COMS 4995-8: Advanced Algorithms (Spring'21)	Feb 02, 2021
Lecture 7: Dimension Reduction	
Instructor: *Alex Andoni*	Scribes: *Zixuan Zhang, Linxiao Wu*

1 Introduction:

This lecture mainly focuses on dimension reduction: Johnson-Linderstrauss Lemma and especially its distributional version. Chi-squared distribution is introduced when there is a sum of Gaussian distributed variables.

2 Last time Recap

Tug-of-War+ Algorithm

1. Frequency vector $f \in \mathbb{R}_+^n$
2. Goal: estimate $F_2 = \|f\|_2^2$
3.

Algorithm 1: Tug-of-War+ Algorithm: repeat ToW k times, and take the average of the estimators.

for $i = 1, 2, \ldots, k$, *where* $k = O(\frac{\lg n}{\epsilon^2})$ **do**
 pick $r_{ij} \in \{\pm 1\}$, $j \in [n]$
 sketch: $Z_i = \sum_{j=1}^{n} r_{ij} f_j$
end
return Estimator: $Z := \frac{1}{k} \sum_{i=1}^{k} Z_i^2$

Better complexity

If probability fails $\leq \delta$ where δ is some parameter.

$$Pr[Z^2 = (1 \pm \epsilon) F_2] \geq 1 - \delta \tag{1}$$

as long as $k \geq \Omega(\frac{1}{\delta} \cdot \frac{1}{\epsilon^2})$
Question to today's lecture: Can we get better dependence on δ?

Claim 1. *Yes, and in fact we can get* $k = O(\frac{\lg(1/\delta)}{\epsilon^2})$

3 Dimension Reduction

Definition 1:(Sketch function). *For $x \in \mathbb{R}^n$, a sketching function $\varphi : \mathbb{R}^n - > \mathbb{R}^k$ is defined as:*

$$\varphi(x) = \sigma \cdot x = \frac{1}{\sqrt{k}}(\sum \sigma_{1i} x_i, \sum \sigma_{2i} x_i, ..., \sum \sigma_{ki} x_i)$$

Note that $\sigma \in \mathbb{R}^{k \times n}$. We can use φ as an estimator of the L2 norm of its argument:

$$\|\varphi(f)\|_2^2 = (1 \pm \epsilon) F_2$$

Definition 2:(Linearity). *If sketching function φ is linear, the following properties hold true:*

$$\varphi(x \pm y) = \varphi(x) \pm \varphi(y)$$

3.1 Johnson-Lindenstrauss Lemma

Lemma: $\forall \epsilon > 0$, *there exists a linear, randomized function φ such that: $\forall x, y \in \mathbb{R}^n$:*

$$Pr[\|\varphi(x) - \varphi(x)\| \in (1 \pm \epsilon)\|x - y\|] \geq 1 - e^{\frac{-\epsilon^2 k}{9}}$$

Proof. Because φ is linear, we have $\varphi(x) - \varphi(x) = \varphi(x - y)$. If we define $z = x - y$, the original lemma is equivalent to:

$$Pr[\|\varphi(z)\| \in (1 \pm \epsilon)\|z\|] \geq 1 - e^{\frac{-\epsilon^2 k}{9}}$$

Note that there could be various implementation of φ:

- we could use Tug-of-war algorithm where $\sigma_{ij} = \pm 1$
- φ maps $\mathbb{R}^n$ to a random k-dim linear subspace
- Similar to Tug-of-war but $\sigma_{ij} \sim N(0, 1)$

Recall: We proved the correctness of Tug-of-War by showing:

- $\mathbb{E}[\sigma_i] = 0$
- $\mathbb{E}[\sigma_i^2] = 1$
- $\mathbb{E}[\sigma_i^4] = 1$

Here we adopt the third implementation, which also satisfies the above conditions. Before moving onto the rest of the proof, we must understand the following properties:

Definition(Stability Property):

$$\sum_{i=1}^{n} g_i x_i \sim \|x\|_2 \cdot a = \sum x_i^2 \cdot a; \ where\ a \sim N(0,1)$$

