The Blurse of Dimensionality

Dr Huw Day

Jean Golding Institute, University of Bristol

Motivation and Background

The Curse of Dimensionality is a phenomena frequently referred to in mathematics, statistics and machine learning and refers to the spread of data as you increase the size of the feature space the data sits within. Contrary to intuition, having more information (i.e. features) is typically not optimal for constructing models. Intuitively appreciating this fact becomes harder when you are tasked with visualising high dimensional space. As my linear algebra lecturer put it to me once:

"The best way to visualise 9 dimensional space is to simply visualise n dimensional space and then let n=9."

Easier said than done!

A rather excitable sphere

A p-dimensional hypersphere of radius $R > 0$ in $\mathcal{R}^p$ is defined as:

$$B_p(R) := \{x = (x_1, \dots, x_p) \in \mathcal{R}^p : x_1^2 + \dots + x_p^2 \leq R^2\}. \quad (1)$$

and it can be shown that the volume of this hypersphere is:

$$vol(B_p(R)) = \int \dots \int_{x_1^2+\dots+x_p^2 \leq R^2} dx_1 \dots dx_p = \frac{\pi^{p/2}}{\Gamma(p/2+1)} R^p. \quad (2)$$

p	$\text{vol}(\text{B}_p)$	
1	2	
2	π	≈ 3.14
3	$\frac{4\pi}{3}$	≈ 4.19
4	$\frac{\pi^2}{2}$	≈ 4.93
5	$\frac{8\pi^2}{15}$	≈ 5.24
6	$\frac{\pi^3}{6}$	≈ 5.17
10	$\frac{\pi^5}{120}$	≈ 2.55
15	$\dots$	≈ 0.38

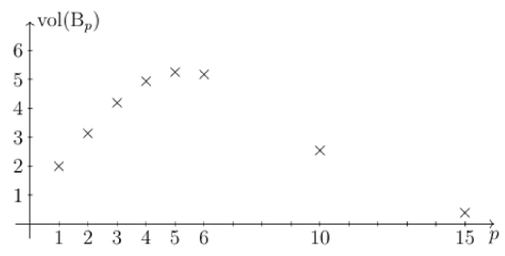

Figure 1: Comparison of the volumes of the hyperball in higher dimensions.

Figure 1: How does the volume of a p-dimensional hypersphere/hyperball of radius $R = 1$ vary for different dimensions. Original figure taken from [1].

We are going to see that how vectors and volumes behave inside high dimensional hyperspheres is...weird.

How can we think about high dimensional space?

In high dimensions, the following statements are true:

1. Any two vectors in the hypersphere are likely to be almost orthogonal
2. The volume of a hypersphere is concentrated close to its surface

1. Any two vectors in the hypersphere are likely to be almost orthogonal

Consider a p-dimensional unit radius hypersphere and consider two vectors $x, y \in \mathcal{R}^p$ on the surface of this sphere (i.e. they both have unit norm). Let $z := x \cdot y \in [-1, 1]$ be their dot product, what is the density function $f_p(z)$?

Because of rotational symmetry, it suffices to consider the case where $x = (1, 0, \dots, 0) \in \mathcal{R}^p$. Then: $z = x \cdot y = \sum_{i=1}^p x_i y_i = y_1$.

So the p.d.f of z is just the marginal of the first coordinate of y; y_1.

What does this look like? We know that because y lies on the surface of the unit radius, $p-1$ dimensional hypersphere which means $\sum_{i=1}^p y_i^2 = 1$.

So we want to know: *How likely is it that the first coordinate has value z, given that the whole vector lies on the surface of that sphere?*

Because every point on the hypersphere is equally likely, if we fix $z = y_1$ then we have the equation: $\sum_{i=2}^p y_i^2 = 1 - z^2$ which tells us that the remaining vectors $y_2, \dots, y_p$ lie in the $p-2$ dimensional hypersphere of radius $\sqrt{1-z^2}$ embedded in $\mathcal{R}^{p-1}$.

This means the marginal density $f_p(z)$ is proportional to the surface area of the of a hypersphere of radius $\sqrt{1-z^2}$ in $\mathcal{R}^{p-1}$.

The surface area of a $(k-1)$ dimensional hypersphere in radius R is proportional to: $A \propto r^{k-2}$ which tells us that:

$$f_p(z) = C_p(1-z^2)^{\frac{p-3}{2}}, z \in [-1, 1] \quad (3)$$

(where C_p is our normalising constant that ensures that our density integrates to 1. It can be shown that $C_p = \frac{\Gamma(\frac{p}{2})}{\sqrt{\pi}\Gamma(\frac{p-1}{2})}$).

