Dimension Reduction and Surprises in High Dimensions II

Orestis Plevrakis

Before proving the JL lemma, we will see some alternative constructions of $\tilde{e}_i$'s, that rely on some useful properties of the multivariate Gaussian distribution.

1 Multivariate Gaussian Distribution

In your probability class, you saw that multivariate Gaussians $\mathcal{N}(\mu, \Sigma)$ are parametrized with their mean μ and their covariance matrix $\Sigma \succcurlyeq 0$. Here, we will only need the case $\mu = 0, \Sigma = I$. Remember that a random vector $X \sim \mathcal{N}(0, I)$ has coordinates independent and identically distributed according to the standard normal distribution $\mathcal{N}(0, 1)$.

1.1 Sampling uniformly from the unit sphere

We define $S^{\kappa-1} := \{x \in \mathbb{R}^\kappa : \|x\| = 1\}$. How can we sample a point uniformly at random from $S^{\kappa-1}$? A very efficient way is to first sample $X \sim \mathcal{N}(0, I)$ in $\mathbb{R}^\kappa$, and then normalize to get a unit vector $\hat{X} := X/\|X\|$. Why $\hat{X}$ is uniformly distributed on $S^{\kappa-1}$? Remember that since $X = (X_1, \ldots, X_\kappa)$ and X_ℓ's are independent, the probability density function of X at a point $x \in \mathbb{R}^\kappa$ is

$$f_X(x) = f_{X_1}(x_1) \cdot \cdots \cdot f_{X_\kappa}(x_\kappa) = \frac{1}{\sqrt{2\pi}} e^{-x_1^2/2} \cdots \frac{1}{\sqrt{2\pi}} e^{-x_\kappa^2/2} = \frac{1}{(2\pi)^{\kappa/2}} e^{-\|x\|^2/2}$$

Since the density depends only on the norm of $\|x\|$ and not on its direction, it follows that the distribution $\mathcal{N}(0, I)$ has no bias toward any particular direction, and so $X/\|X\|$ is uniformly distributed on $S^{\kappa-1}$.

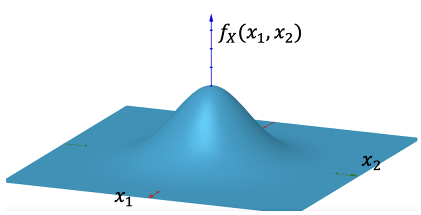

Figure 1: The case $\kappa = 2$. Observe the rotational symmetry of the graph.

First alternative construction of $\tilde{e}_i$'s. Previously, we chose $\tilde{e}_i$'s to be random vectors from $S^{\kappa-1}$ (but not uniformly distributed). Using arguments similar to the ones from the previous lecture, it can be shown that if we choose large enough $\kappa = O\left(\frac{\log d}{\epsilon^2}\right)$ and $\tilde{e}_i$'s independent and uniformly distributed on $S^{\kappa-1}$, then the requirements of the special case will be satisfied with high probability.

2 Norm of a Gaussian vector

Let $X \sim \mathcal{N}(0, I)$ in $\mathbb{R}^\kappa$. Then, $\mathbb{E}[\|X\|^2] = \mathbb{E}[\sum_{\ell=1}^{\kappa} X_\ell^2] = \sum_{\ell=1}^{\kappa} \mathbb{E}[X_\ell^2] = \kappa$. Thus, $\mathbb{E}[\|X/\sqrt{\kappa}\|^2] = 1$. The following simulation indicates that for large κ, the random vector $X/\sqrt{\kappa}$ is very close to the unit sphere with high probability:

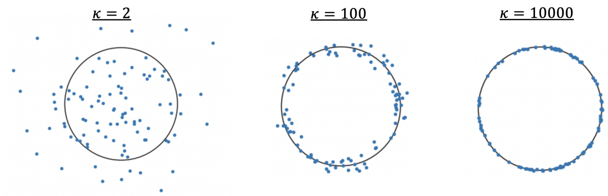

Figure 2: One hundred samples from $X/\sqrt{\kappa}$ for $\kappa = 2,\ 100$ and $100,000$. The second and third figure illustrate the distance from the unit sphere.

