lambda) < \frac{\mathbb{E}(X - \mathbb{E}\,X)^2}{\lambda^2}$$

P[|X−EX| > ϵ EX] < Var[X] / (EX)^2 ϵ^2

k= $O(\frac{m^{1-\frac{1}{k}}}{\epsilon^2})$

با k بار
تکرار

P[|X−EX| > ϵ EX] < 1/k . Var[X] / (EX)^2 ϵ^2

روش ۱:

$$
\begin{aligned}
nF_{2k-1} &= n||x||_{2k-1}^{2k-1} \leq n||x||_k^{2k-1} = ||x||_1||x||_k^{2k-1} \\
&\leq m^{1-\frac{1}{k}}||x||_k||x||_k^{2k-1} \\
&= m^{1-\frac{1}{k}}F_k^2
\end{aligned}
$$

روش ۱:

$$nF_{2k-1}$$

$$\begin{aligned} nF_{2k-1} &= n\|x\|_{2k-1}^{2k-1} \leq n\|x\|_k^{2k-1} = \|x\|_1\|x\|_k^{2k-1} \\ &\leq m^{1-\frac{1}{k}}\|x\|_k\|x\|_k^{2k-1} \\ &= m^{1-\frac{1}{k}}F_k^2 \end{aligned}$$

روش ۱:

$$nF_{2k-1} \leq m^{1-\frac{1}{k}}(F_k)^2.$$

$$\begin{aligned} nF_{2k-1} &= n||x||_{2k-1}^{2k-1} \leq n||x||_k^{2k-1} = ||x||_1||x||_k^{2k-1} \\ &\leq m^{1-\frac{1}{k}}||x||_k||x||_k^{2k-1} \\ &= m^{1-\frac{1}{k}}F_k^2 \end{aligned}$$

روش ۱:

قسمت باقی‌مانده

$$nF_{2k-1} \leq m^{1-\frac{1}{k}}(F_k)^2.$$

$$nF_{2k-1} = n||x||_{2k-1}^{2k-1} \leq n||x||_k^{2k-1} = ||x||_1||x||_k^{2k-1}$$

$$\leq m^{1-\frac{1}{k}}||x||_k||x||_k^{2k-1}$$

$$= m^{1-\frac{1}{k}}F_k^2$$

روش ۱:

قسمت
باقی‌مانده

$$nF_{2k-1} \le m^{1-\frac{1}{k}}(F_k)^2.$$

به ازای هر جمله در $||x||_{2k-1}^{2k-1}$ یک جمله
در $||x||_{k}^{2k-1}$ داریم

$$\begin{aligned} nF_{2k-1} &= n||x||_{2k-1}^{2k-1} \le n||x||_{k}^{2k-1} = ||x||_1 ||x||_k^{2k-1} \\ &\le m^{1-\frac{1}{k}} ||x||_k ||x||_k^{2k-1} \\ &= m^{1-\frac{1}{k}} F_k^2 \end{aligned}$$

روش ۱:

قسمت باقی‌مانده

$$nF_{2k-1} \leq m^{1-\frac{1}{k}}(F_k)^2.$$

به ازای هر جمله در $||x||_{2k-1}^{2k-1}$ یک جمله در $||x||_{k}^{2k-1}$ داریم

$$nF_{2k-1} = n||x||_{2k-1}^{2k-1} \leq n||x||_{k}^{2k-1} = ||x||_1||x||_{k}^{2k-1}$$

n همان نرم ۱ است.

$$\leq m^{1-\frac{1}{k}}||x||_k||x||_{k}^{2k-1}$$

$$= m^{1-\frac{1}{k}}F_k^2$$

روش ۲: یک مرحله‌ای

Pass 1: Pick a stream element $i = i_j$ uniformly at random

Pass 2: Compute x_i

از آن‌جا به بعد

Return $Y = nx_i^{k-1}$

$Y' = n(r^k - (r-1)^k)$

r: تقریب از i بعد ==> $E[r] = \frac{x_i+1}{2}$

بهتر

تحلیل:

$$E[Y'] = nE[(r^k - (r-1)^k)]$$

$$= n\frac{1}{n}\sum_i \sum_{j=1}^{x_i} \left[j^k - (j-1)^k\right]$$

$$= \sum_i x_i^k$$

ادامه تحلیل:

$$Y' = n(r^k - (r-1)^k) \le nkr^{k-1} \le kY$$

$$E[Y'^2] \le k^2 E[Y^2] \le k^2 m^{1-\frac{1}{k}} F_k^2$$

جمع‌بندی: الگوریتم برای همه kهای صحیح

تقریب $1 \pm \epsilon$

حافظه: $O(\frac{k^2 m^{1-\frac{1}{k}}}{\epsilon^2})$ عدد

بعد نرم = k