CSIS8601: Probabilistic Method & Randomized Algorithms
Lecture 5: Johnson-Lindenstrauss Lemma: Dimension Reduction
Date: 30 Sept 2009

These lecture notes are supplementary materials for the lectures. They are by no means substitutes for attending lectures or replacement for your own notes!

1 Dimension Reduction in Euclidean Space

Consider n vectors in Euclidean space of some large dimension. These n vectors reside in an n dimensional subspace. By rotation, we can assume that n vectors lie in $\mathbb{R}^n$. On the other hand, it is easy to see that n mutually orthogonal unit vectors cannot reside in a space with dimension less than n.

Moreover, it is not possible to have three mutually almost orthogonal vectors placed in 2 dimensions.

Definition 1.1 *We say two unit vectors u and v are ϵ-orthogonal to one another if their dot product satisfies $|u \cdot v| \leq \epsilon$.*

One might think that n mutually almost orthogonal vectors require n dimensions. Hence, it might come as a surprise that n vectors that are mutually ϵ-orthogonal can be placed in a Euclidean space with $O(\frac{\log n}{\epsilon^2})$ dimensions.

Observer that for any three points, if the three distances between them are given, then the three angles are fixed. Given $n-1$ vectors, the vectors together with the origin form a set of n points. In fact, given any n points in Euclidean space (in $n-1$ dimensions), the Johnson-Lindenstrauss Lemma states that the n points can be placed in $O(\frac{\log n}{\epsilon^2})$ dimensions such that distances are preserved with multiplicative error ϵ, for any $0 < \epsilon < 1$.

Theorem 1.2 (Johnson-Lindenstrauss Lemma) *Suppose U is a set of n points in Euclidean space $\mathbb{R}^n$. Then, for any $0 < \epsilon < 1$, there is a mapping $f : U \to \mathbb{R}^T$, where $T = O(\frac{\log n}{\epsilon^2})$, such that for all $x, y \in U$,*

$(1-\epsilon)||x-y||^2 < ||f(x) - f(y)||^2 < (1+\epsilon)||x-y||^2.$

Remark 1.3
1. Since for small ϵ, $(1+\epsilon)^2 = 1 + \Theta(\epsilon)$ and $(1-\epsilon)^2 = 1 - \Theta(\epsilon)$, it follows that the squared of the distances are preserved *iff* the distances themselves are.
2. Note that $||x-y||$ is a norm between 2 vectors in Euclidean space $\mathbb{R}^n$ and $||f(x) - f(y)||$ is one between 2 vectors in $\mathbb{R}^T$. Be careful that, $||x - f(x)||$ is not well-defined.

Corollary 1.4 (Almost Orthogonal Vectors) *Suppose $u_1, u_2, \ldots, u_n$ are mutually orthogonal unit vectors in $\mathbb{R}^n$. Then, for any $0 < \epsilon < 1$, there exists a mapping $f : U \to \mathbb{R}^T$, where $T = O(\frac{\log n}{\epsilon^2})$ such that $|\frac{f(u_i)}{||f(u_i)||} \cdot \frac{f(u_j)}{||f(u_j)||}| \leq \epsilon$.*

Proof: We apply Johnson-Lindenstrauss' Lemma with error $\frac{\epsilon}{8}$ to the set U of vectors $u_1, u_2, \ldots, u_n$ together with the origin to obtain $f : U \to \mathbb{R}^T$, where $T = O(\frac{\log n}{\epsilon^2})$.

Hence, it follows that for all i, $1 - \frac{\epsilon}{8} \leq ||f(u_i)||^2 \leq 1 + \frac{\epsilon}{8}$.

Moreover, for $i \neq j$, $(1-\frac{\epsilon}{8})||u_i - u_j||^2 < ||f(u_i) - f(u_j)||^2 < (1+\frac{\epsilon}{8})||u_i - u_j||^2$.

Observe that $||u_i - u_j||^2 = 2$ and $||f(u_i) - f(u_j)||^2 = ||f(u_i)||^2 + ||f(u_j)||^2 - 2f(u_i) \cdot f(u_j)$.

So, from $(1-\frac{\epsilon}{8})||u_i - u_j||^2 < ||f(u_i) - f(u_j)||^2$, we conclude $f(u_i) \cdot f(u_j) \leq \frac{\epsilon}{4}$.

