# Lecture 6 9/11/2025

## Johnson-Lindenstrauss Lemma and Dimensionality Reduction

A fundamental yet simple result from convex geometry that has found many applications in data analysis and algorithms.

Let $\bar{v}_1, \bar{v}_2, \ldots, \bar{v}_n$ be $n$ vectors/points in $R^d$ where $d$ is say large.
The lemma says that one can project these vectors to a lower dimensional space $R^k$ to create

new vectors $\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_n$ such that $\forall\, i,j \in [n]$

$$(1-\varepsilon)\|v_i - v_j\|_2 \leq \|u_i - u_j\|_2 \leq (1+\varepsilon)\|v_i - v_j\|_2$$

ie the pairwise Euclidean distance is approximately preserved. And

$$k = O\left(\frac{\log n}{\varepsilon^2}\right).$$

Thus $d$ does not show up.

Clearly this is useful only if $d$ was originally larger than $k$.

Further the projection is via a linear map $A \in \mathbb{R}^{k \times n}$ that is randomized and <u>oblivious</u> to the data.

The core of the result is about a single vector and we then use a union bound.

## Distributional JL Lemma

Fix a vector $\bar{x} \in \mathbb{R}^d$. Let $\Pi$ be a $k \times d$ matrix where $\Pi_{i,j}$ is chosen independently ~~p~~ from a stand Gaussian distribution $N(0,1)$.

If $k = \Omega\left(\frac{1}{\varepsilon} \log \frac{1}{\delta}\right)$ then with probability $\geq 1-\delta$

$$(1-\varepsilon)\|\bar{x}\|_2 \leq \left\| \frac{1}{\sqrt{k}} \Pi \bar{x} \right\|_2 \leq (1+\varepsilon)\|\bar{x}\|_2 .$$

Assuming the DJL the dimensionality reduction is easy.

Choose $K = \Omega\left(\frac{1}{\varepsilon^2} \log n\right)$.

Then $\forall i, j$ look at vector $\bar{x} = \bar{v}_i - \bar{v}_j$

With probability $1 - \frac{1}{n^3}$

$$\frac{1}{\sqrt{K}} \| \Pi \bar{v}_i - \Pi \bar{v}_j \|_2 \text{ is within}$$

$$(1 \pm \varepsilon) \text{ of } \| \bar{v}_i - \bar{v}_j \|.$$

So by union bound all pairs are preserved with prob $1 - \frac{1}{n}$.

So $\bar{u}_i = \frac{1}{\sqrt{K}} \Pi \bar{v}_i$.

Oblivious since $\Pi$ was chosen without

Considering the data $\bar{V}_1, \bar{V}_2, \ldots, \bar{V}_n$.

$\Rightarrow$ can be used for "future" data.

Proof of the DJL Lemma.

Relies on some nice properties of the Gaussian distribution. One can use other distributions and there are some advantages to do so but analysis is more involved.

Sum of Independent Normally distributed Random Variables

Recall $N(\mu, \sigma^2)$ is normal with mean $\mu$ and variance $\sigma^2$: $f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$.

<u>Lemma</u>: Suppose $Z_1 \sim N(\mu_1, \sigma_1^2)$ and $Z_2 \sim N(\mu_2, \sigma_2^2)$ and $Z_1, Z_2$ are independent. Then.

$$Z_1 + Z_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

<u>Corollary</u>: Suppose $\bar{Z} = (Z_1, Z_2, \ldots, Z_d)$ is a a vector of independent $N(0,1)$ random variables. Let $\bar{X} \in \mathbb{R}^d$ and $\|X\|_2 = 1$ ie a unit vector.

Then $\sum_{i=1}^{d} x_i Z_i$ is distributed as $N(0,1)$.

<u>Proof</u>: ~~Th~~ Variance is $\sum_{i=1}^{d} x_i^2 = 1$ since $\bar{X}$ is a unit vector. $\square$

Proof of Lemma:

We will consider the moment generating function of $N(\mu, \sigma^2)$.

Suppose $X \sim N(\mu, \sigma^2)$.

What is $E[e^{tX}]$ ?

$$= \int_{-\infty}^{\infty} e^{tx} \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}} \cdot dx$$

$$= \int_{-\infty}^{\infty} \frac{e^{-\frac{1}{2\sigma^2}\left[(x-(\mu+\sigma^2 t))^2 - 2\mu\sigma^2 t - \sigma^4 t^2\right]}}{\sqrt{2\pi\sigma^2}} dx$$

$$= e^{\mu t + \frac{\sigma^2 t^2}{2}}$$

Thus MGF is $e^{\mu t + \frac{\sigma^2 t^2}{2}}$

Suppose we have $k$ independent

random variables $X_1, X_2, \dots, X_k$
with distributions $\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_k$.
If $M_1(t), M_2(t), \dots, M_k(t)$ are
the moment generating functions of
$X_1, X_2, \dots, X_k$ then the
moment generating function of
$X = \sum X_i$ is

$$M(t) = M_1(t) M_2(t) \dots M_k(t)$$

because

$$E[e^{tX}] = \prod_{i=1}^{k} E[e^{tX_i}]$$

If $X_1, X_2, \dots, X_k$ are $N(\mu_1, \sigma_1^2)$
$\dots N(\mu_k, \sigma_k^2)$ then

$M(t) = e^{\sum_i \mu_i t_i + \frac{1}{2}\sum_i \sigma_i^2}$ which is
the MGF a $N(\sum \mu_i, \sum \sigma_i^2)$.

