Chapter 23

Dimension Reduction

By Sariel Har-Peled, April 26, 2022[1]

23.1. Introduction to dimension reduction

Given a set P of n points in $\mathbb{R}^d$, we need nd numbers to describe them. In many scenarios, d might be quite large, or even larger than n (in some applications, where the access to the points is given only through dot-product, it is useful to think about the dimension as being unbounded). If we care only about the distances between any pairs of points, then all we need to store are the pairwise distances between the points. This would require roughly n^2 numbers, if we just write down the distance matrix.

But can we do better? (I.e., use less space.) A natural idea is to reduce the dimension of the points. Namely, replace the ith point $\mathsf{p}_i \in P$, by a point $\mathsf{u}_i \in \mathbb{R}^k$, where $k \ll d$ and $k \ll n$. We would like k to be small. If we can do that, then we compress the data from size dn to size kn, which might be a large compression.

Of course, one can do such compression of information without losing some information. In particular, we are willing to let the distances to be a bit off. Formally, we would like to have the property that $(1-\varepsilon)\|\mathsf{p}_i - \mathsf{p}_j\| \leq \|\mathsf{u}_i - \mathsf{u}_j\| \leq (1+\varepsilon)\|\mathsf{p}_i - \mathsf{p}_j\|$, for all i, j, where u_i is the image of $\mathsf{p}_i \in P$ after the dimension reduction.

To this end, we generate a random matrix M of dimensions $d \times k$, where $k = \Theta(\varepsilon^{-2} \log n)$ (the exact details of how to generate this matrix are below, but informally every entry is going to be picked from a normal distribution and scaled appropriately). We then set $\mathsf{u}_i = \mathsf{M}\mathsf{p}_i$, for all $\mathsf{p}_i \in P$.

Before dwelling on the details, we need to better understand the normal distribution.

23.2. Normal distribution

The ***standard normal distribution*** has

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\bigl(-x^2/2\bigr) \tag{23.1}$$

as its density function. We denote that X is distributed according to such distribution, using $X \sim \mathbf{N}(0,1)$. It is depicted in Figure 23.1.

Somewhat strangely, it would be convenient to consider two such independent variables X and Y together. Their probability space (X, Y) is the plane, and it defines a two dimensional density function

$$g(x,y) = f(x)f(y) = \frac{1}{2\pi} \exp\bigl(-(x^2+y^2)/2\bigr). \tag{23.2}$$

The key property of this function is that $g(x,y) = g(x',y') \iff \|(x,y)\|^2 = x^2 + y^2 = \|(x',y')\|^2$. Namely, $g(x,y)$ is symmetric around the origin (i.e., all the points in the same distance from the origin

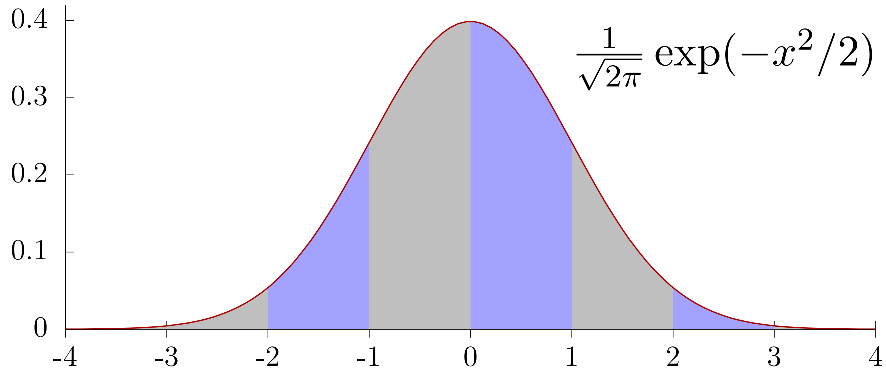

Figure 23.1

have the same density). We next use this property in verifying that $f(\cdot)$ it is indeed a valid density function.

