COMS 4995-3: Advanced Algorithms Feb 08, 2017

# Lecture 7 – Dimension Reduction and Johnson-Linderstrauss Lemma

Instructor: *Alex Andoni* Scribes: *Jiahui Liu, David Pu*

## 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

**Tug-of-War:**

- for frequency vector $f \in \mathbb{R}$:
- pick random $\sigma_i \in \{\pm 1\}$
- $z_i = \sum_{i=1} \sigma_i f_i$
- Estimator: $z^2$

**Tug-of-War+ : k estimators**

$$z_j = \sum_{j=1} \sigma_{ij} f_i, j = 1, ...., k$$

Estimator:

$$\frac{1}{k} \sum {z_j}^2$$

## 3 Dimension Reduction

**Definition 1** (Sketching function). *For $\bar{x} \in \mathbb{R}^n$, $\bar{x} = (x_1, x_2, ...., x_n)$, a sketching function $\varphi : \mathbb{R}^n \to \mathbb{R}^k$ is defined as*

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

**Definition 2** (Linear Property). *$\varphi$ is linear if:*

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

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

Estimator:

$$\varphi(x) \to \|\varphi(x)\|^2 = \frac{1}{k}\Sigma_{j=1}{z_j}^2$$

$$\varphi(x) = \frac{1}{\sqrt{k}}(z_i, z_2, ...z_k)$$

Given sketches $\varphi(x)$ and $\varphi(y)$:
we can compute

$$\|\varphi(x) - \varphi(y)\|_2{}^2 = \|\varphi(x-y)\|_2{}^2 = (1 \pm \epsilon)\|x-y\|^2{}_2 = (1 \pm \epsilon)\sum_{i=1}^{n}(x_i - y_i)^2$$

## 3.1 Johnson-Lindenstrauss Lemma

**Lemma 3** (Distributional Johnson-Lindenstrauss 1984)**.**

$$\forall \epsilon > 0, \ \textit{there is a randomized } \varphi : (R)^n \to (R)^k \textit{ such that } \forall x, y \in (R)^n$$

*we have*

$$P\left[\|\varphi(x) - \varphi(y)\| \in (1 \pm \epsilon)\|x - y\|_2\right] \geq 1 - e^{\frac{\epsilon^2 k}{9}}$$

*(*$e^{\frac{\epsilon^2 k}{9}}$ *is the failure probability.)*

In original Johnson-Lindenstrauss lemma: $\varphi :$ a random k-dimensional subspace.

*Proof.*

Take

$$\varphi(x) = (\sum_{i=1}^{n} g_{1i}x_i, \sum_{i=1}^{n} g_{2i}x_i, ....., \sum_{i=1}^{n} g_{ki}x_i)\frac{1}{\sqrt{k}}$$

Each $g_{ji}$ is a Gaussian/normal N(0,1):

$$pdf(g) = \frac{1}{2\pi}e^{-\frac{g^2}{2}}$$

Recall: What did we use to prove the correctness of Tug-of-War?

(1) $E[\sigma_i] = 0$

(2) $E[{\sigma_i}^2] = 1$

(3) $E[{\sigma_i}^4] = 1$

This is satisfied by $\sigma_i \in \{\pm 1\}$, but also by the Gaussian/normal random variable.

Consider $k = 1$:

$$\varphi(x) = \sum g_i x_i$$

**Definition 4** (Stability Property)**.**

$$\sum_{i=1}^{k} g_i x_i \sim \|x\|_2 \cdot a = (\sum x_i^2)^{\frac{1}{2}} \cdot a$$

*a is another Gaussian N(0,1)*

□

The probability density distribution for a centrally spherically symmetric vector $\bar{g} = (g_1, \cdots, g_n)$

$$pdf(\bar{g}) = (\frac{1}{\sqrt{2\pi}})^n \cdot e^{\frac{-g_1^2}{2}} \cdot e^{\frac{-g_2^2}{2}} \cdot \cdots \cdot e^{\frac{-g_n^2}{2}} = (\frac{1}{\sqrt{2\pi}})^n \cdot e^{\frac{-\sum_{i=1}^{n} g_i^2}{2}}$$

$\bar{g} \cdot x$ is distributed as $\bar{g}' \cdot \big(||X||_x, 0, 0, \cdots, 0\big) = g_1' \cdot ||x||_2$

