MATH 7251: High-Dimensional Probability **Spring 2022**

Lecture 7

Name: Tyler LaBonte **Professor:** Konstantin Tikhomorov

1 The Johnson-Lindenstrauss Lemma

The Johnson-Lindenstrauss lemma maps a metric space into a subset of Euclidean space of relatively low dimension, and it follows nicely from Bernstein-type inequalities. Suppose we have a collection $\mathcal{C} = \{\boldsymbol{x}_i\}_{i=1}^m \subseteq \mathbb{R}^n$ all distinct. Then, $(\mathcal{C}, \|\cdot\|_2)$ is a metric space. In order to compress signal, work in lower dimension, etc., we want to map this space into $\mathbb{R}^k$ where $k \ll n$ so that the metric structure is not distorted too much.

The idea is to use a random $k \times n$ matrix for the dimension reduction. Denote this linear map by $\Phi : \mathbb{R}^n \to \mathbb{R}^k$. We would like

$$1 - \epsilon \leq \frac{\|\Phi(\boldsymbol{x}) - \Phi(\boldsymbol{y})\|_2}{\|\boldsymbol{x} - \boldsymbol{y}\|_2} \leq 1 + \epsilon \tag{1}$$

for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{C}$. To summarize, we have parameters ϵ, k, n, m, and the question is what restriction on these parameters is necessary to find such a mapping Φ.

Lemma 1.1: Johnson-Lindenstrauss

There exists a universal constant $c > 0$ with the following property. Let $k \geq c\epsilon^{-2}\log(m)$ for $\epsilon \in (0, \frac{1}{2}]$. Then, there exists a linear mapping Φ satisfying condition (1). In particular, if $\mathbf{A}$ is a $k \times n$ matrix with iid entries of zero mean, unit variance, and constant subgaussian moment, then $\Phi = \frac{1}{\sqrt{k}}\mathbf{A}$ works with high probability.

The proof idea is given by the corollaries from the last lecture. For every $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{C}$, the vector

$$\boldsymbol{q} = \frac{\mathbf{A}(\boldsymbol{x} - \boldsymbol{y})}{\|\boldsymbol{x} - \boldsymbol{y}\|_2} \tag{2}$$

is isotropic, centered, and subgaussian with iid subgaussian components. This implies that

$$\Pr\left[\left|\|\boldsymbol{q}\|_2 - \sqrt{k}\right| \geq t\right] \leq 2\exp\left(-\tilde{c}t^2\right) \tag{3}$$

for all $t > 0$, where $\tilde{c} > 0$ may only depend on the subgaussian moment of the entries. If t is taken as an appropriate constant multiple of $\sqrt{\log(m)}$, then we get a quantity much less than m^{-2} on the right-hand side. So we can take a union bound over pairs of $\boldsymbol{x}, \boldsymbol{y}$ to obtain the result.

An interesting example is that if $\mathcal{C}$ is the standard basis in $\mathbb{R}^n$, applying Φ gives us m vectors in $\mathbb{R}^{c\log(n)}$ such that $\|\Phi(\boldsymbol{x}_i)\| \in [0.99, 1.01]$ for all i and $|\langle \Phi(\boldsymbol{x}_i), \Phi(\boldsymbol{x}_j)\rangle| \le 0.01$. So, we have just mapped exponentially many vectors into the space in such a way that they are almost unit and almost pairwise orthogonal.

2 Concentration of Sparse Linear Combinations

Theorem 2.1: Chernoff

Let $b_1, b_2, \ldots, b_n$ be iid $\mathrm{Ber}(p)$ random variables. Then,

$$\Pr\left[\sum_{i=1}^{n} b_i - pn \ge tpn\right] \le \left(\frac{\exp(t)}{(1+t)^{(1+t)}}\right)^{pn} \tag{4}$$

for $t > 0$. For the other tail,

$$\Pr\left[\sum_{i=1}^{n} b_i - pn \le -tpn\right] \le \left(\frac{\exp(-t)}{(1-t)^{(1-t)}}\right)^{pn} \tag{5}$$

for $t \in (0, 1)$.