Lemma 23.2.1. *We have* $I = \int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$*, where* $f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-x^2/2\right)$.

Proof: Observe that

$$I^2 = \left(\int_{x=-\infty}^{\infty} f(x)\ \mathrm{d}x\right)^2 = \left(\int_{x=-\infty}^{\infty} f(x)\ \mathrm{d}x\right)\left(\int_{y=-\infty}^{\infty} f(y)\ \mathrm{d}y\right) = \int_{x=-\infty}^{\infty}\int_{y=-\infty}^{\infty} f(x)f(y)\,dx\,dy$$
$$= \int_{x=-\infty}^{\infty}\int_{y=-\infty}^{\infty} g(x,y)\,\mathrm{d}x\,\mathrm{d}y.$$

Change the variables to $x = r\cos\alpha$, $y = r\sin\alpha$, and observe that the determinant of the Jacobian is

$$J = \det\begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \alpha} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \alpha} \end{vmatrix} = \det\begin{vmatrix} \cos\alpha & -r\sin\alpha \\ \sin\alpha & r\cos\alpha \end{vmatrix} = r\left(\cos^2\alpha + \sin^2\alpha\right) = r.$$

As such,

$$I^2 = \frac{1}{2\pi}\int_{r=0}^{\infty}\int_{\alpha=0}^{2\pi} \exp\left(-\frac{r^2}{2}\right)|J|\ d\alpha\,dr = \frac{1}{2\pi}\int_{r=0}^{\infty}\int_{\alpha=0}^{2\pi} \exp\left(-\frac{r^2}{2}\right) r\,d\alpha\,dr$$
$$= \int_{r=0}^{\infty} \exp\left(-\frac{r^2}{2}\right) r\,dr = \left[-\exp\left(-\frac{r^2}{2}\right)\right]_{r=0}^{r=\infty} = -\exp(-\infty) - (-\exp(0)) = 1. \qquad \blacksquare$$

Lemma 23.2.2. *For* $X \sim \mathbf{N}(0,1)$*, we have that* $\mathbb{E}[X] = 0$ *and* $\mathbb{V}[X] = 1$.

Proof: The density function of X, see Eq. (23.2) is symmetric around 0, which implies that $\mathbb{E}[X] = 0$. As for the variance, we have

$$\mathbb{V}[X] = \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2 = \mathbb{E}\left[X^2\right] = \int_{x=-\infty}^{\infty} x^2\,\mathbb{P}[X = x]\,\mathrm{d}x = \frac{1}{\sqrt{2\pi}}\int_{x=-\infty}^{\infty} x^2\exp(-x^2/2)\,\mathrm{d}x.$$

Observing that

$$x^2\exp\left(-x^2/2\right) = \left(-x\exp(-x^2/2)\right)' + \exp\left(-x^2/2\right),$$

implies (using integration by guessing) that

$$\mathbb{V}[X] = \frac{1}{\sqrt{2\pi}}\left[-x\exp(-x^2/2)\right]_{x=-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp(-x^2/2)\,\mathrm{d}x = 0 + 1 = 1. \qquad \blacksquare$$

23.2.1. The standard multi-dimensional normal distribution

The ***multi-dimensional normal distribution***, denoted by $\mathbf{N}^d$, is the distribution in $\mathbb{R}^d$ that assigns a point $\mathsf{p} = (\mathsf{p}_1, \ldots, \mathsf{p}_d)$ the density $g(\mathsf{p}) = \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{d} \mathsf{p}_i^2\right)$.

It is easy to verify, using the above, that $\int_{\mathbb{R}^d} g(\mathsf{p}) d\mathsf{p} = 1$. Furthermore, we have the following useful but easy properties.[②]

Lemma 23.2.3. *We have the following properties:*

(A) Consider d independent variables $X_1, \ldots, X_d \sim \mathbf{N}(0,1)$, the point $\mathsf{u} = (X_1, \ldots, X_d)$ has the multi-dimensional normal distribution $\mathbf{N}^d$.

