Lecture 9

TTIC 41000: Algorithms for Massive Data

Toyota Technological Institute at Chicago

Spring 2021

Instructor: Sepideh Mahabadi

Announcements

- Project proposals are due on April 30th
- Problem Set 1 is due on May 8th

This Lecture

- ❑ Johnson-Lindenstrauss
- ❑ Lower Bound
- ❑ Fast JL

Johnson-Lindenstrauss Lemma

- Given a set of n points P in $\mathbb{R}^d$, for any $\epsilon \in (0, \frac{1}{2})$, there exists an embedding of the points to $f: \mathbb{R}^d \to \mathbb{R}^m$ where $m = O(\frac{\log n}{\epsilon^2})$ such that
- $\forall x, y \in P, \quad (1-\epsilon)\|x-y\|_2 \leq \|f(x) - f(y)\|_2 \leq (1+\epsilon)\|x-y\|_2$

- $\forall \epsilon, \delta \in (0, \frac{1}{2})$, there exists $D_{\epsilon,\delta}$ on $\mathbb{R}^{m \times d}$ such that $\forall x \in \mathbb{R}^d$, we have that
- $\Pr_{A \sim D_{\epsilon,\delta}}[\|Ax\|_2 \notin [1-\epsilon, 1+\epsilon] \cdot \|x\|_2] \leq \delta$
- $m = O\left(\frac{1}{\epsilon^2} \cdot \log \frac{1}{\delta}\right)$

➢ It is enough to apply this on all $u = x - y$ where $x, y \in P$

Mapping

- $\forall \epsilon, \delta \in (0, \frac{1}{2})$, there exists $D_{\epsilon,\delta}$ on $\mathbb{R}^{m \times d}$ such that $\forall x \in \mathbb{R}^d$, we have that
- $\Pr_{A \sim D_{\epsilon,\delta}}[\|Ax\|_2 \notin [1-\epsilon, 1+\epsilon] \cdot \|x\|_2] \leq \delta$
- $m = O(\frac{1}{\epsilon^2} \cdot \log \frac{1}{\delta})$

Examples of such distributions:

- Projection onto a random m dimensional subspace (best constants)
- Take A to be a matrix where each entry is chosen iid from $\mathcal{N}(0,1)$ (then normalized)
- … (more in this lecture)

Normal Distribution

- $\mathcal{N}\left(\mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma}} \exp(-\frac{(x-\mu)^2}{2\sigma^2})$
- If X and Y are independent random variable with normal distribution then $X + Y$ has normal distribution $\mathcal{N}\,(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$

Mapping

- Let A be a matrix where every entry is picked iid from $\mathcal{N}(0,1)$
- Then output $\|Ax\|_2^2/m$ as an approximation for $\|x\|_2^2$
- $\mathbb{E}\left[\frac{\|Ax\|_2^2}{m}\right] = \frac{1}{m} \cdot \mathbb{E}\left(x^T A^T A x\right) = \frac{1}{m} x^T \mathbb{E}\left(A^T A\right) x = \|x\|_2^2$
- $\mathbb{E}\left(A^T A\right)$ is a diagonal matrix with all entries on the diagonal are m, i.e., for any j, $\mathbb{E}\left[\sum_i A_{i,j}^2\right] = \sum_i \mathbb{E}\left[A_{i,j}^2\right] = m$ as the mean is 0 and variance is 1.
- Off diagonal entries are 0, $\mathbb{E}\left[\sum_i A_{j,i} A_{i,h}\right] = \sum_i \mathbb{E}[A_{j,i} A_{i,h}]$ as they are independent and the means are 0

Concentration

- Let A be a matrix where every entry is picked iid from $\mathcal{N}(0,1)$
- Then output $\|Ax\|_2^2/m$ as an approximation for $\|x\|_2^2$
- $\Pr\left[\left|\|Ax\|_2^2 - m\|x\|_2^2\right| \geq \epsilon m\|x\|_2^2\right] \leq \exp\left(-C\epsilon^2 m\right) \leq \delta$
- One side: $\Pr[\|Ax\|_2^2 \geq (1+\epsilon)m\|x\|_2^2]$
- Assume $\|x\|_2^2 = 1$, let $Z = Ax$ then $\Pr[\|Z\|_2^2 \geq (1+\epsilon)m] \leq \exp(-\epsilon^2 m + O\left(m\epsilon^3\right))$
- Let $Y = \|Z\|_2^2$, then $\Pr(Y > \alpha) = \Pr[\exp(sY) > \exp(s\alpha)] \leq \exp(-s\alpha)\, \mathbb{E}[\exp(sY)]$ by Markov
- By independence, $\mathbb{E}(\exp(sY)) = \prod_i \mathbb{E}(\exp(sZ_i^2))$
- Z_i has also normal distribution, we can compute $\mathbb{E}(Z_i) = 0$ and $\mathrm{Var}(Z_i) = 1$