Proof: Of the first inequality. Choose $\lambda > 0$ (we will optimize over λ later). Using MGF,

$$\Pr\left[\sum_{i=1}^{n} b_i - pn \ge tpn\right] = \Pr\left[\exp\left(\lambda\sum_{i=1}^{n} b_i - \lambda pn\right) \ge \exp(\lambda tpn)\right]. \tag{6}$$

Applying Markov's inequality, this is at most

$$\frac{\mathbb{E}[\exp(\lambda\sum_{i=1}^{n} b_i - \lambda pn)]}{\exp(\lambda tpn)} = \frac{(\mathbb{E}[\exp(\lambda b_1 - \lambda p)])^n}{\exp(\lambda tpn)}. \tag{7}$$

Since the variables are Bernoulli we can easily and directly compute the MGF. So the above is equal to

$$\frac{((1-p)\exp(-\lambda p) + p\exp(\lambda - \lambda p))^n}{\exp(\lambda tpn)} = \frac{\exp(-\lambda pn)}{\exp(\lambda tpn)}(1 - p + p\exp(\lambda))^n. \tag{8}$$

If we take $\lambda = \ln(1+t)$ and use $1 + s \le \exp(s)$, this simplifies to

$$\frac{(1+pt)^n}{\exp(tpn\ln(1+t) + pn\ln(1+t))} \le \frac{\exp(ptn)}{(1+t)^{tpn+pn}} = \left(\frac{\exp(t)}{(1+t)^{(1+t)}}\right)^{pn}. \tag{9}$$

■

Corollary 2.1

For a universal constant $c > 0$,

$$\Pr\left[\left|\sum_{i=1}^{n} b_i - pn\right| \geq t\right] \leq 2\exp\left(-\frac{ct^2}{pn}\right). \tag{10}$$

Note that this Chernoff-based bound is stronger than Hoeffding's inequality in the regime $p \to 0$. If we apply the version of Hoeffding for variables supported on bounded intervals, we would get

$$\Pr\left[\left|\sum_{i=1}^{n} b_i - pn\right| \geq t\right] \leq 2\exp\left(-\frac{ct^2}{n}\right). \tag{11}$$

This is similar for the other version of Hoeffding and the general Bernstein inequality.

Proof: For every $t \in (0, pn]$ we have by Chernoff

$$\Pr\left[\left|\sum_{i=1}^{n} b_i - pn\right| \geq t\right] \leq \Pr\left[\sum_{i=1}^{n} b_i - pn \geq t\right] + \Pr\left[\sum_{i=1}^{n} b_i - pn \leq -t\right] \tag{12}$$

$$\leq \left(\frac{\exp\left(\frac{t}{pn}\right)}{\left(1+\frac{t}{pn}\right)^{1+\frac{t}{pn}}}\right)^{pn} + \left(\frac{\exp\left(-\frac{t}{pn}\right)}{\left(1-\frac{t}{pn}\right)^{1-\frac{t}{pn}}}\right)^{pn} \tag{13}$$

$$= \exp\left(t - pn\left(1+\frac{t}{pn}\right)\ln\left(1+\frac{t}{pn}\right)\right) + \exp\left(-t - pn\left(1-\frac{t}{pn}\right)\ln\left(1-\frac{t}{pn}\right)\right). \tag{14}$$

We now apply the fact from calculus that for all t such that $\frac{t}{pn} \in (0, 1]$, we have

$$\frac{t}{pn} - \left(1+\frac{t}{pn}\right)\ln\left(1+\frac{t}{pn}\right) \leq -\frac{1}{4}\frac{t^2}{p^2n^2}. \tag{15}$$

This can be checked directly by taking derivatives. So, the first term in (14) is at most

$$\exp\left(-\frac{pn}{4}\frac{t^2}{p^2n^2}\right) = \exp\left(-\frac{1}{4}\frac{t^2}{pn}\right). \tag{16}$$