Why is this happening? Law of large numbers again!

$$\left\| \frac{X}{\sqrt{\kappa}} \right\|^2 = \frac{1}{\kappa} \sum_{\ell=1}^{\kappa} X_\ell^2 \approx 1$$

for large κ. This is made quantitative using the Chernoff-bound for the χ^2 distribution (a random variable Y has distribution χ^2 if $Y = X^2$, where $X \sim \mathcal{N}(0, 1)$). This Chernoff bound is

$$\mathbb{P}\left(\left| \frac{1}{\kappa} \sum_{\ell=1}^{\kappa} X_\ell^2 - 1 \right| \geq t \right) \leq 2 \exp\left(-\frac{\kappa t^2}{8} \right) \tag{1}$$

for all $t \in (0, 1)$. Thus, for large κ, if we want to normalize an $X \sim \mathcal{N}(0, I)$ to make it a unit vector, we can divide with $\sqrt{\kappa}$ instead of $\|X\|$ and we will make it unit up to a small error.

Second alternative construction of $\tilde{e}_i$'s. Combining the last comment with the first alternative construction, it seems natural to consider $\tilde{e}_i = \frac{1}{\sqrt{\kappa}} g_i$, where $g_i \sim \mathcal{N}(0, I)$ and the different g_i's are chosen independently. This construction indeed satisfies the requirements of the special case for large enough $\kappa = O\left(\frac{\log d}{\epsilon^2}\right)$. We will not prove this, as the proof of the general case (which we will see) implies this statement.

3 The General Case

We now prove the JL lemma. Let's remember the statement.

Theorem 1. *Let $n \leq poly(d)$ and $v_1, \ldots, v_n \in \mathbb{R}^d$. Let $\epsilon \in (0, 1)$. Then, there exists a $\kappa = O\left(\frac{\log d}{\epsilon^2}\right)$ and $\tilde{v}_1, \ldots, \tilde{v}_n \in \mathbb{R}^\kappa$ such that for all $i \neq j$,*

$$(1 - \epsilon)\|v_i - v_j\| \leq \|\tilde{v}_i - \tilde{v}_j\| \leq (1 + \epsilon)\|v_i - v_j\| \tag{2}$$

As you will see in the proof, without assuming $n \leq \text{poly}(d)$, we will get $\kappa = O\left(\frac{\log n}{\epsilon^2}\right)$.

Proof. Fix an $\epsilon \in (0,1)$. The dimension κ will be chosen in a bit. We will construct a dimension-reduction map $\phi : \mathbb{R}^d \to \mathbb{R}^\kappa$ and we will set $\tilde{v}_i := \phi(v_i)$. Since we know how to reduce the dimension of the standard basis, a natural map ϕ is

$$\phi(x) = x_1\tilde{e}_1 + \cdots + x_d\tilde{e}_d$$

We will choose the $\tilde{e}_i$ according to the second alternative construction (the other two constructions also work, but this one is the easiest to analyze). Observe that for an $x \in \mathbb{R}^d$, $\phi(x) = \frac{1}{\sqrt{\kappa}}Gx$, where $G \in \mathbb{R}^{\kappa \times d}$ is a random matrix with independent entries drawn from $\mathcal{N}(0,1)$. We start with the following lemma for ϕ.

Lemma 2. *Let $\delta \in (0,1)$. If $\kappa \geq 8\frac{\log(2/\delta)}{\epsilon^2}$, then for all $x \in \mathbb{R}^d$,*

$$\mathbb{P}\left((1-\epsilon)\|x\| \leq \|\phi(x)\| \leq (1+\epsilon)\|x\|\right) \geq 1-\delta \tag{3}$$

Note that the ϵ in the lemma is the ϵ we fixed in the beginning of the proof. From Lemma 2, we can get the theorem as follows: let $\delta := \frac{1}{100n^2}$ (this choice will be justified in a bit). Let $\kappa := \left\lceil 8\frac{\log(200n^2)}{\epsilon^2}\right\rceil$. Then, for any fixed pair $i \neq j$, by applying the lemma for $x \leftarrow v_i - v_j$, we get

$$\mathbb{P}\left((1-\epsilon)\|v_i - v_j\| \leq \|\phi(v_i) - \phi(v_j)\| \leq (1+\epsilon)\|v_i - v_j\|\right) \geq 1-\delta$$

where we used that ϕ is linear. This says that every fixed pair's distance is preserved with high probability. We want to know that with high probability all the distances are preserved, so we apply union bound:

$$\mathbb{P}\left(\forall i \neq j, \quad (1-\epsilon)\|v_i - v_j\| \leq \|\phi(v_i) - \phi(v_j)\| \leq (1+\epsilon)\|v_i - v_j\|\right) \geq 1 - \binom{n}{2}\delta \geq 0.99$$