Concentration

- Assume $\|x\|_2^2 = 1$, let $Z = Ax$ then $\Pr[\|Z\|_2^2 \geq (1+\epsilon)m] \leq \exp(-\epsilon^2 m + O\left(m\epsilon^3\right))$
- Let $Y = \|Z\|_2^2$, then $\Pr(Y > \alpha) = \Pr[\exp(sY) > \exp(s\alpha)] \leq \exp(-s\alpha)\, \mathbb{E}[\exp(sY)]$ by Markov
- By independence, $\mathbb{E}(\exp(sY)) = \prod_i \mathbb{E}(\exp(sZ_i^2))$
- Z_i has also normal distribution, we can compute $\mathbb{E}(Z_i) = 0$ and $\mathrm{Var}(Z_i) = 1$
- We can analytically compute $\mathbb{E}\left[\exp\left(sZ_i^2\right)\right] = \frac{1}{\sqrt{2\pi}} \int \exp\left(st^2\right) \exp(-\frac{t^2}{2})\, dt \ = \frac{1}{\sqrt{1-2s}}$
- So we get $\Pr[Y \geq \alpha] = \exp(-s\alpha) \cdot (1-2s)^{-\frac{m}{2}}$
- Set $s = \frac{1}{2} - \frac{m}{2\alpha}$ so $1 - 2s = \frac{m}{\alpha}$

Mapping - concentration

- Assume $\|x\|_2^2 = 1$, let $Z = Ax$ then $\Pr[\|Z\|_2^2 \geq (1+\epsilon)m] \leq \exp(-\epsilon^2 m + O(m\epsilon^3))$
- Let $Y = \|Z\|_2^2$, then $\Pr(Y > \alpha) = \Pr[\exp(sY) > \exp(s\alpha)] \leq \exp(-s\alpha)\,\mathbb{E}[\exp(sY)]$ by Markov
- So we get $\Pr[Y \geq \alpha] = \exp(-s\alpha) \cdot (1-2s)^{-\frac{m}{2}}$
- Set $s = \frac{1}{2} - \frac{m}{2\alpha}$ so $1 - 2s = \frac{m}{\alpha}$
- $\Pr[Y \geq \alpha] = \exp(-\frac{\alpha}{2}\left(1 - \frac{m}{\alpha}\right)) \cdot \left(\frac{m}{\alpha}\right)^{-\frac{m}{2}} = \exp\left(\frac{m-\alpha}{2}\right)\left(\frac{m}{\alpha}\right)^{-\frac{m}{2}}$
- and set $\alpha = m(1+\epsilon)^2$
- $\Pr[Y \geq \alpha] = \exp\left(\frac{m-\alpha}{2}\right)\left(\frac{m}{\alpha}\right)^{-\frac{m}{2}} = \mathrm{e}^{-\epsilon m - \frac{\epsilon^2}{2}m} e^{-\frac{m}{2}\ln(\frac{m}{\alpha})} = \mathrm{e}^{-\epsilon m - \frac{\epsilon^2}{2}m} e^{-\frac{m}{2}\ln(\frac{1}{(1+\epsilon)^2})} = \mathrm{e}^{-\epsilon m - \frac{\epsilon^2}{2}m} e^{m\ln(1+\epsilon)} = \mathrm{e}^{m(-\epsilon - \frac{\epsilon^2}{2} + \epsilon - \frac{1}{2}\epsilon^2 + O(\epsilon^3))}$ using Taylor's expansion for $\ln(1+x) = x - \frac{x^2}{2} + O(x^3)$
- $\Pr[Y \geq \alpha] = \mathrm{e}^{m(-\epsilon - \frac{\epsilon^2}{2} + \epsilon - \frac{1}{2}\epsilon^2 + O(\epsilon^3))} = e^{-m\epsilon^2 + mO(\epsilon^3)}$

JL Lowerbound

- Consider pointset $X = \{0, e_1, \dots, e_n\} \subseteq \mathbb{R}^n$

Claim. If we embed these m-dimensional space and preserve distances up to a factor of c, the target dimension has to be at least $\frac{\log n}{\log(2c+1)}$.