Let $\Pi$ be a $k \times d$ Gaussian matrix and let $\bar{x}$ be a unit vector wlog.

Let $\bar{y} = \Pi \bar{x}$ $\qquad$ $\bar{y} = (Y_1, Y_2, \ldots, Y_k)$

~~Ea~~ where $Y_i \sim N(0,1)$.

and $Y_1, Y_2, \ldots, Y_k$ are independent.

What is $\|\bar{y}\|_2^2$ ?

$$= \sum_{i=1}^{k} Y_i^2$$

Hence $E[\|y\|_2^2] = \sum_{i=1}^{k} E[Y_i^2] = k.$

Thus $\frac{1}{\sqrt{k}} E[\|\bar{y}\|_2] = 1.$

Our goal is to prove that

$$Y = \sum_{i=1}^{k} Y_i^2 \quad \text{is concentrated.}$$

Note that $Y_i \sim N(0,1)$ and hence $Y_i^2$ is the square.

The distribution of the square of a Normal distribution is called the $\chi^2$-distribution and is important in statistics.

The distribution of the sum of squares of $t$ independent $N(0,1)$ random variables is called the $\chi^2$-distribution with parameter $t$. $\chi^2(t)$.

It is known that the $\chi^2$-distribution with parameter $t$ has exponential tails.

<u>Lemma</u>: Let $Y_1, Y_2, \ldots, Y_k$ be independent $N(0,1)$ random variables and let $Y = \sum_{i=1}^{k} Y_i^2$.

Then for $\varepsilon \in (0, \frac{1}{2})$

$$\Pr\left[(1-\varepsilon)^2 k \le Y \le (1+\varepsilon)^2 k\right] \geq 1 - 2e^{-c\varepsilon^2 k}$$

where $c$ is an absolute constant.

The preceding concentration lemma implies the DJL lemma

By choosing $k = \frac{c'}{\varepsilon^2} \ln \frac{1}{\delta}$

$$\Pr\left[(1-\varepsilon) \le \left\| \frac{1}{\sqrt{k}} \Pi X \right\|_2 \le (1+\varepsilon)\right] \geq 1-\delta .$$

---

Proof of the concentration lemma

We will follow the exponential moment method. For $t \geq 0$

$$\Pr[Y \geq \alpha] = \Pr[e^{tY} > e^{t\alpha}]$$

$$\leq \frac{E[e^{tY}]}{e^{t\alpha}} .$$

The difficult part is understanding

$$\mathbb{E}[e^{tY}].$$

It turns out that for $\chi^2(k)$ there is a nice closed form for $t \in [0, \frac{1}{2})$

$$\mathbb{E}[e^{tY}] = (1-2t)^{-k/2}.$$

Not defined for $t \geq \frac{1}{2}$.

Then we can plug it in

$$\Pr[X > \alpha] \leq (1-2t)^{-k/2} \cdot e^{-t\alpha}$$

as long as $t \in [0, \frac{1}{2})$.

Choose $t = \left(1 - \frac{k}{\alpha}\right)\frac{1}{2}$.

$$\Pr[X > \alpha] \le \left(\frac{k}{\alpha}\right)^{-\frac{k}{2}} e^{-\frac{(\alpha - k)}{2}}$$

For $\alpha = (1+\varepsilon)^2 k$.

$$\le (1+\varepsilon)^k \cdot e^{-\frac{k[(1+\varepsilon)^2 - 1]}{2}}$$

$$\le e^{-k\left[\frac{(1+\varepsilon)^2 - 1}{2} - \ln(1+\varepsilon)\right]}$$

$$\ln(1+\varepsilon) = \varepsilon - \frac{\varepsilon^2}{2} + \frac{\varepsilon^3}{3!} \cdots \le \varepsilon$$

$$\le e^{-k\left[\frac{\varepsilon^2}{2} + \varepsilon - \varepsilon\right]}$$

$$\le e^{-\frac{k\varepsilon^2}{2}}$$

Lower tail is similar. $\square$

How do we get the closed form for $e^{tY}$?

Recall $Y = \sum_{i=1}^{k} Y_i^2$

$Y_i$ are independent and each $Y_i \sim N(0,1)$

Hence MGF of $Y_i^2$

$$= E\left[e^{tY_i^2}\right]$$

$$-\frac{y^2}{2}(1-2t)$$

$$= \int_{-\infty}^{\infty} e^{ty^2} \cdot \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}} \, dy$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2 \cdot \frac{1}{(1-2t)}}} \cdot dy$$

$$= \frac{1}{\sqrt{1-2t}} \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi \frac{1}{\sqrt{1-2t}}}} \cdot e^{-\frac{y^2}{2\frac{1}{1-2t}}} dy$$

$$= \frac{1}{\sqrt{1-2t}} = (1-2t)^{-\frac{1}{2}}.$$

Since $Y = \sum_{i=1}^{k} Y_i^2$ and independent identical

$$MGF(Y) = (1-2t)^{-\frac{k}{2}}$$

D.

