CS 388R: Randomized Algorithms Fall 2017

Lecture 15 — October 24, 2017

Prof. Eric Price *Scribe: Naren Manoj, Ryan Rice*

NOTE: THESE NOTES HAVE NOT BEEN EDITED OR CHECKED FOR CORRECTNESS

1 Overview

In this lecture we state the Johnson Lindenstrauss Lemma and the Distributional Johnson Lindenstrauss Lemma. We then develop some of the probability machinery required to prove these, such as Moment Generating Functions, Subgaussian variables, and the derivation of the additive Chernoff bound.

2 JL Lemma

Note: All norms are ℓ^2 norms.

Johnson Lindenstrauss Lemma For any set of n vectors $x_1, \ldots, x_n \in \mathbb{R}^d$ there exists an embedding $y_1, \ldots, y_n \in \mathbb{R}^m$ such that

$$\forall i, j \in [n], \quad \|y_i - y_j\|_2 \in (1 \pm \epsilon) \|x_i - x_j\|_2$$

As long as $m = O(\frac{1}{\epsilon^2} \log n)$.

Notice that we can get an exact embedding ($\epsilon = 0$) in n dimensions by projecting the vectors onto the subspace spanned by $x_1, \ldots, x_n$.

Distributional Johnson Lindenstrauss Lemma If we fix $0 < \epsilon, \delta < 1/2$, then there exists distributions on matrices $A \in \mathbb{R}^{m \times d}$ such that for any $x \in \mathbb{R}^d$, $\|Ax\|_2 \in (1 \pm \epsilon) \|x\|_2$ with probability $1 - \delta$ with $m = O(\frac{1}{\epsilon^2} \log \frac{1}{\delta})$.

The Johnson Lindenstrauss Lemma follows from this because

$$\|y_i - y_j\|_2 = \|Ax_i - Ax_j\|_2 = \|A(x_i - x_j)\|_2 \in (1 \pm \epsilon) \|x_i - x_j\|_2$$

3 Probability Background

Here, we'll develop some of the probability tools that are useful both in proving the JL Lemma and Chernoff bounds.

3.1 Concentration Inequalities

3.1.1 Markov's Inequality

Recall Markov's Inequality. For a nonnegative random variable X and positive t, we have:

$$\mathbb{P}[X \geq t] \leq \frac{\mathbb{E}[X]}{t}$$

3.1.2 Chebyshev's Inequality and kth Moments

We can use Markov's Inequality to derive Chebyshev's Inequality. Suppose X has mean 0. Then,

$$\begin{aligned} \mathbb{P}[|X| \geq t] &= \mathbb{P}[X^2 \geq t^2] \\ &\leq \frac{\mathbb{E}[X^2]}{t^2} \end{aligned}$$

Letting $X = Y - \mathbb{E}[Y]$ gives us the inequality we're familiar with.

Notice that this exponentiation trick works for even powers k:

$$\begin{aligned} \mathbb{P}[|X| \geq t] &= \mathbb{P}[X^k \geq t^k] \\ &\leq \frac{\mathbb{E}[X^k]}{t^k} \end{aligned}$$

Equivalently, we can write these in terms of a failure probability δ, where the failure is marked by some deviation from the mean. Concretely, we set $\mathbb{P}[|X| > t] = \delta$, which yields:

$$|x| \leq \left(\frac{\mathbb{E}[X^k]}{\delta}\right)^{1/k}$$

3.1.3 Moment Bounds on the Sums of Random Variables

Now, let's say that X is the sum of several independent random variables $X_i \in [-1, 1]$ where $\mathbb{E}[X_i] = 0$. In particular, set:

$$X = \sum_{i=1}^{n} X_i$$

We can try to come up with some tail bound for X in terms of the number of observations we have (n). Using our previous result, we can say:

$$\mathbb{P}[|X| \geq t] \leq \frac{\mathbb{E}[X^2]}{t^2}$$

Note that it suffices to determine an upper bound for $\mathbb{E}[X^2]$ in terms of n. Doing so will allow us to substitute this bound into the above inequality. To do so, consider the below:

$$\begin{aligned}
\mathbb{E}[X^2] &= \mathbb{E}\left[\left(\sum_{i=1}^{n} X_i\right)^2\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{n} X_i^2 + \sum_{i \neq j} X_i X_j\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{n} X_i^2\right] + \mathbb{E}\left[\sum_{i \neq j} X_i X_j\right] \\
&= \sum_{i=1}^{n} \mathbb{E}\left[X_i^2\right] + \sum_{i \neq j} \mathbb{E}\left[X_i X_j\right] \\
&= \sum_{i=1}^{n} \mathbb{E}\left[X_i^2\right] + \sum_{i \neq j} \mathbb{E}\left[X_i\right] \mathbb{E}\left[X_j\right] \\
&= \sum_{i=1}^{n} \mathbb{E}\left[X_i^2\right] \\
&\leq \sum_{i=1}^{n} 1 \\
&= n
\end{aligned}$$