Since $z \in [-1, 1]$, as $p \to \infty$ we see that: $f_p(z) \to 1_{z=1}$. This means that in the large p limit, the dot product of two randomly sampled vectors converges to 1 which is equivalent to saying they are orthogonal.

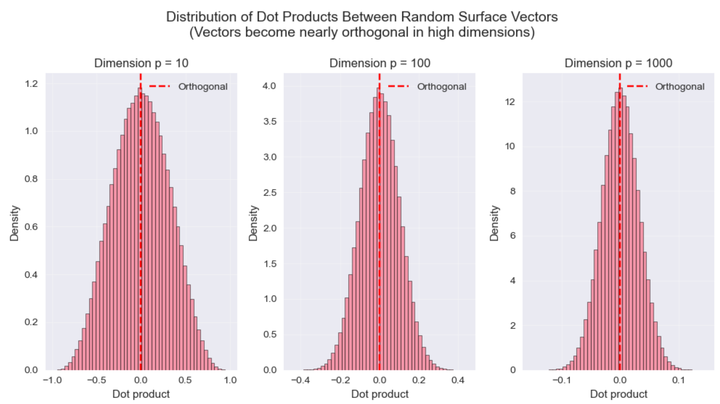

Figure 2: If you randomly sample 10000 vectors on the surface of a unit radius hypersphere, what is the distributions of their dot products? Note that as the dimension of the p hypersphere grows larger, the distribution of the dot product gets more concentrated, implying vectors get closer and closer to being orthogonal with high probability [2].

2. The volume of a hypersphere is concentrated close to its surface

"If one peels an orange in high dimensions, almost nothing is left."

Fix $0 < \varepsilon < 1$ small and consider a random vector $x \in B_p$:

$$Prob(||x|| > 1-\varepsilon) = \frac{vol(\{x \in B_p : ||x|| > 1-\varepsilon\})}{vol(B_p)}$$
$$\dots = \frac{vol(B_p) - vol(\{x \in B_p : ||x|| \leq 1-\varepsilon\})}{vol(B_p)} = 1 - \frac{B_p(1-\varepsilon)}{B_p(1)}$$
$$\dots = 1 - (1-\varepsilon)^p \longrightarrow_{p\to\infty} 1.$$

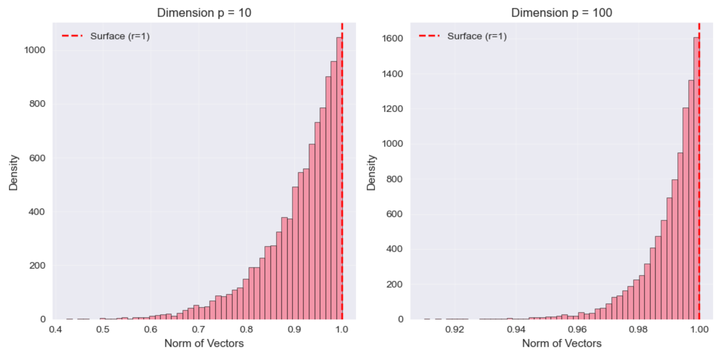

Figure 3: If you randomly sample 10000 vectors in a hypersphere of unit dimenison, what is the distribution of their norms? The higher the dimension p, the more vectors are closer to the surface [2].

Blessing or a Curse?

When considering high dimensional data analysis problems, we frequently run into the curse of dimensionality whereby data gets sparse and it becomes increasingly difficult to sample the densities of distributions in high dimensions.

But it's not all bad news! As a consequence of this we are blessed with the fact that many random vectors and random matrices behave close to deterministically, making it much more routine to sample smooth, 1 dimensional functions of high dimensional vectors.

References + Acknowledgments

The content of this poster is inspired by the notes and recordings of Roland Speicher's graduate course "High Dimensional Analysis: Random Matrices and Machine Learning" (https://rolandspeicher.com/lectures/course-on-high-dimensional-analysis-random-matrices-and-machine-learning-summer-term-2023/) [1].

Figures 2 and 3 were generated by Fahd Abdelazim from the Ask-JGI data science helpdesk (ask-jgi@bristol.ac.uk) code available on GitHub (https://github.com/HuwWDay/BlurseOfDimensionality) [2].