On the other hand, from $||f(u_i) - f(u_j)||^2 < (1+\frac{\epsilon}{8})||u_i - u_j||^2$, we have $f(u_i) \cdot f(u_j) \geq -\frac{\epsilon}{4}$.

Hence, we have $|f(u_i) \cdot f(u_j)| \leq \frac{\epsilon}{4}$. However, observe that $f(u_i)$ and $f(u_j)$ might not be unit vectors. We know that $||f(u_i)|| \cdot ||f(u_j)|| \geq (1-\frac{\epsilon}{8})^2 \geq \frac{1}{4}$. Therefore, we have $|\frac{f(u_i)}{||f(u_i)||} \cdot \frac{f(u_j)}{||f(u_j)||}| \leq \epsilon$. ■

2 Random Projection

For point x, suppose $f(x) := (f_i(x))_{i \in [T]}$. Then, $||f(x) - f(y)||^2 = \sum_{i \in [T]} |f_i(x) - f_i(y)|^2$.

We have learned that the sum of independent random variables concentrate around its mean. Hence, the goal is to design a random mapping $f_i : U \to \mathbb{R}$ such that $E[|f_i(x) - f_i(y)|^2] = \frac{1}{T} \cdot ||x - y||^2$, in which case we have $E[||f(x) - f(y)||^2] = ||x - y||^2$.

Note that f_i takes a vector and returns a number. Observe that Euclidean space is equipped with dot product. Note that dot product with a unit vector gives the magnitude of the projection on the unit vector. Hence, we can take a random vector r in space $\mathbb{R}^n$, and let f_i have the form $f_i(x) := r \cdot x$.

Suppose we fix two points x and y. Since dot product is linear, we have $f_i(x) - f_i(y) = f_i(x - y)$. Hence, we consider $v := x - y = (v_0, v_1, \ldots, v_{n-1})$, and let $\nu := ||v|| = \sqrt{\sum_i v_i^2}$. Recall the goal is to define f_i, and hence find a random vector r such that $E[(r \cdot v)^2] = \frac{1}{T} \cdot ||v||^2 = \frac{\nu^2}{T}$.

Using Random Bits to Define a Random Projection For each $j \in [n]$, suppose $\gamma_j \in \{-1, +1\}$ is a uniform random bit such that γ's are independent. Define the random vector $r := \frac{1}{\sqrt{T}}(\gamma_0, \gamma_1, \ldots, \gamma_{n-1})$. Hence, $f_i(v) = \frac{1}{\sqrt{T}} \sum_j \gamma_j v_j$.

Check that $E[(f_i(v))^2] = \frac{1}{T} \sum_j v_j^2 = \frac{\nu^2}{T}$. Hence, we have found the required random mapping $f_i : \mathbb{R}^n \to \mathbb{R}^T$.

Remark 2.1 Observe that the mapping $f : \mathbb{R}^n \to \mathbb{R}^T$ is linear.

3 Proof of Johnson-Lindenstrauss Lemma

We define $X_i := f_i(v)^2 = \frac{1}{T}(\sum_j \gamma_j v_j)^2$, and let $Y := \sum_i X_i$. Recall $E[X_i] = \frac{\nu^2}{T}$ and $E[Y] = \nu^2$. Then, the desirable event can be expressed as:

$Pr[(1-\epsilon)||x - y||^2 < ||f(x) - f(y)||^2 < (1+\epsilon)||x - y||^2] = Pr[|Y - E[Y]| < \epsilon E[Y]]$.

The goal is to first find a T large enough such that the failing probability $Pr[|Y - E[Y]| \geq \epsilon E[Y]]$ is at most $\frac{1}{n^2}$. Since there are $\binom{n}{2}$ such pairs of points, using union bound, we can show that with probability at least $\frac{1}{2}$, the distances of all pairs of points are preserved.

We again use the method of moment generating function.

3.1 JL as a Measure Concentration Result

Using the method of moment generating function described in previous classes, the failure probability in question is at most the sum of the following two probabilities.

1. $Pr[Y \leq (1-\epsilon)\nu^2] \leq \exp(-t(1-\epsilon)\nu^2) \cdot \prod_i E[\exp(tX_i)]$, for all $t < 0$.
2. $Pr[Y \geq (1+\epsilon)\nu^2] \leq \exp(-t(1+\epsilon)\nu^2) \cdot \prod_i E[\exp(tX_i)]$, for all $t > 0$.

We next derive an upper bound for $E[e^{tX_i}]$.