Property(Spherically Symmetric): For a vector $b = (b_1, b_2, ..., b_n)$ where $b_i \sim N(0,1)$ are i.i.d

$$pdf(b) = \Pi \frac{1}{\sqrt{2\pi}} e^{\frac{-b_i^2}{2}} = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{\frac{-\|g\|_2^2}{2}}$$

Now, let's rewrite the definition of φ:

$$\begin{aligned}
\varphi(z) &= \frac{1}{\sqrt{k}}(\sum_j g_{1j}z_j, \sum_j g_{2j}z_j, ..., \sum_j g_{kj}z_j) \\
&= \frac{1}{\sqrt{k}}(\|z\|g^{(1)}, \|z\|g^{(2)}, ..., \|z\|g^{(k)}) \\
&= \frac{\|z\|}{\sqrt{k}}(g^{(1)}, g^{(2)}, ..., g^{(k)}) \rightarrow k - dim\ Gaussian \\
\|\varphi(z)\|^2 &= \frac{1}{k}\sum_{j=1}^{k}(g^{(j)})^2 \cdot \|z\|^2 \\
&= \|z\|^2 \frac{1}{k}\sum_j (g^{(j)})^2 \\
&= \|z\|^2 \boldsymbol{\chi_k^2}, where\ \chi\ is\ the\ Chi\ square\ distribution
\end{aligned}$$

Fact:

$$P[\chi_k^2 \notin (1 \pm \epsilon)] \leq e^{\frac{-\epsilon^2 k}{9}}$$

Therefore, we have:

$$\begin{aligned}
Pr[\|\varphi(z)\| \in (1 \pm \epsilon)\|z\|] &= 1 - Pr[\|\varphi(z)\| \notin (1 \pm \epsilon)\|z\|] \\
&\geq 1 - e^{\frac{-\epsilon^2 k}{9}}
\end{aligned}$$

Q.E.D □

Corollary 2. *(of Johnson-Linderstrauss '84)*

- *fix* N *vectors* $x_1, x_2, \ldots, x_N \in \mathbb{R}^n$*, pick* Φ *as in Johnson-Linderstrauss, with* $k = \Theta(\frac{\lg N}{\epsilon^2})$
- $Pr_\Phi[\forall i, j \in [N], \|\Phi x_i - \Phi x_j\| = (1 \pm \epsilon)\|x_i - x_j\|] \geq 1 - \frac{1}{N}$

Proof. fix $k = \frac{3 \cdot 9 \cdot \ln N}{\epsilon^2}$

Then, by Johnson-Linderstrauss, $\forall i, j \in [N]$

$$Pr[\|\Phi x_i - \Phi x_j\| \notin (1 \pm \epsilon)\|x_i - x_j\|] \leq e^{\frac{-\epsilon^2 k}{9}} = \frac{1}{N^3}$$

Recall the union bound theorem

$$\mathbb{P}\left(\bigcup_i A_i\right) \leq \sum_i \mathbb{P}(A_i).$$

By Union bound over all pairs for $i, j \in [N]$

$$\begin{aligned} Pr_\Phi[\exists i, j \ pair \in [N], \|\Phi x_i - \Phi x_j\| \notin (1 \pm \epsilon) \|x_i - x_j\|] &\leq \sum_{i,j \ pairs} Pr[\|\Phi x_i - \Phi x_j\| \notin (1 \pm \epsilon) \|x_i - x_j\|] \\ &\leq N^2 \times \frac{1}{N^3} \\ &= \frac{1}{N} \end{aligned}$$

□

Rephrase: For $\forall$ N vectors in $\mathbb{R}^n$, can map them in to $\mathbb{R}^k$, where k = $O(\frac{\lg N}{\epsilon^2})$ while preserving distance $1 \pm \epsilon$

Dimension Reduction in Other space

Remark: Can we do the same for l_1?

Recall that

$$l_1^d : \mathbb{R}^d \ where \ \|x - y\|_1 = \sum_{i=1}^{n} \|x_i - y_i\|$$

For l_1: N vectors into lower dimension l_1:

$$k = N^{\Omega(\frac{1}{\alpha})}$$

for α-approximation.