(B) The multi-dimensional normal distribution is symmetric. For any two points $\mathsf{p}, \mathsf{u} \in \mathbb{R}^d$ such that $\|\mathsf{p}\| = \|\mathsf{u}\|$, we have that $g(\mathsf{p}) = g(\mathsf{u})$, where $g(\cdot)$ is the density function of the multi-dimensional normal distribution $\mathbf{N}^d$.

(C) The projection of the normal distribution on any direction (i.e., any vector of length 1) is a one-dimensional normal distribution.

Proof: (A) Let $f(\cdot)$ denote the density function of $\mathbf{N}(0,1)$, and observe that the density function of u is $f(X_1)f(X_2)\cdots f(X_d), = \frac{1}{\sqrt{2\pi}}\exp\left(-X_1^2/2\right)\cdots\frac{1}{\sqrt{2\pi}}\exp\left(-X_d^2/2\right)$, which readily implies the claim.

(B) Readily follows from observing that $g(\mathsf{p}) = \frac{1}{(2\pi)^{d/2}}\exp\left(-\|\mathsf{p}\|^2/2\right)$.

(C) Let $\mathsf{p} = (X_1, \ldots, X_d)$, where $X_1, \ldots, X_d \sim \mathbf{N}(0,1)$. Let v be any unit vector in $\mathbb{R}^d$, and observe that by the symmetry of the density function, we can (rigidly) rotate the space around the origin in any way we want, and the measure of sets does not change. In particular rotate space so that v becomes the unit vector $(1, 0, \ldots, 0)$. We have that

$$\mathbb{P}[\langle v, \mathsf{p}\rangle \leq \alpha] = \mathbb{P}[\langle (1, 0, \ldots, 0), \mathsf{p}\rangle \leq \alpha] = \mathbb{P}[X_1 \leq \alpha],$$

which implies that $\langle v, \mathsf{p}\rangle \sim X_1 \sim \mathbf{N}(0,1)$. ■

The generalized multi-dimensional distribution is a ***Gaussian***. Fortunately, we only need the simpler notion.

23.3. Dimension reduction

23.3.1. The construction

The input is a set $P \subseteq \mathbb{R}^d$ of n points (where d is potentially very large), and let $\varepsilon > 0$ be an approximation parameter. For

$$k = \left\lceil 24\varepsilon^{-2}\ln n\right\rceil \tag{23.3}$$

we pick k vectors $u_1, \ldots, u_k$ independently from the d-dimensional normal distribution $\mathbf{N}^d$. Given a point $\mathsf{p} \in \mathbb{R}^d$, its image is

$$h(v) = \frac{1}{\sqrt{k}}\left(\langle u_1, \mathsf{p}\rangle, \cdots, \langle u_k, \mathsf{p}\rangle\right).$$

[②] The normal distribution has such useful properties that it seems that the only thing normal about it is its name.

In matrix notation, let

$$\mathsf{M} = \frac{1}{\sqrt{k}} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_k \end{pmatrix}.$$

For every point $\mathsf{p}_i \in P$, we set $\mathsf{u}_i = h(\mathsf{p}_i) = \mathsf{M}\mathsf{p}_i$.

23.3.2. Analysis

23.3.2.1. A single unit vector is preserved

Consider a vector v of length one in $\mathbb{R}^d$. The natural question is what is the value of k needed, so that the length of $h(v)$ is a good approximation to v. Since $\langle u_i, v\rangle \sim \mathbf{N}(0,1)$, by Lemma 23.2.3, this question can boil down to the following: Given k variables $X_1, \ldots, X_k \sim \mathbf{N}(0,1)$, sampled independently, how concentrated is the random variable

$$Y = \|(X_1, \ldots, X_k)\|^2 = \sum_{i=1}^{k} X_i^2.$$

We have that $\mathbb{E}[Y] = k\,\mathbb{E}\big[X_i^2\big] = k\,\mathbb{V}[X_i] = k$, since $X_i \sim \mathbf{N}(0,1)$, for any i. The distribution of Y is known as the ***chi-square distribution with k degrees of freedom***.