4 Upper Bound for $E[e^{tX_i}]$

For notational convenience, we drop the subscript i, and write $X := \frac{1}{T}(\sum_j \gamma_j v_j)^2$, where $\nu^2 = \sum_j v_j^2$, where $\gamma_j \in \{-1, 1\}$ are uniform and independent. Hence, we have

$E[e^{tX}] = E[\exp(\frac{t}{T}(\sum_j v_j^2 + \sum_{i \neq j} \gamma_i \gamma_j v_i v_j))]$.

Although the γ_j's are independent, the cross-terms $\gamma_i\gamma_j$'s are not. In particular, $\gamma_i\gamma_j$ and $\gamma_{i'}\gamma_{j'}$ are not independent if $i = i'$ or $j = j'$.

We compare X with another variable $\widehat{X}$, which we can analyze.

4.1 Normal Distribution

Suppose g is a random variable having standard normal distribution $N(0,1)$, with mean 0 and variance 1. In particular, it has the following probability density function:

$\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, for $x \in \mathbb{R}$.

Suppose γ is a $\{-1, 1\}$ is a random variable that takes value -1 or 1, each with probability $\frac{1}{2}$. Then, the random variables g and γ have some common properties.

Fact 4.1 *Suppose γ is a uniform $\{-1,1\}$-random variable and g is a random variable with normal distribution $N(0,1)$.*

1. $E[\gamma] = E[g] = 0$.
2. $E[\gamma^2] = E[g^2] = 1$.

For higher moments we have,

1. *For odd $n \geq 3$, $E[\gamma^n] = E[g^n] = 0$.*
2. *For even $n \geq 4$, $1 = E[\gamma^n] \leq E[g^n]$.*

Normal distributions have the following important property.

Fact 4.2 *Suppose g_i's are independent random variables, each having standard normal distribution $N(0,1)$. Define $Z := \sum_j g_j v_j$, where v_j's are real numbers. Then, Z has normal distribution $N(0,\nu^2)$ with mean 0 and variance $\nu^2 := \sum_i v_i^2$.*

We define $\widehat{X} := \frac{1}{T}(\sum_j g_j v_j)^2$ and let $Z := \sum_j g_j v_j$. Notice that we have $Z \sim N(0,\nu^2)$.

Using Fact 4.1, we can compare the moments of X and $\widehat{X}$.

Lemma 4.3 *Define X and $\widehat{X}$ as above.*

1. *For all integers $n \geq 0$, $E[X^n] \leq E[\widehat{X}^n]$.*
2. *Using the Taylor expansion $\exp(y) := \sum_{i=0}^{\infty} \frac{y^i}{i!}$, we have $E[\exp(tX)] \leq E[\exp(t\widehat{X})]$, for $t > 0$.*

Lemma 4.4 *For $t < \frac{T}{2\nu^2}$, $E[\exp(t\widehat{X})] \leq (1 - \frac{2t\nu^2}{T})^{-\frac{1}{2}}$.*

Sketch Proof: Observe that $\widehat{X} = \frac{1}{T}Z^2$, where Z has normal distribution $N(0,\nu^2)$.

Hence, it follows that $E[e^{t\widehat{X}}] = E[\exp(\frac{t}{T} \cdot Z^2)]$. We leave the rest of the calculation as a homework exercise.

Therefore, for $t > 0$, we conclude that $E[\exp(tX)] \leq E[\exp(t\widehat{X})] \leq (1 - \frac{2t\nu^2}{T})^{-\frac{1}{2}}$, for $t < \frac{T}{2\nu^2}$.

Claim 4.5 *Suppose $X := \frac{1}{T}(\sum_j \gamma_j v_j)^2$, where $\nu^2 = \sum_j v_j^2$.*

Then, for $0 < t < \frac{T}{2\nu^2}$, $E[\exp(tX)] \leq (1 - \frac{2t\nu^2}{T})^{-\frac{1}{2}}$.

For negative t, we cannot argue that $E[\exp(tX)] \leq E[\exp(t\widehat{X})]$. However, we can still obtain an upper bound using another method.

Claim 4.6 *For $t < 0$, $E[\exp(tX)] \leq 1 + \frac{t\nu^2}{T} + \frac{3}{2} \cdot (\frac{t\nu^2}{T})^2$.*

Proof:

We use the inequality: for $y < 0$, $e^y \leq 1 + y + \frac{y^2}{2}$.