Thus, we have:

$$\mathbb{P}[|X| \geq t] \leq \frac{\mathbb{E}[X^2]}{t^2} \leq \frac{n}{t^2}$$

In fact, we can generalize the above result to arbitrary even integers k:

$$\mathbb{E}\left[\left(\sum_{i=1}^{n} X_i\right)^k\right] \leq k^{k/2} n^{k/2} \tag{1}$$

Using **??**, we have:

$$\begin{aligned}
\mathbb{P}[|X| \geq y] &\leq \frac{\mathbb{E}\left[X^k\right]}{y^k} \\
&\leq \frac{k^{k/2} n^{k/2}}{y^k}
\end{aligned}$$

3.1.4 Obtaining a Chernoff-like Bound

Note that our previous inequality is true for all even k. As a result, we can select some k which makes the bound as tight as possible. To do so, set $y = t\sqrt{n}$ below:

$$\begin{aligned}\mathbb{P}[|X| \geq t\sqrt{n}] &\leq \frac{k^{k/2}n^{k/2}}{t^k n^{k/2}} \\ &\leq \left(\frac{\sqrt{k}}{t}\right)^k\end{aligned}$$

Now, we can use some calculus to determine the value of k which minimizes the right hand side. Once we do this, we find that $k = t^2/e$. Substituting this back gives us:

$$\begin{aligned}\mathbb{P}[|X| \geq t\sqrt{n}] &\leq \left(\frac{\sqrt{t^2/e}}{t}\right)^{t^2/e} \\ &= \left(e^{-1/2}\right)^{t^2/e} \\ &= \exp\left(-\frac{t^2}{2e}\right)\end{aligned}$$

Now, set $u = t/\sqrt{n}$. This yields:

$$\mathbb{P}[|X| \geq u\sqrt{n}] = \mathbb{P}[|X| \geq t] \leq \exp\left(-\frac{t^2}{2ne}\right) = \exp\left(-\Omega\left(\frac{t^2}{n}\right)\right)$$

which gives us something that looks like a Chernoff bound.

3.2 Moment Generating Functions and Subgaussian Variables

Observe that our previous analysis is somewhat faulty. Our moment bound is only known to hold true for even k, but in our computations, we set $k = t^2/e$. This is clearly not always an integer (and much less an even integer). We can construct a more sound analysis, but in order to do so, we have to introduce a new tool.

Definition 1 (Moment Generating Function)**.** *The Moment Generating Function (MGF) of a random variable X and parametrized by $\lambda \in \mathbb{R}$ is:*

$$M_X(t) = \mathbb{E}\left[e^{\lambda(X - \mathbb{E}[X])}\right]$$

To motivate the MGF, it helps to take a look at the below manipulations (for X being mean 0):

$$
\begin{aligned}
\mathbb{E}\left[e^{\lambda X}\right] &= \mathbb{E}\left[1 + \lambda X + \frac{\lambda^2 X^2}{2} + \frac{\lambda^3 X^3}{6} + \cdots + \frac{\lambda^r X^r}{r!} + \ldots\right] \\
&= \sum_{r=0}^{\infty} \mathbb{E}\left[\frac{\lambda^r X^r}{r!}\right] \\
&= \sum_{r=0}^{\infty} \frac{\lambda^r}{r!}\, \mathbb{E}\left[X^r\right]
\end{aligned}
$$

where we obtain the above from linearity of expectation and the Taylor Series expansion of $e^{\lambda X}$. Notice that this expansion results in each moment of X having some nonzero coefficient. In other words, all moments are "generated" (hence the name of the function). Furthermore, notice that selecting a varying value of λ results in different moments being prioritized. This is analogous to selecting some optimal value of k above, and we can use very similar ideas to obtain our desired bounds.

3.2.1 Moment Generating Function of a Gaussian

Let's put our new machinery to use by computing the MGF of a distribution.

Quite often, we work with Gaussians. We also get relatively well-bounded tail behavior with a Gaussian; hence, computing the MGF of a Gaussian seems like it might be useful. Consider the below:

$$
\begin{aligned}
M_X(\lambda) &= \int_{-\infty}^{\infty} p(t) e^{\lambda t}\, \mathrm{d}t \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-t^2/2\sigma^2} \cdot e^{\lambda t}\, \mathrm{d}t
\end{aligned}
$$