Lemma 23.3.1. *Let $\varphi \in (0,1)$, and $\varepsilon \in (0,1/2)$ be parameters, and let $k \geq \left\lceil \frac{16}{\varepsilon^2} \ln \frac{2}{\varphi} \right\rceil$ be an integer. Then, for k independent random variables $X_1, \ldots, X_k \sim \mathbf{N}(0,1)$, we have that $Z = \sum_i X_i^2/k$ is strongly concentrated. Formally, we have that $\mathbb{P}[Z \leq 1+\varepsilon] \geq 1-\varphi$.*

Proof: Arguing as in the proof of Chernoff's inequality, using $t = \varepsilon/4 < 1/2$, we have

$$\mathbb{P}[Z \geq 1+\varepsilon] \leq \mathbb{P}\Big[\exp(tkZ) \geq \exp\big(tk(1+\varepsilon)\big)\Big] \leq \frac{\mathbb{E}[\exp(tkZ)]}{\exp\big(tk(1+\varepsilon)\big)} = \prod_{i=1}^{k} \frac{\mathbb{E}\big[\exp(tX_i^2)\big]}{\exp\big(t(1+\varepsilon)\big)}.$$

Using the substitution $x = \frac{y}{\sqrt{1-2t}}$ and $\mathrm{d}x = \frac{1}{\sqrt{1-2t}}\,\mathrm{d}y$, we have

$$\begin{aligned}
\mathbb{E}\big[\exp(tX_i^2)\big] &= \int_{x=-\infty}^{\infty} \frac{\exp(tx^2)}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \mathrm{d}x = \frac{1}{\sqrt{2\pi}} \int_{x=-\infty}^{\infty} \exp\left(-(1-2t)\frac{x^2}{2}\right) \mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}} \int_{y=-\infty}^{\infty} \exp\left(-\frac{1-2t}{2}\left(\frac{y}{\sqrt{1-2t}}\right)^2\right) \frac{1}{\sqrt{1-2t}}\,\mathrm{d}y = \frac{1}{\sqrt{1-2t}} \cdot \frac{1}{\sqrt{2\pi}} \int_{y=-\infty}^{\infty} \exp\left(-\frac{y^2}{2}\right) \mathrm{d}y \\
&= \frac{1}{\sqrt{1-2t}}.
\end{aligned}$$

We have that $\frac{1}{1-z} = \sum_{i=0}^{\infty} z^i$, for $0 \leq z < 1$, and thus

$$\tfrac{1}{1-\varepsilon/2} = \sum_i \left(\frac{\varepsilon}{2}\right)^i \leq \left[1 + \frac{1}{2}\sum_{i=1}^{\infty}\left(\frac{\varepsilon}{2}\right)^i\right]^2 \leq \exp\left[\frac{1}{2}\sum_{i=1}^{\infty}\left(\frac{\varepsilon}{2}\right)^i\right]^2.$$

Since $t = \varepsilon/4$, we have

$$\mathbb{E}\Big[\exp(tX_i^2)\Big] = \frac{1}{\sqrt{1-2t}} = \frac{1}{\sqrt{1-\varepsilon/2}} \le \exp\left(\frac{1}{2}\sum_{i=1}^{\infty}\left(\frac{\varepsilon}{2}\right)^i\right).$$

As such, we have

$$\mathbb{P}[Z \ge 1+\varepsilon] \le \exp\left(\frac{1}{2}\sum_{i=1}^{\infty}\left(\frac{\varepsilon}{2}\right)^i - \frac{\varepsilon}{4}(1+\varepsilon)\right)^k = \exp\left(-\frac{\varepsilon^2}{8} + \frac{1}{2}\sum_{i=3}^{\infty}\left(\frac{\varepsilon}{2}\right)^i\right)^k \le \exp\left(-\frac{k\varepsilon^2}{16}\right) \le \frac{\varphi}{2},$$

since, for $\varepsilon < 1/2$, we have $\frac{1}{2}\sum_{i=3}^{\infty}(\varepsilon/2)^i \le (\varepsilon/2)^3 \le \varepsilon^2/16$. The last step in the above inequality follows by substituting in the lower bound on the value of k. ■