Similar ideas work for the second term in (14), and the corollary follows. ■

Lemma 2.1: Sparse Bernstein Inequality

Let $b_1, b_2, \ldots, b_n$ be iid $\text{Ber}(p)$ random variables and $a_1, a_2, \ldots, a_n \in \mathbb{R}^n$ such that $\|\boldsymbol{a}\|_2 = 1$. Then, for all $t > 0$ we have

$$\Pr\left[\sum_{i=1}^n a_i b_i - p \sum_{i=1}^n a_i \geq tp\right] \leq \exp\Big(-cp \min\left(t^2, \|\boldsymbol{a}\|_\infty^{-1} t\right)\Big). \tag{17}$$

and

$$\Pr\left[\sum_{i=1}^n a_i b_i - p \sum_{i=1}^n a_i \leq -tp\right] \leq \exp\Big(-cp \min\left(t^2, \|\boldsymbol{a}\|_\infty^{-1} t\right)\Big). \tag{18}$$

This statement can be viewed as a generalization of the last corollary (it is exactly the scenario with all $a_i = 1/\sqrt{n}$).

Proof: Of the first inequality. We use the MGF/Markov strategy:

$$\begin{aligned}
\Pr\left[\sum_{i=1}^n a_i b_i - p \sum_{i=1}^n a_i \geq tp\right] &= \Pr\left[\exp\left(\lambda \sum_{i=1}^n a_i b_i - \lambda p \sum_{i=1}^n a_i\right) \geq \exp(\lambda tp)\right] && (19)\\
&\leq \frac{\mathbb{E}[\exp(\lambda \sum_{i=1}^n a_i b_i - \lambda p \sum_{i=1}^n a_i)]}{\exp(\lambda tp)} && (20)\\
&= \frac{\prod_{i=1}^n \mathbb{E}[\exp(\lambda a_i b_i - \lambda p a_i)]}{\exp(\lambda tp)}. && (21)
\end{aligned}$$

We can compute the numerator explicitly to find

$$\frac{\prod_{i=1}^n [(1-p)\exp(-\lambda p a_i) + p \exp(\lambda a_i - \lambda p a_i)]}{\exp(\lambda tp)}. \tag{22}$$

Again using $1 + x \leq \exp(x)$,

$$\frac{\prod_{i=1}^n [(1-p) + p\exp(\lambda a_i)]}{\exp(\lambda tp + \lambda p \sum_{i=1}^n a_i)} \leq \frac{\exp(p(\exp(\lambda a_i) - 1))}{\exp(\lambda tp + \lambda p \sum_{i=1}^n a_i)}. \tag{23}$$

This is at most

$$\exp\left(p \sum_{i=1}^n (\exp(\lambda a_i) - 1) - \lambda tp - \lambda p \sum_{i=1}^n a_i\right) = \exp\left(\lambda p \left(\sum_{i=1}^n \frac{\exp(\lambda a_i) - \lambda a_i - 1}{\lambda} - t\right)\right). \tag{24}$$

Note that the difference in the sum is not too large and can be approximated by the square of the number in the exponent. That is, if $\lambda \leq \|\boldsymbol{a}\|_\infty^{-1}$ so that $|\lambda a_i| \leq 1$ for all i, then we have

$$\sum_{i=1}^n \frac{\exp(\lambda a_i) - \lambda a_i - 1}{\lambda} \leq \tilde{c} \sum_{i=1}^n \frac{\lambda^2 a_i^2}{\lambda} = \tilde{c}\lambda. \tag{25}$$

So (24) is at most

$$\exp(\lambda p(\tilde{c}\lambda - t)) \tag{26}$$

provided that $0 \le \lambda \le \|\boldsymbol{a}\|_\infty^{-1}$. Now we just need the right choice of λ based on t. If $0 \le t \le \|\boldsymbol{a}\|_\infty^{-1}$, we take $\lambda = \frac{t}{2\tilde{c}}$ which gives

$$\exp\left(\frac{t}{2\tilde{c}}p(-\frac{t}{2})\right) = \exp\left(-\frac{t^2}{4\tilde{c}}p\right). \tag{27}$$

This is the subgaussian tail. If $t \ge \|\boldsymbol{a}\|_\infty^{-1}$, we take $\lambda = \frac{1}{2}\|\boldsymbol{a}\|_\infty^{-1}$ to find

$$\exp\left(\frac{1}{2\tilde{c}}\|\boldsymbol{a}\|_\infty^{-1} p(-\frac{t}{2})\right) = \exp\left(-\frac{t\|\boldsymbol{a}\|_\infty^{-1} p}{4\tilde{c}}\right). \tag{28}$$

This is the subexponential tail. ■