To simplify things a bit, we can complete the square in the exponents and manipulate that result:

$$
\begin{aligned}
& -\frac{t^2}{2\sigma^2} + \lambda t \\
&= -\left(\frac{t^2}{2\sigma^2} - \lambda t\right) \\
&= -\left(\left(\frac{t}{\sigma\sqrt{2}} - \frac{\sqrt{2}\sigma\lambda}{2}\right)^2 - \frac{\sigma^2\lambda^2}{2}\right) \\
&= -\left(\frac{t}{\sigma\sqrt{2}} - \frac{\sqrt{2}\sigma\lambda}{2}\right)^2 + \frac{\sigma^2\lambda^2}{2} \\
&= -\left(\frac{t}{\sigma\sqrt{2}} - \frac{\sigma^2\lambda}{\sigma\sqrt{2}}\right)^2 + \frac{\sigma^2\lambda^2}{2} \\
&= -\frac{\left(t - \sigma^2\lambda\right)^2}{2\sigma^2} + \frac{\sigma^2\lambda^2}{2}
\end{aligned}
$$

Define $p'(t)$ below:

$$p'(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-(t-\sigma^2\lambda)^2/2\sigma^2}$$

And note that we can rewrite our messy integral in terms of $p'(t)$:

$$\int_{-\infty}^{\infty} p'(t) e^{\sigma^2\lambda^2/2} \,\mathrm{d}t$$

Observe that $p'(t)$ represents a Gaussian that is horizontally shifted. Furthermore, observe that $\exp(\sigma^2\lambda^2)$ is constant in t. We now have:

$$
\begin{aligned}
& \int_{-\infty}^{\infty} p'(t) e^{\sigma^2\lambda^2/2} \,\mathrm{d}t \\
&= e^{\sigma^2\lambda^2} \int_{-\infty}^{\infty} p'(t) \,\mathrm{d}t \\
&= e^{\sigma^2\lambda^2/2}
\end{aligned}
$$

We therefore conclude that if X is Gaussian with variance σ^2, then:

$$M_X(\lambda) = e^{\lambda^2\sigma^2/2}$$

3.2.2 Subgaussian Random Variables

Subgaussian random variables are, intuitively, random variables which yield better tail bounds than Gaussian variables. We can formally define them here:

Definition 2. *A random variable with mean 0 is subgaussian(σ) if its moment generating function is not greater than that of a Gaussian with variance σ^2. A random variable X with nonzero mean is subgaussian if $X - \mathbb{E}[X]$ is subgaussian.*

Up to constant coefficients in σ, it can be shown that the following definitions are equivalent for a random variable which is subgaussian with parameter σ:

1. MGF Definition: $M_X(\lambda) \leq \exp(\lambda^2\sigma^2/2)$
2. Tail Definition: $\mathbb{P}[|X| > t] \leq 2\exp(-t^2/2\sigma^2)$
3. Moment Definition: $\mathbb{E}[|X|^k] \leq k^{k/2}\sigma^k$ for all integers $k \geq 1$

Using this, we can state the two following useful lemmas.

Lemma 3. *Any bounded random variable $X \in [a, a+c]$ is subgaussian with parameter $c/2$. [**?**]*

It directly follows from the above that if $X \in [-1, 1]$, then $\sigma = 1$.

Lemma 4 (Addition of Subgaussians)**.** *If we have two independent Subgaussian variables X with parameter σ_1 and Y with parameter σ_2, then $X + Y$ is subgaussian with parameter $\sqrt{\sigma_1^2 + \sigma_2^2}$.*

Proof.

$$\begin{aligned}
\mathbb{E}[e^{\lambda(X+Y-\mathbb{E}[X+Y])}] &= \mathbb{E}[e^{\lambda(X-\mathbb{E}[X])}]\,\mathbb{E}[e^{\lambda(Y-\mathbb{E}[Y])}] \\
&\leq e^{\frac{\lambda^2\sigma_1^2}{2}} e^{\frac{\lambda^2\sigma_2^2}{2}} \\
&= e^{\frac{\lambda^2(\sigma_1^2+\sigma_2^2)}{2}} \\
&= e^{\frac{\lambda^2(\sqrt{\sigma_1^2+\sigma_2^2})^2}{2}}
\end{aligned}$$

□

Corollary 5. *The additive Chernoff bound on independent random variables bounded in* $[0, 1]$ *holds.*

Proof. Let $X_1, \ldots, X_n \in [0, 1]$ be independent random variables. Then X_i are subgaussian with parameter $\sigma_i^2 = 1/4$ by Lemma **??**. Therefore $X = \sum_{i=1}^n X_i$ is subgaussian with parameter $\sigma^2 = n/4$ by Lemma **??**. Therefore

$$\begin{aligned}
\mathbb{P}[|X - \mathbb{E}[X]| \geq t] &\leq 2e^{\frac{-t^2}{2(n/4)}} \\
&= 2e^{-\frac{2t^2}{n}}
\end{aligned}$$

by the tail definition of a subgaussian variable. This is exactly the additive Chernoff bound. □

References

[1] Rajeev Motwani, Prabhakar Raghavan. Randomized Algorithms. *Cambridge University Press*, 0-521-47465-5, 1995.

[2] David Joyce. Moments and the moment generating function. *Math 217 Probability and Statistics*. Clark University, 2014.

[3] Philippe Rigollet. Lecture Notes. *High Dimensional Statistics*. MIT, 2015.