The other direction we need follows in a similar fashion. We state the needed result without proof [LM00, Lemma 1] (which also yields better constants):

Lemma 23.3.2. *Let $Y_1, \ldots, Y_k$ be k independent random variables with $Y_i \sim \mathbf{N}(0,1)$. Let $Z = \sum_{i=1}^k Y_i^2/k$. For any $x > 0$, we have that*

$$\mathbb{P}\Big[Z \le 1 - 2\sqrt{x/k}\Big] \le \exp(-x) \qquad \textit{and} \qquad \mathbb{P}\Big[Z \ge 1 + 2\sqrt{x/k} + 2x/k\Big] \le \exp(-x).$$

For our purposes, we require that $\exp(-x) \le \varphi/2$, which implies $x = \ln(2/\varphi)$. We further require that $2\sqrt{x/k} \le \varepsilon$ and $2\sqrt{x/k} + 2x/k \le \varepsilon$, which hold for $k = 8\varepsilon^{-2}\ln\frac{2}{\varphi}$, for $\varepsilon \le 1$. We thus get the following result.

Corollary 23.3.3. *Let $\varphi \in (0,1)$, and $\varepsilon \in (0,1/2)$ be parameters, and let $k \ge \left\lceil \frac{8}{\varepsilon^2}\ln\frac{2}{\varphi}\right\rceil$ be an integer. Then, for k independent random variables $X_1, \ldots, X_k \sim \mathbf{N}(0,1)$, we have for $Z = \sum_i X_i^2/k$ that that $\mathbb{P}[1-\varepsilon \le Z \le 1+\varepsilon] \ge 1-\varphi$.*

Remark 23.3.4. The result of Corollary 23.3.3 is surprising. It says that if we pick a point according to the k-dimensional normal distribution, then its distance to the origin is strongly concentrated around $\sqrt{k}$. Namely, the normal distribution "converges" to a sphere, as the dimension increases. The mind boggles.

Lemma 23.3.5. *Let v be a unit vector in $\mathbb{R}^d$, then*

$$\mathbb{P}\Big[1-\varepsilon \le \|\mathsf{M}v\| \le 1+\varepsilon\Big] \ge 1 - \frac{1}{n^2}.$$

Proof: Observe that if for a number x, if $1-\varepsilon \le x^2 \le 1+\varepsilon$, then $1-\varepsilon \le x \le 1+\varepsilon$. As such, the claim holds if $1-\varepsilon \le \|\mathsf{M}v\|^2 \le 1+\varepsilon$. By Corollary 23.3.3, setting $\varphi = 1/n^2$, we need

$$k \ge 8\varepsilon^{-2}\ln(2/\varphi) = 8\varepsilon^{-2}\ln(2n^2) = 24\varepsilon^{-2}\ln n,$$

which holds for the value picked for k in Eq. (23.3). ■

23.3.3. All pairwise distances are preserved

Lemma 23.3.6. *With probability at least half, for all points* $\mathsf{p}, \mathsf{p}' \in P$*, we have that*

$$(1-\varepsilon)\left\|\mathsf{p}-\mathsf{p}'\right\| \leq \left\|\mathsf{M}\mathsf{p}-\mathsf{M}\mathsf{u}\right\| \leq (1+\varepsilon)\left\|\mathsf{p}-\mathsf{u}\right\|.$$