Thus, with 99% probability over the choice of G, the function ϕ will preserve all distances up to error ϵ.

We now prove Lemma 2. Fix a nonzero $x \in \mathbb{R}^d$ (for $x = 0$ the statement trivially holds). Letting G_ℓ be the ℓ^{th} row of G, we have

$$\|\phi(x)\|^2 = \frac{1}{\kappa}\sum_{\ell=1}^{\kappa}(G_\ell \cdot x)^2 = \|x\|^2 \cdot \frac{1}{\kappa}\sum_{\ell=1}^{\kappa}\left(G_\ell \cdot \frac{x}{\|x\|}\right)^2$$

Let $Z_\ell := G_\ell \cdot \frac{x}{\|x\|}$. Since $G_\ell \sim \mathcal{N}(0, I)$ we have $Z_\ell \sim \mathcal{N}(0,1)$ (why?). Furthermore, since G_ℓ's are independent random vectors, the Z_ℓ's are independent random variables. From the Chernoff bound from the χ^2-distribution, we get that for all $\epsilon \in (0,1)$,

$$\mathbb{P}\left(\left|\frac{1}{\kappa}\sum_{\ell=1}^{\kappa} Z_\ell^2 - 1\right| \geq \epsilon\right) \leq 2\exp\left(-\frac{\kappa\epsilon^2}{8}\right)$$

Note that this bound is at most δ for $\kappa \geq 2\frac{\log(2/\delta)}{\epsilon^2}$. Thus, with probability at least $1-\delta$, $(1-\epsilon)\|x\|^2 \leq \|\phi(x)\|^2 \leq (1+\epsilon)\|x\|^2$ which gives

$$\sqrt{1-\epsilon}\|x\| \leq \|\phi(x)\| \leq \sqrt{1+\epsilon}\|x\|$$

Using that $1-\epsilon < \sqrt{1-\epsilon} < 1 < \sqrt{1+\epsilon} < 1+\epsilon$, we are done. □

We conclude with a geometric interpretation of ϕ.

3.1 Geometric Interpretation: Random Projection

Let $u_\ell := G_\ell/\sqrt{d}$ and observe that

$$\phi(x) = \sqrt{\frac{d}{\kappa}} \begin{pmatrix} u_1 \cdot x \\ \vdots \\ u_\kappa \cdot x \end{pmatrix}$$

From Section 2, we know that $u_1, \ldots, u_\kappa$ are nearly unit and nearly pairwise orthogonal. Let's assume for the moment that the last statement is exact, i.e., $u_1, \ldots, u_\kappa$ form an orthonormal basis of the subspace S that they span. Then, the map $x \mapsto (u_1 \cdot x)u_1 + \cdots + (u_\kappa \cdot x)u_\kappa$ projects the inputs on S. So, the first part of ϕ, i.e., $\Pi(x) = ((u_1 \cdot x), \ldots, (u_\kappa \cdot x))$ projects x to S and then keeps the coordinates of the projection (with respect to the basis of the u_i's). If we had used Π in place of ϕ, then each initial distance $\|v_i - v_j\|$ would either have decreased or remained the same. This is general fact: if we project a bunch of points on a subspace, then each distance will either decrease or stay the same (prove this to yourself!). However, a consequence of JL lemma is that (with high probability) Π shrinks all distances almost by the *same* factor: $\sqrt{\frac{\kappa}{d}}$, and thus by having in ϕ the scaling factor $\sqrt{\frac{d}{\kappa}}$ we approximately preserve them. Thus, up to a small error (remember that u_i's are not exactly orthonormal), $\phi(x)$ projects on a random κ-dimensional subspace, keeps the coordinates, and then rescales.