Hence, for $t < 0$,

$E[\exp(tX)] \leq E[1 + tX + \frac{t^2}{2}X^2] = 1 + \frac{t\nu^2}{T} + \frac{t^2}{2}E[X^2]$.

We use the fact that $E[X] = \frac{\nu^2}{T}$. We next obtain an upper bound for $E[X^2]$. From Lemma 4.3, we have $E[X^2] \leq E[\widehat{X}^2]$.

Observe that $\widehat{X}^2 = \frac{Z^4}{T^2}$, where Z has the normal distribution $N(0,\nu^2)$. Hence, $E[\widehat{X}^2] = \frac{\nu^4}{T^2}E[g^4]$, where g has the standard normal distribution $N(0,1)$.

Through a standard calculation, we have $E[g^4] = 3$, hence achieving the required bound. ■

4.2 Finding the right value for t.

We now have an upper bound for $E[e^{tX_i}]$ and hence we can finish the proof.

Positive t. For $t > 0$, we have $Pr[Y \geq (1+\epsilon)\nu^2] \leq \exp(-t(1+\epsilon)\nu^2) \cdot \prod_i E[\exp(tX_i)]$

$\leq \exp(-t(1+\epsilon)\nu^2) \cdot (1 - \frac{2t\nu^2}{T})^{-\frac{T}{2}}$, where t has to satisfy $t < \frac{T}{2\nu^2}$ too.

Remark 4.7 In this case, the upper bound is not of the form $E[\exp(tX_i)] \leq \exp(g_i(t))$. Instead of trying to find the best value of t by calculus, sometimes another valid value of t is good enough. We try $t := \frac{T}{2\nu^2} \cdot \frac{\epsilon}{1+\epsilon}$. In this case, we have $(1 - \frac{2t\nu^2}{T})^{-\frac{1}{2}} \leq \sqrt{1+\epsilon}$. Hence,

$Pr[Y \geq (1+\epsilon)\nu^2] \leq (\sqrt{e^{-\epsilon}(1+\epsilon)})^T \leq \exp(-\frac{\epsilon^2 T}{12})$,

where the last inequality comes from the fact that for $0 < \epsilon < 1$,

$\sqrt{e^{-\epsilon}(1+\epsilon)} = \exp(\frac{1}{2}(-\epsilon + \ln(1+\epsilon))) \leq \exp(-\frac{\epsilon^2}{12})$.

Negative t. For negative t, we use the bound $E[e^{tX}] \leq 1 + \frac{t\nu^2}{T} + \frac{3}{2} \cdot (\frac{t\nu^2}{T})^2$.

We can pick any negative t. So, we try $t := -\frac{\epsilon}{2(1+\epsilon)} \cdot \frac{T}{\nu^2}$.

$Pr[Y \leq (1-\epsilon)\nu^2] \leq [(1 - \frac{\epsilon}{2(1+\epsilon)} + \frac{3\epsilon^2}{8(1+\epsilon)^2}) \exp(\frac{\epsilon(1-\epsilon)}{2(1+\epsilon)})]^T$.

We apply the inequality $1 + x \leq e^x$, for any real x to obtain the following upper bound.

$[\exp(-\frac{\epsilon}{2(1+\epsilon)} + \frac{3\epsilon^2}{8(1+\epsilon)^2} + \frac{\epsilon(1-\epsilon)}{2(1+\epsilon)})]^T \leq \exp(-\frac{\epsilon^2 T}{12})$.

One can check that $-\frac{\epsilon}{2(1+\epsilon)} + \frac{3\epsilon^2}{8(1+\epsilon)^2} + \frac{\epsilon(1-\epsilon)}{2(1+\epsilon)} \leq -\frac{\epsilon^2}{12}$, for $0 < \epsilon < 1$.

Hence, in conclusion, for $0 < \epsilon < 1$,

$Pr[|Y - \nu^2| \geq \epsilon\nu^2] \leq 2\exp(-\frac{\epsilon^2 T}{12})$. This probability is at most $\frac{1}{n^2}$, if we choose $T := \left\lceil \frac{12 \ln 2n^2}{\epsilon^2} \right\rceil$.

5 Lower Bound

We show that if we want to maintain the distances of n points in Euclidean space, in some cases, the number of dimension must be at least $\Omega(\log n)$.

