Mathematical Toolkit **Autumn 2021**

Lecture 17: November 26, 2019

Lecturer: Madhur Tulsiani

1 Gaussian Random Variables

Recall that a Gaussian random variable X is defined through the density function

$$\gamma(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

where μ is its mean and σ^2 is its variance, and we write $X \sim \mathcal{N}(\mu, \sigma^2)$. Before proceeding to applications of Gaussian random variables, we prove the following fact which we will use repeatedly.

Proposition 1.1 *Let $Z = c_1X_1 + c_2X_2$, where $X_1 \sim \mathcal{N}(0,1)$ and $X_2 \sim \mathcal{N}(0,1)$ are independent. Then $Z \sim \mathcal{N}(0, c_1^2 + c_2^2)$.*

Proof: By a simple change of variable, we can check that the density function for c_1X_1 is $\frac{1}{\sqrt{2\pi}|c_1|}e^{-\frac{x^2}{2c_1^2}}$, which shows that $c_1X_1 \sim \mathcal{N}(0, c_1^2)$, and similarly $c_2X_2 \sim \mathcal{N}(0, c_2^2)$.
Next, we can check that if X and Y are independent random variables with densities f and g, then for $Z = X + Y$, we have

$$\mathbb{P}[Z \le t] = \int_{\infty}^{t} \left(\int_{-\infty}^{\infty} f(x) \cdot g(z-x) dx \right) dz,$$

which gives the density of Z as $h(z) = \int_{-\infty}^{\infty} f(x) \cdot g(z-x)dx$. Taking $X = c_1X$ and $Y = c_2X_2$, we get the density of $Z = c_1X_1 + c_2X_2$ is

$$h(z) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,|c_1|} \cdot e^{-\frac{x^2}{2c_1^2}} \cdot \frac{1}{\sqrt{2\pi}\,|c_2|} \cdot e^{-\frac{(z-x)^2}{2c_2^2}} dx.$$

We leave it as an exercise to show that the above integral gives

$$h(z) = \frac{1}{\sqrt{2\pi(c_1^2 + c_2^2)}} \cdot e^{-\frac{z^2}{2(c_1^2+c_2^2)}},$$

which implies $c_1X_1 + c_2X_2 \sim \mathcal{N}(0, c_1^2 + c_2^2)$. ■

One can obtain the following corollary using an inductive application of the above proposition.

Corollary 1.2 *Let $X_1, \ldots, X_n \sim \mathcal{N}(0,1)$ be independent standard Gaussian random variables. Then, for any vector of coefficients $c = (c_1, \ldots, c_n)$, we have*

$$Z \;=\; c_1 X_1 + \cdots + c_n X_n \;\sim\; \mathcal{N}(0, \|c\|^2)\,,$$

where $\|c\|^2 = c_1^2 + \cdots + c_n^2$.

2 Johnson–Lindenstrauss Lemma

We will use concentration bounds on Gaussian random variables to prove the following important lemma.

Lemma 2.1 (Johnson–Lindenstrauss [JL84]) *Let $\mathcal{P}$ be a set of n points in $\mathbb{R}^d$. Let $0 < \varepsilon < 1$. For $k = \frac{8 \ln n}{\varepsilon^2/2 - \varepsilon^3/3}$, there exists a mapping $\varphi : \mathcal{P} \to \mathbb{R}^k$ such that for all $u, v \in \mathcal{P}$*

$$(1-\varepsilon)\|u-v\|^2 \leq \|\varphi(u) - \varphi(v)\|^2 \leq (1+\varepsilon)\|u-v\|^2\,.$$

The above lemma is useful for dimensionality reduction, especially when a problem has an exponential dependence on the number of dimensions.

We construct the mapping φ as follows. First choose a matrix $G \in \mathbb{R}^{k \times d}$ such that each $G_{ij} \sim \mathcal{N}(0,1)$ is independent. Define

$$\varphi(u) = \frac{Gu}{\sqrt{k}}\,.$$

Note that by the above construction φ is oblivious, meaning that it doesn't depend on the points in $\mathcal{P}$, and it is linear.

The strategy of proving the lemma is to first prove that with high probability the lemma holds for any fixed two points and then apply union bounds to get the result for all pairs of points.

