Mathematical Toolkit	Autumn 2015

Lecture 17: November 25, 2015

Lecturer: Madhur Tulsiani

1 Real-valued random variables and distributions

Let μ be any probability measure over the space $\mathbb{R}$ equipped with the Borel σ-algebra. Define the function F as

$$F(x) \ := \ \mu((-\infty, x])\,,$$

which is well defined since the interval $(-\infty, x]$ is in the σ-algebra. This can be used to define a random variable X such that $\mathbb{P}[X \leq x] = F(x)$. The function F is known as the distribution function or the cummulative density function of X.

When the function F has the form

$$F(x) \ = \ \int_{-\infty}^{x} f(z)dz\,,$$

then f is called the density function of the random variable X.

2 Gaussian Random Variables

A Gaussian random variable X is defined through the density function

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

where μ is its mean and σ^2 is its variance, and we write $X \sim \mathcal{N}(\mu, \sigma^2)$. To see the definition gives a valid probability distribution, we need to show $\int_{-\infty}^{\infty} \gamma(x)dx = 1$. It suffices to show for the case that $\mu = 0$ and $\sigma^2 = 1$. First we show the integral is bounded.

Claim 2.1 *$I = \int_{-\infty}^{\infty} e^{-x^2/2}dx$ is bounded.*

Proof: We see that

$$I = \int_{-\infty}^{\infty} e^{-x^2/2}dx = 2\int_{0}^{\infty} e^{-x^2/2}dx \leq 2\int_{0}^{2} 1dx + 2\int_{2}^{\infty} e^{-x}dx = 4 + 2e^{-2}\,,$$

where we use the fact that I is even and after $x = 2$, $e^{-x^2/2}$ is upper bounded by e^{-x}. ■

Next we show that the normalization factor is $\sqrt{2\pi}$.

Claim 2.2 $I^2 = 2\pi$.

Proof:

$$\begin{aligned} I^2 &= \int_{-\infty}^{\infty} e^{-x^2/2} dx \int_{-\infty}^{\infty} e^{-y^2/2} dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)/2} dx dy \\ &= \int_0^{\infty}\int_0^{2\pi} e^{-r^2/2} r dr d\theta \quad (\text{let } x = r\cos\theta \text{ and } y = r\sin\theta) \\ &= 2\pi \int_0^{\infty} e^{-s} ds \quad (\text{let } s = r^2/2) \\ &= 2\pi\,. \end{aligned}$$

■

This completes the proof that the definition gives a valid probability distribution. We prove a useful lemma for later use.

Lemma 2.3 *For* $X \sim \mathcal{N}(0,1)$ *and* $\lambda \in (0, 1/2)$,

$$\mathbb{E}\left[e^{\lambda \cdot X^2}\right] = \frac{1}{\sqrt{1-2\lambda}}\,.$$

Proof:

$$\begin{aligned} \mathbb{E}\left[e^{\lambda \cdot X^2}\right] &= \int_{-\infty}^{\infty} e^{\lambda \cdot x^2} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(1-2\lambda)x^2/2} dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \frac{dy}{\sqrt{1-2\lambda}} \quad (\text{let } y = \sqrt{1-2\lambda}x) \\ &= \frac{1}{\sqrt{1-2\lambda}} \end{aligned}$$

■

3 Johnson–Lindenstrauss Lemma

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

Lemma 3.1 (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* $u, v \in \mathcal{P}$

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

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. Before we prove the lemma, we will use the following fact several times.

Fact 3.2 *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 $Z \sim \mathcal{N}(0, c_1^2 + c_2^2)$.*

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 3.3 *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 3.2. 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.

$$\begin{aligned}
\mathbb{P}\left[Z \geq (1+\varepsilon)k\right] &\leq \mathbb{P}\left[e^{tZ} \geq e^{\lambda\cdot(1+\varepsilon)k}\right] \\
&\leq \frac{\mathbb{E}\left[e^{\lambda\cdot Z}\right]}{e^{\lambda\cdot(1+\varepsilon)k}} && \text{(by Markov's inequality)} \\
&= \frac{\mathbb{E}\left[e^{\lambda\cdot\sum_{i=1}^{k} X_i^2}\right]}{e^{\lambda\cdot(1+\varepsilon)k}} = \frac{\prod_{i=1}^{k}\mathbb{E}\left[e^{\lambda\cdot X_i^2}\right]}{e^{\lambda\cdot(1+\varepsilon)k}} && \text{(by the independence of } X_1,\ldots,X_k) \\
&= \frac{\prod_{i=1}^{k}\frac{1}{\sqrt{1-2\lambda}}}{e^{\lambda\cdot(1+\varepsilon)k}} && \text{(by Lemma 2.3)} \\
&\leq \left(\frac{e^{-2(1+\varepsilon)\lambda}}{1-2\lambda}\right)^{k/2} && \text{(assume } \lambda < 1/2) \\
&\leq \left(e^{-\varepsilon}(1+\varepsilon)\right)^{k/2} && \text{(let } \lambda = \frac{\varepsilon}{2(1+\varepsilon)}) \\
&\leq \left((1-\varepsilon+\frac{\varepsilon^2}{2})(1+\varepsilon)\right)^{k/2} && \text{(by Taylor expansion of } e^{-x}) \\
&\leq e^{-\left(\frac{\varepsilon^2}{2}-\frac{\varepsilon^3}{2}\right)\frac{k}{2}} && \text{(by } 1+x \leq e^x)
\end{aligned}$$

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

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

■

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).