5.1 Simple Volume Argument

Consider a set $V = \{u_1, u_2, \ldots, u_n\}$ of n points in n-dimensional Euclidean space. For instance, let $u_i := \frac{e_i}{\sqrt{2}}$, where e_i is the standard unit vector, where the ith position is 1 and 0 elsewhere. Then, for $i \neq j$, $||u_i - u_j|| = 1$.

We show the following result.

Theorem 5.1 *Let $0 < \epsilon < 1$. Suppose $f : V \to \mathbb{R}^T$ such that for all $i \neq j$,*

$1 \leq ||f(u_i) - f(u_j)|| \leq 1 + \epsilon$.

Then, T is at least $\Omega(\log n)$.

Remark 5.2 Observe that if we have $1 - \epsilon \leq ||f(u_i) - f(u_j)|| \leq 1 + \epsilon$, then we can divide the mapping by $(1-\epsilon)$, i.e. $f' := \frac{f}{1-\epsilon}$. Then, we have $1 \leq ||f'(u_i) - f'(u_j)|| \leq \frac{1+\epsilon}{1-\epsilon} = 1 + \Theta(\epsilon)$.

Proof:

For each i, consider a ball $B(f(u_i), \frac{1}{2})$ of radius $\frac{1}{2}$ around the center $f(u_i)$. Since for $i \neq j$, $||f(u_i) - f(u_j)|| \geq 1$, the balls are disjoint (except maybe for only 1 point of contact between two balls).

On the other hand, for all $i > 1$, $||f(u_1) - f(u_i)|| \leq (1 + \epsilon)$. Hence, it follows the big ball $B(f(u_1), \frac{3}{2} + \epsilon)$ centered at $f(u_1)$ contains all the n smaller balls.

Note that the volume of a ball with radius r in $\mathbb{R}^T$ is proportional to r^T. Since there are n disjoint smaller balls in the big ball, the ratio of the volume of the big ball to that of a smaller ball is at least n.

Hence, we have $n \leq \frac{(\frac{3}{2}+\epsilon)^T}{(\frac{1}{2})^T} \leq 5^T$, for $\epsilon < 1$. Therefore, it follows that $T \geq \Omega(\log n)$. ■

6 Homework Preview

1. Suppose g is a random variable with normal distribution $N(0, 1)$. Prove the following.

   (a) For odd $n \geq 1$, $E[g^n] = 0$.

   (b) For even $n \geq 2$, $E[g^n] \geq 1$.

   (Hint: Use induction. Let $I_n := E[g^n] = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x^n e^{-\frac{x^2}{2}} dx$. Use integration by parts to show that $I_{n+2} = (n + 1)I_n$.)

2. Suppose γ_j's are independent uniform $\{-1, 1\}$-random variables and g_j's are independent random variables, each having normal distribution $N(0, 1)$. Suppose v_j's are real numbers, and define $X := (\sum_j \gamma_j v_j)^2$ and $\widehat{X} := (\sum_j g_j v_j)^2$. Show that for all integers $n \geq 1$, $E[X^n] \leq E[\widehat{X}^n]$.

3. Suppose Z is a random variable having normal distribution $N(0, \nu^2)$. Compute $E[e^{tZ^2}]$. For what values of t is your expression valid?

4. In this question, we investigate if Johnson-Lindenstrauss Lemma can preserve area.

   (a) Suppose the distances between three points are preserved with multiplicative error ϵ. Is the area of the corresponding triangle also always preserved with multiplicative error $O(\epsilon)$, or even some constant multiplicative error?

   (b) Suppose u and v are mutually orthogonal unit vectors. Observe that the vectors u and v together with the origin form a right-angled isosceles triangle with area $\frac{1}{2}$. Suppose the lengths of the triangle are distorted with multiplicative error at most ϵ. What is the multiplicative error for the area of the triangle?

   (c) Suppose a set V of n points are given in Euclidean space $\mathbb{R}^n$. Let $0 < \epsilon < 1$. Give a randomized algorithm that produces a low-dimensional mapping $f : V \to \mathbb{R}^T$ such that the areas of all triangles formed from the n points are preserved with multiplicative error ϵ. What is the value of T for your mapping? Please give the exact number and do not use big O notation.
   (Hint: If two triangles lie in the same plane (a 2-dimensional affine space) in $\mathbb{R}^n$, then under a linear mapping their areas have the same multiplicative error. For every triangle, add an extra point to form a right-angled isosceles triangle in the same plane.)