EE270
Large scale matrix computation, optimization and learning

Instructor : Mert Pilanci

Stanford University

Tuesday, Jan 21 2020

Randomized Linear Algebra
Lecture 5: Randomized Dimension Reduction: Johnson Lindenstrauss Lemma

Recap: Verifying Matrix Multiplication

- Given three $n \times n$ matrices A, B, M
- verify whether
$$AB = M$$
- Sample a random vector $r = [r_1, ..., r_n]^T$
- Compute ABr by first computing Br and then $A(Br)$
- Compute Mr
- If $A(Br) \neq Mr$, then $AB \neq M$
- Otherwise, return $AB = M$

Recap: Verifying Matrix Multiplication

- Given three $n \times n$ matrices A, B, M
- verify whether
$$AB = M$$
- Sample a random vector $r = [r_1, ..., r_n]^T$
- Compute ABr by first computing Br
  and then $A(Br)$
- Compute Mr
- If $A(Br) \neq Mr$, then $AB \neq M$
- Otherwise, return $AB = M$
- Complexity: three matrix-vector multiplications $O(n^2)$
  Freivalds' Algorithm (1977)

Recap: Failure Probability

- Let $r = [r_1, ..., r_n]^T$ be i.i.d. from a discrete distribution taking k distinct values each with probability $\frac{1}{k}$
- Lemma $\mathbb{P}[ABr = Mr] \leq \frac{1}{k}$

Dimension Reduction

- map a high dimensional vector to low dimensions such that certain properites are preserved
- examples so far:
- Approximate Matrix Multiplication $AS^TSB \approx AB$ where S is random
- Freivalds Algorithm $ABr - Mr$ where r is random
- Trace estimation $r^TMr \approx \mathbf{tr}(M)$ where r is random

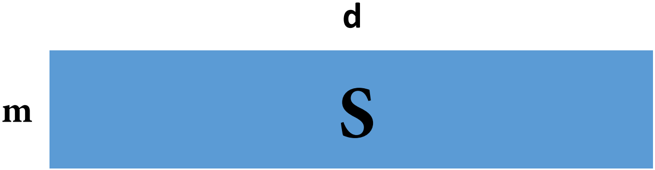
Dimension Reduction

- map a high dimensional vector to low dimensions such that certain properites are preserved
- examples so far:
- Approximate Matrix Multiplication $AS^TSB \approx AB$ where S is random
- Freivalds Algorithm $ABr - Mr$ where r is random
- Trace estimation $r^TMr \approx \mathbf{tr}(M)$ where r is random

- Generic dimension reduction problem
- Given vectors $x_1, ..., x_n \in \mathbb{R}^d$, compress the data points into low dimensional representation $y_1, ..., y_n \in \mathbb{R}^m$ where $m < d$
- another instance is Principal Component Analysis

Randomized Dimension Reduction

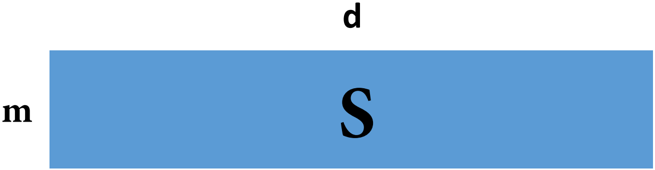
- Given vectors $x_1, \ldots, x_n \in \mathbb{R}^d$, compress the data points into low dimensional representation $y_1, \ldots, y_n \in \mathbb{R}^m$ where $m < d$



- Linear transformation $y_i = Sx_i$ for $i = 1, \ldots, n$
- S is chosen randomly

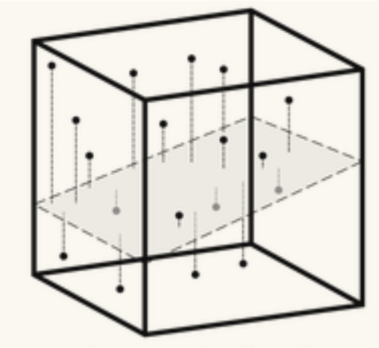
Randomized Dimension Reduction

- Given vectors $x_1, \dots, x_n \in \mathbb{R}^d$, compress the data points into low dimensional representation $y_1, \dots, y_n \in \mathbb{R}^m$ where $m < d$



- Linear transformation $y_i = Sx_i$ for $i = 1, \dots, n$
- S is chosen randomly
- Approximate Matrix Multiplication: $AS^TSB \approx AB$
  where S is random matrix

Geometry of Random Projections



Johnson Lindenstrauss Lemma

- Let $\epsilon \in (0, \frac{1}{2})$. Given any set of points $\{x_1, ..., x_n\}$ in $\mathbb{R}^d$, there exists a map $S : \mathbb{R}^n \rightarrow \mathbb{R}^m$ with $m = \frac{9\log(n)}{\epsilon^2 - \epsilon^3}$ such that

$$1 - \epsilon \leq \frac{\|Sx_i - Sx_j\|_2^2}{\|x_i - x_j\|_2^2} \leq 1 + \epsilon$$

Johnson Lindenstrauss Lemma

- Let $\epsilon \in (0, \frac{1}{2})$. Given any set of points $\{x_1, ..., x_n\}$ in $\mathbb{R}^d$, there exists a map $S : \mathbb{R}^n \to \mathbb{R}^m$ with $m = \frac{9\log(n)}{\epsilon^2 - \epsilon^3}$ such that

$$1 - \epsilon \leq \frac{\|Sx_i - Sx_j\|_2^2}{\|x_i - x_j\|_2^2} \leq 1 + \epsilon$$

- Note that the target dimension m is **independent of the original dimension** d, and depends **only on the number of points** n and the accuracy parameter.