Proof: The key observation is that M is a linear operator. As such, let $v = (\mathsf{p}-\mathsf{p}')/\|\mathsf{p}-\mathsf{p}'\|$ be a unit vector, and observe that

$$\begin{aligned}
& (1-\varepsilon)\left\|\mathsf{p}-\mathsf{p}'\right\| \leq \left\|\mathsf{M}\mathsf{p}-\mathsf{M}\mathsf{p}'\right\| = \left\|\mathsf{M}(\mathsf{p}-\mathsf{p}')\right\| \leq (1+\varepsilon)\left\|\mathsf{p}-\mathsf{p}'\right\| \\
\iff \qquad & 1-\varepsilon \leq \left\|\mathsf{M}\frac{\mathsf{p}-\mathsf{p}'}{\|\mathsf{p}-\mathsf{p}'\|}\right\| \leq 1+\varepsilon \\
\iff \qquad & (1-\varepsilon)\left\|v\right\| \leq \left\|\mathsf{M}v\right\| \leq (1+\varepsilon)\left\|v\right\|.
\end{aligned}$$

The probability the later condition does not hold is at most $1/n^2$, by Lemma 23.3.5. As such, for all possible pairs of points, the probability of failure is $\binom{n}{2} \cdot \frac{1}{n^2} \leq 1/2$, as claimed. ■

We thus got the famous JL-Lemma.

Theorem 23.3.7 (The Johnson-Lindenstrauss Lemma). *Given a set P of n points in $\mathbb{R}^d$, and a parameter ε, one can reduce the dimension of P to $k = O(\varepsilon^{-2}\log n)$ dimensions, such that all pairwise distances are $1 \pm \varepsilon$ preserved.*

23.4. Even more on the normal distribution

The following is not used anywhere in the above, and is provided as additional information about the normal distribution.

Lemma 23.4.1. *Let $X \sim \mathbf{N}(0,1)$, and let $\sigma > 0$ and μ be two real numbers. The random variable $Y = \sigma X + \mu$ has the density function*

$$f_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right). \tag{23.4}$$

The variable Y has the **normal distribution** *with variance σ^2, and expectation μ, denoted by $Y \sim \mathbf{N}(\mu, \sigma^2)$.*

Proof: We have $\mathbb{P}[Y \leq \alpha] = \mathbb{P}[\sigma X + \mu \leq \alpha] = \mathbb{P}\left[X \leq \frac{\alpha-\mu}{\sigma}\right] = \int_{y=-\infty}^{(\alpha-\mu)/\sigma} f(y)\,\mathrm{d}y$, where $f(x) = \frac{1}{\sqrt{2\pi}}\exp\left(-x^2/2\right)$. Substituting $y = (x-\mu)/\sigma$, and observing that $\mathrm{d}y/\mathrm{d}x = 1/\sigma$, we have

$$\mathbb{P}[Y \leq \alpha] = \int_{x=-\infty}^{\alpha} f\left(\frac{x-\mu}{\sigma}\right)\frac{1}{\sigma}\,\mathrm{d}x = \frac{1}{\sqrt{2\pi}\sigma}\int_{x=-\infty}^{\alpha}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\mathrm{d}x,$$

as claimed.

As for the second part, observe that $\mathbb{E}[Y] = \mathbb{E}[\sigma X + \mu] = \sigma\,\mathbb{E}[X] + \mu = \mu$ and $\mathbb{V}[Y] = \mathbb{V}[\sigma X + \mu] = \mathbb{V}[\sigma X] = \sigma^2\,\mathbb{V}[X] = \sigma^2$. ■

Lemma 23.4.2. *Consider two independent variables $X \sim \mathbf{N}(0,1)$ and $Y \sim \mathbf{N}(0,1)$. For $\alpha, \beta > 0$, we have $Z = \alpha X + \beta Y \sim \mathbf{N}(0, \sigma^2)$, where $\sigma = \sqrt{\alpha^2 + \beta^2}$.*

Proof: Consider the region in the plane $H^- = \left\{(x, y) \in \mathbb{R}^2 \;\middle|\; \alpha x + \beta y \leq z\right\}$ – this is a halfspace bounded by the line $\ell \equiv \alpha x + \beta y = z$. This line is orthogonal to the vector $(-\beta, \alpha)$. We have that $\ell \equiv \frac{\alpha}{\sigma} x + \frac{\beta}{\sigma} y = \frac{z}{\sigma}$. Observe that $\left\|\left(\frac{\alpha}{\sigma}, \frac{\beta}{\sigma}\right)\right\| = 1$, which implies that the distance of ℓ from the origin is $d = z/\sigma$.