- Wlog, assume that zero is mapped to zero (otherwise translate the instance)
- Distances are preserved; points should have distance in [1,c] from zero and distance in $[\sqrt{2}, c\sqrt{2}]$ from each other. This means that the ball of radius $\frac{1}{2}$ around all points and zero are disjoint.
- By a volume argument, $n\ vol_m\left(B\left(\frac{1}{2}\right)\right) \leq vol_m(B(c+\frac{1}{2}))$ which implies that $n \leq \frac{vol_m\left(B\left(c+\frac{1}{2}\right)\right)}{vol_m\left(B\left(\frac{1}{2}\right)\right)} = (2c+1)^m$.
- Thus, $m \geq \frac{\log n}{\log(2c+1)}$

Fast JL

- $A = \sqrt{\frac{d}{m}} \cdot SHD$
  - D is a $d \times d$ diagonal with iid ± 1 on the diagonal (Rademacher)
  - H is the $d \times d$ normalized Hadamard matrix (divide entries by $1/\sqrt{d}$)
    - Every entry of H is ± 1
    - Every two rows are perpendicular
    - In all rows except the first one, the number of $+1$ is equal to -1
    - $H_1 = (1), H_2 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \; H_{2t} = \begin{pmatrix} H_t & H_t \\ H_t & -H_t \end{pmatrix}$
  - S is a $m \times d$ sampling matrix with replacement (each row has a 1 at a uniformly random location and zeroes elsewhere).

➢ Time: D can be applied in $O(d)$ time and H in $O(d \log d)$ time, and S in $O(m)$ time
  - In compare to previous case where it takes $O(md)$ time

➢ $m = O(\frac{1}{\epsilon^2} \cdot \log\left(\frac{1}{\delta}\right) \cdot \log\frac{d}{\delta})$.
  ➢ for optimal m, instead use $\Pi'\Pi$ where Π is FJLT and Π' is an optimal JL with $m' = O(\epsilon^{-2}\log(1/\delta))$. Runtime increases by additive $m \cdot m'$.

Proof of Fast JL

- Define $y = HDx$. We show that $\|y\|_\infty = O(\sqrt{\log(d/\delta)/d})$ w.p. $1 - \delta/2$.
- $y_i = (HDx)_i = \sum_{j=1}^{d} \sigma_j \cdot \left(\frac{1}{\sqrt{d}} \gamma_{i,j} x_j\right) = \langle \sigma, z^i \rangle$
  - $|\gamma_{i,j}| = 1$ and z^i is a vector with $\left(z^i\right)_j = \frac{1}{\sqrt{d}} \gamma_{i,j}\, x_j$

Khinchine's inequality. $X_1, \dots, X_n$ i.i.d Rademacher, for $a_1, \dots, a_n \in \mathbb{R}$, and $\lambda > 0$, $\Pr\left(\left|\sum_{i=1}^{n} a_i X_i\right| > \lambda \|a\|_2\right) \le 2e^{-\lambda^2/2}$

- By Khintchine's inequality, setting $X_j = \sigma_j$, $a_j = \left(z^i\right)_j = \frac{1}{\sqrt{d}} \gamma_{i,j}\, x_j$, $\|a\|_2 = \left(\frac{1}{\sqrt{d}}\right) \|x\|$, $\lambda = \sqrt{2\log(4d/\delta)\,/d}$
  - $\forall i, \Pr[|y_i| > \sqrt{2\log(4d/\delta)\,/d}] < 2e^{-\log\left(\frac{d}{\delta}\right)} = \frac{\delta}{2d}$
- By union bound, $\Pr[\|y\|_\infty > \sqrt{2\log(4d/\delta)\,/d}] < \frac{\delta}{2}$, and thus $\|y\|_\infty^2 \le \frac{2\log\left(\frac{4d}{\delta}\right)}{d} := \frac{\tau}{d}$
- Using Chernoff –type arguments
  - $\|y\|_2^2 = \|x\|_2^2$ (as D changes the sign of entries in x, and H can be viewed as change of basis matrix)
  - Each y_i is small and thus small variance.
  - Each row of S is sampling one y_i uniformly at random.
  - $\Pr\left[\frac{d}{m}\|Sy\|_2^2 \approx (1+\epsilon)\|y\|_2^2\right] \ge 1 - \delta$

Next Lecture

- Approximate Nearest Neighbor