Johnson Lindenstrauss Lemma

- Let $\epsilon \in (0, \frac{1}{2})$. Given any set of points $\{x_1, ..., x_n\}$ in $\mathbb{R}^d$, there exists a map $S : \mathbb{R}^n \to \mathbb{R}^m$ with $m = \frac{9\log(n)}{\epsilon^2 - \epsilon^3}$ such that

$$1 - \epsilon \leq \frac{\|Sx_i - Sx_j\|_2^2}{\|x_i - x_j\|_2^2} \leq 1 + \epsilon$$

- Note that the target dimension m is **independent of the original dimension** d, and depends **only on the number of points** n and the accuracy parameter.
- more surprises: picking an $m \times d$ random matrix $S = \frac{1}{\sqrt{m}} G$ with $G_{ij} \sim N(0, 1)$ standard normal works with high probability!

Johnson Lindenstrauss (JL) Lemma

- Define $u_{ij} \triangleq \frac{x_i - x_j}{\|x_i - x_j\|_2}$.
- note that $\|u_{ij}\|_2 = 1$

- JL Lemma:
  $\mathbb{P}\left[\|Su_{ij}\|_2^2 \in (1 \pm \epsilon) \text{ for all } i, j \in \{1, ..., n\}\right] \geq 1 - \delta$

  where $\delta \in (0, 1)$ for large enough m

Proof of JL Lemma

- We need to show $\|Su_{ij}\|_2^2$ is concentrated around 1
- **Lemma** Let $S_{ij} \sim \frac{1}{\sqrt{m}} N(0,1)$ and u be any fixed vector. Then

$$\mathbb{E}\|Su\|_2^2 = \|u\|$$

- implies that the distance between two points is preserved in expectation
- **Proof:**

Concentration of Measure for Uniform Distribution on the Sphere

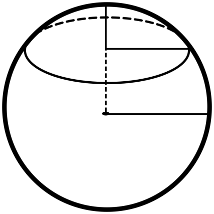
- Suppose $m = 1$, i.e., we project to dimension one
- S is a row vector $S = g^T \in \mathbb{R}^d \sim N(0, I)$
- $\mathbb{P}\left[|g^T u| \geq \epsilon\right] = \mathbb{P}\left[|g^T e_1| \geq \epsilon\right] = \mathbb{P}\left[|g_1| \geq \epsilon\right]$ where e_1 is the first ordinary basis vector

Concentration of Measure for Uniform Distribution on the Sphere

- Suppose $m = 1$, i.e., we project to dimension one
- S is a row vector $S = g^T \in \mathbb{R}^d \sim N(0, I)$
- $\mathbb{P}\left[|g^T u| \geq \epsilon\right] = \mathbb{P}\left[|g^T e_1| \geq \epsilon\right] = \mathbb{P}\left[|g_1| \geq \epsilon\right]$ where e_1 is the first ordinary basis vector
- **Lemma:** $\mathbb{P}\left[|s_1| \geq \frac{t\|g\|_2}{\sqrt{d}}\right] \leq 2e^{-\frac{t^2}{2}}$.
- Note that $\frac{g}{\|g\|_2}$ is distributed uniformly on the unit sphere

Concentration of Measure for Uniform Distribution on the Sphere

- **Lemma:** $\mathbb{P}\left[|g_1| \geq \frac{t\|g\|_2}{\sqrt{d}}\right] \leq 2e^{-\frac{t^2}{2}}$.
- Note that $\frac{g}{\|g\|_2}$ is distributed uniformly on the unit sphere



- Pythagorean theorem: $\frac{t^2}{d} + R_{\text{cap}}^2 = 1$ implies $R_{\text{cap}} = \sqrt{1 - \frac{t^2}{d}}$
- $\mathbb{P}\left[|\frac{g_1}{\|g\|_2}| \geq \frac{t}{\sqrt{d}}\right] \leq \frac{\text{area of the spherical cap}}{\text{area of the sphere}} \leq \frac{\left(\sqrt{1-\frac{t^2}{d}}\right)^{d-1}}{1^{d-1}}$
- using the fact $(1 - \frac{x}{n})^n \leq e^{-x}$ we get
  $\mathbb{P}\left[|\frac{g_1}{\|g\|_2}| \geq \frac{t}{\sqrt{d}}\right] \leq 2e^{-\frac{t^2}{2}}$.

Proof of JL Lemma

- Back to the general case $S \in \mathbb{R}^{m \times d}$
- Consider the probability that $\|Su\|_2^2$ deviates from 1, i.e., projected vectors are stretched more than their expectation

$$\mathbb{P}\left[\|Su\|_2^2 \geq (1+\epsilon)\|u\|_2^2\ \right]$$

Questions?

References

- Lecture notes on randomized linear algebra, Michael Mahoney https://arxiv.org/pdf/1608.04481
- Lecture notes, Jelani Nelson https://www.sketchingbigdata.org/fall17/lec/lec3.pdf
- Lecture notes, Aleksander Madry https://people.csail.mit.edu/madry/gems/notes/lecture21.pdf