New Topic

Hashing

Requires some background in pseudo randomness / derandomization.

How do we convert randomized algorithms into deterministic algorithms?

Is it always possible?

We will discuss some of these later but our goal today is to introduce the notion of limited independence.

<u>Definition</u>: A set of random variables $X_1, X_2, \ldots, X_n$ are <u>pairwise independent</u> or 2-wise independent if $X_i$ and $X_j$ are independent for any $i \neq j$.

<u>Ex</u>: $X_1, X_2, X_3 \in \{0,1\}$

$X_1, X_2$ independent with prob $\frac{1}{2}$ of $0, 1$.

$X_3 = X_1 \oplus X_2$.

$X_1, X_2, X_3$ are pairwise indep but not independent.

- Pairwise independence suffices in some applications
- Can generate many pairwise indep random bits from a small # of independent random bits.

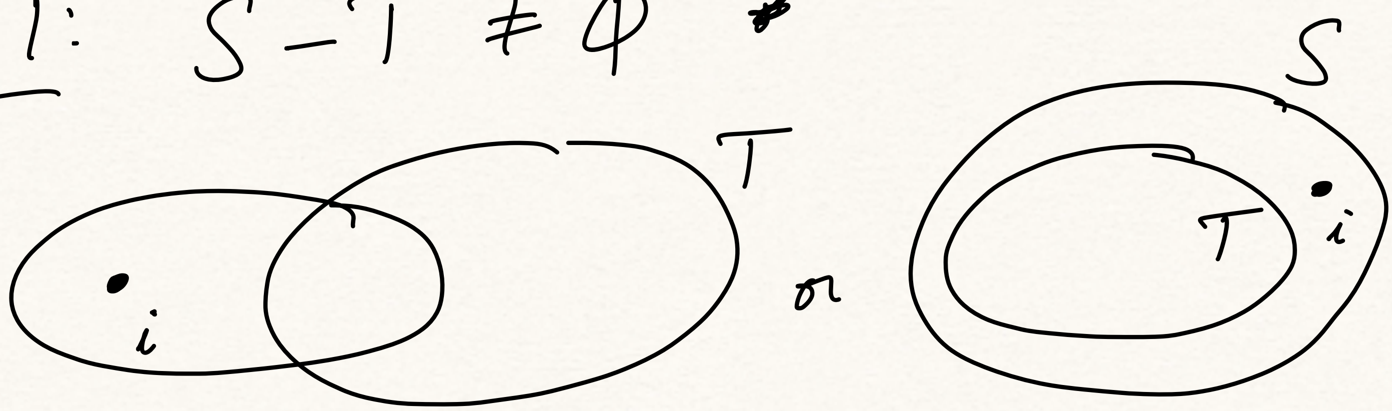
Lemma: Can construct $n = 2^k - 1$ pairwise independent random variables $X_1, X_2, \ldots, X_n$ where each $X_i \in \{0,1\}$ from $k$ random bits $Y_1, Y_2, \ldots, Y_k$.

Construction: Let $S$ be a non-empty subset of $\{1, 2, \ldots, k\}$.

define $X_S = \bigoplus_{i \in S} Y_i$

If $S \neq T$ $X_S$ and $X_T$ are independent.

Case 1: $S - T \neq \phi$




Say $i \in S - T$.

For any choice of bits in $T$, the value of $X_S$ is equally likely to be 0 or 1.

Case 2: $T - S \neq \phi$ same as above.

If $S \neq T$ either $S - T \neq 0$ or $T - S \neq 0$.

D.

Application:

Derandomizing Max-Cut alg.

Recall $G=(V,E)$ want to find Max-Cut

1. $S \leftarrow \emptyset$
2. For each $v \in V$
   add $v$ to $S$ with prob $\frac{1}{2}$.

Recall analysis

$X_v$ indicator for $v \in S$.

$Y_e$ indicator for edge $e$ being cut
i.e. $e \in \delta(S)$.

$$\Pr[Y_e = 1] = \frac{1}{2}.$$

For edge $e = uv$

This only requires $X_u$ and $X_v$ to be independent.

$Y = \sum_e Y_e$ and we used linearity of expectation to claim

$E[Y] = \frac{1}{2} m$.

So if $X_u$ $u \in V$ are pairwise independent

$E[Y] = \frac{1}{2} m$ still holds.

If we have $n$ vertices we can use above construction to generate $n$ pairwise random bits from $\lceil \log n + 1 \rceil$ true random bits.

# of bit strings of length $\lceil \log n + 1 \rceil$
$$= O(n).$$

Try each such bit string, generate $n$ pairwise random bits and run alg on each of these and take the best.

A deterministic algorithm!