Now, we have

$$\mathbb{P}[Z \leq z] = \mathbb{P}[\alpha X + \beta Y \leq z] = \mathbb{P}[H^-] = \int_{p=(x,y)\in H^-} g(x, y)\,\mathrm{d}p,$$

see Eq. (23.2). Since, the two dimensional density function g is symmetric around the origin. any halfspace containing the origin, which its boundary is in distance d from the origin, has the same probability. In particular, consider the halfspace $T = \left\{(x, y) \in \mathbb{R}^2 \;\middle|\; x \leq d\right\}$. We have that

$$\begin{aligned}\mathbb{P}[Z \leq z] = \mathbb{P}[H^-] = \mathbb{P}[T] = \mathbb{P}[X \leq d] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{d} \exp\left(-\frac{x^2}{2}\right) \mathrm{d}x = \frac{1}{\sqrt{2\pi}} \int_{y=-\infty}^{z} \exp\left(-\frac{y^2}{2\sigma^2}\right) \frac{\mathrm{d}x}{\mathrm{d}y}\,\mathrm{d}y,\\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{y=-\infty}^{z} \exp\left(-\frac{y^2}{2\sigma^2}\right) \mathrm{d}y,\end{aligned}$$

by change of variables $x = y/\sigma$, and observing that $\mathrm{d}x/\mathrm{d}y = 1/\sigma$. By Eq. (23.4), the above integral is the probability of a variable distributed $\mathbf{N}(0, \sigma^2)$ to be smaller than z, establishing the claim. ■

Lemma 23.4.3. *Consider two independent variables $X \sim \mathbf{N}(\mu_1, \sigma_1^2)$ and $Y \sim \mathbf{N}(\mu_2, \sigma_2^2)$. We have $Z = X + Y \sim \mathbf{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$,*

Proof: Let $\widehat{X} \sim \mathbf{N}(0, 1)$ and $\widehat{Y} \sim \mathbf{N}(0, 1)$, and observe that we can write $X = \sigma_1 \widehat{X} + \mu_1$ and $Y = \sigma_2 \widehat{Y} + \mu_2$. As such, we have

$$Z = X + Y = \sigma_1 \widehat{X} + \sigma_2 \widehat{Y} + \mu_1 + \mu_2.$$

The variable $W = \sigma_1 \widehat{X} + \sigma_2 \widehat{Y} \sim \mathbf{N}(0, \sigma_1^2 + \sigma_2^2)$, by Lemma 23.4.2. Adding $\mu_1 + \mu_2$ to W, just shifts its expectation, implying the claim. ■

23.5. Bibliographical notes

The original result is due to Johnson and Lindenstrauss [JL84]. By now there are many proofs of this lemma. Our proof follows class notes of Anupam Gupta, which in turn follows Indyk and Motwani [IM98],

References

[IM98] P. Indyk and R. Motwani. *Approximate nearest neighbors: Towards removing the curse of dimensionality*. *Proc. 30th Annu. ACM Sympos. Theory Comput.* (STOC), 604–613, 1998.

[JL84] W. B. Johnson and J. Lindenstrauss. *Extensions of lipschitz mapping into hilbert space*. *Contemporary Mathematics*, 26: 189–206, 1984.

[LM00] B. Laurent and P. Massart. *Adaptive estimation of a quadratic functional by model selection*. *Ann. Statist.*, 28(5): 1302–1338, 2000.