Claim 2.2 *Fix $u, v \in \mathcal{P}$. Let $w = u - v$. With probability greater than $1 - 1/n^3$, the following inequality holds,*

$$(1-\varepsilon) \cdot \|w\|^2 \;\leq\; \|\varphi(w)\|^2 \;\leq\; (1+\varepsilon) \cdot \|w\|^2\,.$$

Proof: Recall that $\varphi(u) = \frac{Gu}{\sqrt{k}}$. Let

$$Z = \frac{k\|\varphi(w)\|^2}{\|w\|^2} = \frac{\sum_{i=1}^k (Gw)_i^2}{\|w\|^2}.$$

We need to show $(1-\varepsilon)k \leq Z \leq (1+\varepsilon)k$. We know that the sum of Gaussian random variables is still a Gaussian random variable, so $(Gw)_i = G_i w = \sum_{j=1}^n G_{ij} w_j$ is a Gaussian variable. Besides, $\mathsf{Var}\left[\sum_{j=1}^n G_{ij} w_j\right] = \sum_j w_j^2 = \|w\|^2$ according to Fact **??**. In other words, $G_i w \sim \mathcal{N}(0, \|w\|^2)$. As a result, $Z = \sum_{i=1}^k \frac{(Gw)_i^2}{\|w\|^2} = \sum_{i=1}^k X_i^2$, where $X_i \sim \mathcal{N}(0,1)$. The expectation of each individual element in Gw is

$$\mathbb{E}\left[(Gw)_i^2\right] = \mathbb{E}\left[(G_i w)^2\right] = \mathbb{E}\left[\left(\sum_{j=1}^n G_{ij} w_j\right)^2\right] = \mathsf{Var}\left[\sum_{j=1}^n G_{ij} w_j\right] = \|w\|^2 .$$

In addition,

$$\mathbb{E}\left[Z\right] = \frac{\sum_{j=1}^k \mathbb{E}\left[(Gw)_i^2\right]}{\|w\|^2} = k .$$

Now we prove the concentration bound for Z. The proof is almost identical to Chernoff bound.

$$\begin{aligned}
\mathbb{P}\left[Z \geq (1+\varepsilon)k\right] &\leq \mathbb{P}\left[e^{tZ} \geq e^{\lambda\cdot(1+\varepsilon)k}\right] \\
&\leq \frac{\mathbb{E}\left[e^{\lambda\cdot Z}\right]}{e^{\lambda\cdot(1+\varepsilon)k}} && \text{(by Markov's inequality)} \\
&= \frac{\mathbb{E}\left[e^{\lambda\cdot\sum_{i=1}^k X_i^2}\right]}{e^{\lambda\cdot(1+\varepsilon)k}} = \frac{\prod_{i=1}^k \mathbb{E}\left[e^{\lambda\cdot X_i^2}\right]}{e^{\lambda\cdot(1+\varepsilon)k}} && \text{(by the independence of } X_1,\ldots,X_k\text{)} \\
&= \frac{\prod_{i=1}^k \frac{1}{\sqrt{1-2\lambda}}}{e^{\lambda\cdot(1+\varepsilon)k}} && \text{(by Lemma ??)} \\
&\leq \left(\frac{e^{-2(1+\varepsilon)\lambda}}{1-2\lambda}\right)^{k/2} && \text{(assume } \lambda < 1/2\text{)} \\
&\leq \left(e^{-\varepsilon}(1+\varepsilon)\right)^{k/2} && \text{(let } \lambda = \frac{\varepsilon}{2(1+\varepsilon)}\text{)} \\
&\leq \left((1-\varepsilon+\frac{\varepsilon^2}{2})(1+\varepsilon)\right)^{k/2} && \text{(by Taylor expansion of } e^{-x}\text{)} \\
&\leq e^{-\left(\frac{\varepsilon^2}{2}-\frac{\varepsilon^3}{2}\right)\frac{k}{2}} && \text{(by } 1+x \leq e^x\text{)}
\end{aligned}$$

We can derive the other side of the inequality in an analogous way. Thus, we have

$$\mathbb{P}\left[|Z-k| \geq \varepsilon k\right] \leq 2\cdot\exp\left(-\left(\frac{\varepsilon^2}{2}-\frac{\varepsilon^3}{2}\right)\frac{k}{2}\right) \leq 2\cdot\exp\left(-3\ln n\right) = \frac{2}{n^3} ,$$

where we choose

$$k = \left\lceil \frac{6\ln n}{\frac{\varepsilon^2}{2}-\frac{\varepsilon^3}{2}} \right\rceil .$$

$\blacksquare$

To prove Johnson–Lindenstrauss Lemma, we apply the union bound and get the desired result

$$\begin{aligned}\mathbb{P}\left[\forall u, v \in \mathcal{P}, (1-\varepsilon)\|u-v\|^2 \leq \|\varphi(u)-\varphi(v)\|^2 \leq (1+\varepsilon)\|u-v\|^2\right] &\geq 1 - \binom{n}{2}\frac{2}{n^3} \\ &\geq 1 - \frac{1}{n}\,.\end{aligned}$$

References

[JL84] W Johnson and J Lindenstrauss, *Extensions of Lipschhitz maps into a Hilbert space*, Contemporary Math **26** (1984). 2