**General** $k$:

$$||\phi(x) - \phi(y)|| = ||\phi(x - y)|| \approx ||x - y||_2 \leftarrow ||z||_2 \text{ where } z = x - y$$

fix $z$:

$$\phi(z) = \frac{1}{\sqrt{k}} \cdot \left( \sum g_{1i} z_i, \cdots \sum g_{ki} z_i \right) \sim \frac{1}{\sqrt{k}} \cdot \left( a_1 \cdot ||z||, a_2 \cdot ||z|, \cdots, a_k \cdot ||z|| \right)$$

where each $a_i$ is Gaussian distributed

$$\begin{aligned} ||\phi(z)||_2^2 &= \frac{1}{k} \sum_{j=1}^{k} a_j^2 \cdot ||z||^2 \\ &= ||z||^2 \cdot \mathbf{\frac{1}{k} \sum_{j=1}^{k} a_j^2} \\ &= ||z||^2 \cdot \mathcal{X}_k^2 \end{aligned}$$

This is $\mathcal{X}^2$ (Chi-squared) distributed with $k$ degrees of freedom.

**Fact:**

$$P\big[\mathcal{X}_k^2 \notin (1 \pm \epsilon)\big] \leq 2 \cdot e^{\frac{-k}{4}(\epsilon^2 - \epsilon^3)}$$

for $\epsilon < \frac{1}{2}$ this gives the DJL

**Corollary 5.** *For all $N$ vectors $(x_1, x_2, \cdots, x_N) \in \mathbb{R}^d$ in d-dimension, there exists a random $\phi$ from DJL such that with $k = O(\frac{\log(N)}{\epsilon^2})$, for all $i \neq j; i, j \in [N]$ :*

$$P\left[ ||\phi(x_i) - \phi(x_j)|| \in (1 \pm \epsilon) ||x_i - x_j|| \right] \geq 1 - \frac{1}{N}$$

*Proof.* Pick $k = c \cdot \frac{\log(N)}{\epsilon^2}$, DJL states:

$$\forall\, x, y \quad P\left[||\phi(x) - \phi(y)|| \in (1 \pm \epsilon)||x - y||\right] \,\geq\, 1 - e^{\frac{-\epsilon^2 k}{9}} \,\geq\, 1 - \frac{1}{N^3}$$

by union bound:

$$P\left[\forall\, i, j : \, ||\phi(x) - \phi(y)|| \in (1 \pm \epsilon)||x - y|| \text{ for } x = x_i, y = x_j\right] \,\geq\, 1 - \binom{N}{2} \cdot \frac{1}{N^3} \,\geq\, 1 - \frac{1}{N}$$

□

For $k \times n$ matrix $\mathbb{G}$ and vector $\mathbf{x}$, where each entry in $\mathbb{G}$ is a Gaussian:

$$\phi(x) = \frac{1}{\sqrt{k}} \cdot \mathbb{G} \cdot \mathbf{x}$$

with $1 \pm \epsilon$ approximation,

$$\phi \,:\, l_2^d \to l_2^k$$

where

$$l_2^d = ||x - y||_2 = \sum_{j=1}^{d} (x_i \cdots y_i)$$

What about $l_1$?

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

$$l_p^d \,:\, \mathbb{R}^d \text{ where } ||x - y||_1 = \left(\sum_{i=1}^{n} |x_i - y_i|^p\right)^{\frac{1}{p}}$$

For $l_1$: $N$ vectors into lower dimensional $l_1$

$$K = N^{\Omega(\frac{1}{D})} \text{ for D-approximation}$$

**Alternative Sketch**:

$$\phi(x) = \frac{1}{k} \cdot \mathbb{C} \cdot \mathbf{x}$$

where $\mathbb{C}$ is a matrix with Cauchy distribution. So given $\phi(x)$, $\phi(y)$ we can estimate $||x-y||$ as the median $(||\phi(x) - \phi(y)||$ of the absolute values of the $k$ coordinates.

It's enough to take

$$k = O(\frac{\log(N)}{\epsilon^2})$$

Cauchy variables are the 1-stable distribution: $\sum c_i x_i$, where $c_i$ are random Cauchy, is distributed as $\|x\|_1 \cdot c$ where $c$ is also Cauchy. In general, for $p \in (0, 2]$, there exist $p$-stable distributions satisfying the above with $\|x\|_1$ replaced by $\|x\|_p$.