Source: Unsplash

The Blessing of Dimensionality: Why the Curse of Dimensionality is a Two-Sided Coin

The benefits of warped & wacky spaces

Andre Ye Jun 14, 2020 · 5 min read ★

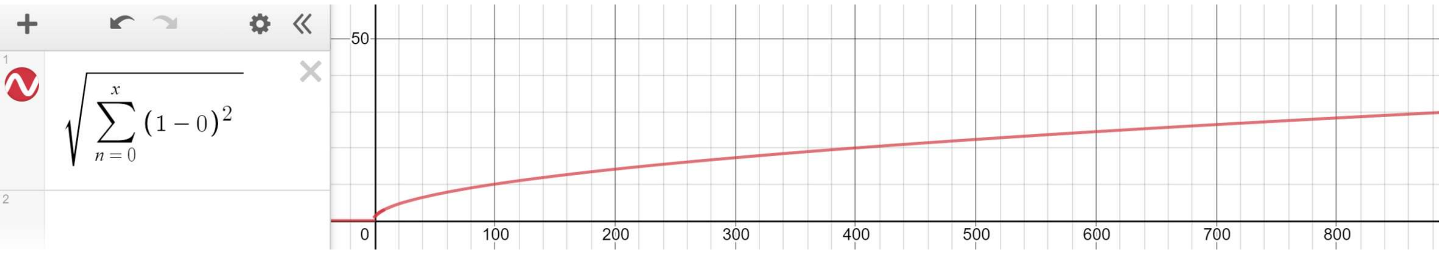
manipulating and analyzing data in high-dimensional spaces that don't occur in low-dimensional spaces.

Consider, for instance, how the concept of distance is distorted in a high dimensional plane. The formula for distance between two points *a* and *b* defined by coordinates (*x*, *y*, …) each with *n* dimensions is as follows:

$$d(a,b) = \sqrt{\sum_{i=1}^{n}(a_i - b_i)^2}$$

You can find a proof for the formula here.

When the formula is plotted, where *x* is the number of dimensions and *y* the distance between the origin and a point (1, 1, 1, …), one can see that distance gradually will inevitably peak, or, more likely, tend to one single value.

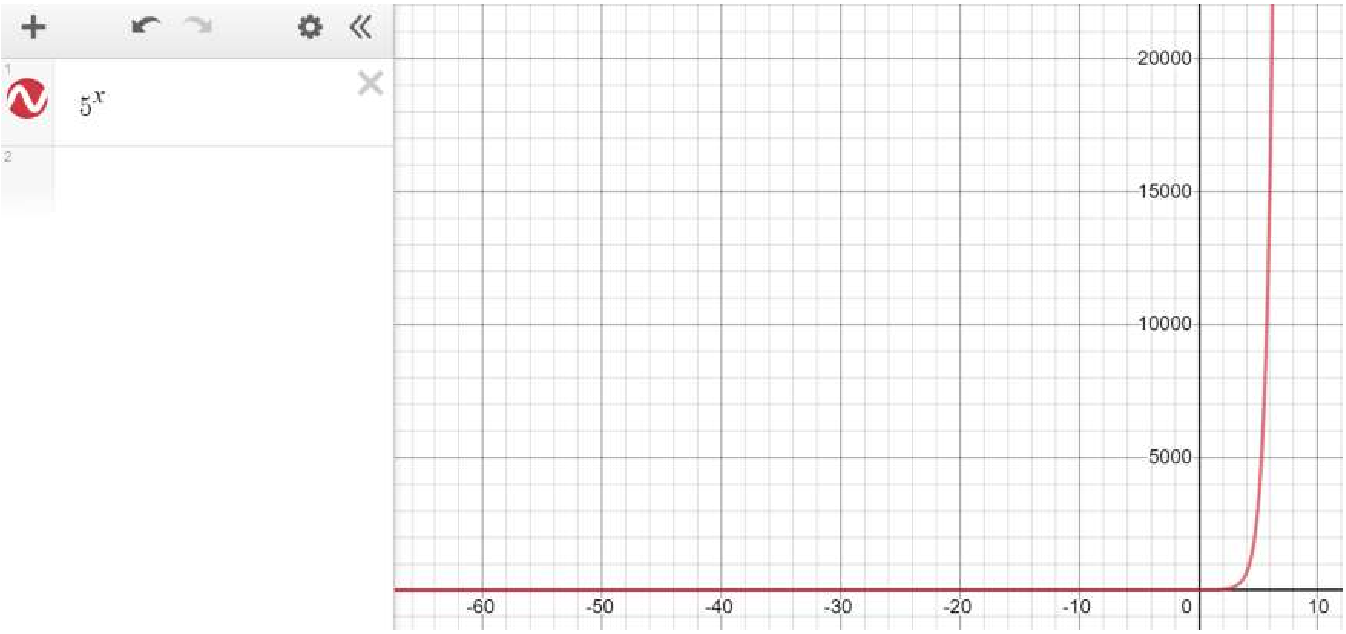


Graphed in Desmos.

This is evidence of the nature of distance in high-dimensional spaces — as dimensionality increases, the importance, or value, of any single one-dimensional line diminishes, as can be seen with the diminishing returns on the *y*-axis. This is to be expected, considering how quickly volume and the possibilities of points grows as dimensionality increases.

This is also the reason why a hypersphere's volume tends towards zero as the dimensionality increases — since a sphere is defined as consisting of all points that are one radius' distance in Euclidean distance away from a center point, the number of dimensions increases but the distance gradually tapers off. Hence, however, growing (or stay constant, if the side length is 1).

Let us consider a hypercube in two dimensions (a square) with side lengths of five units. There are, then $5^2 = 25$ units. A similar hypercube in three dimensions (a cube) has 125 units. From there, it skyrockets. The power of exponents is really very incredible — just within ten dimensions, the hypercube already has a hypervolume of 9,765,625 units. Adding an addition dimension to a space expands the current space by a huge magnitude, so it should be no surprise that a miniscule one-dimensional distance has diminishing value.



Graphed in Desmos.

The mathematics of higher dimensions is odd, but in machine learning, it poses an even larger risk. Because the volume of a high-dimensional space is so unimaginably enormous and the number of data points is never anywhere even comparable to that volume, high-dimensional data is often very sparse. Consider, for instance, a 150 by 150 pixel image — a very common dimension size to work with in machine learning. This image then has 150 times 150 times 3 (three values per pixel to specify RGB — red, green, and blue values) equals 67,500 dimensions.

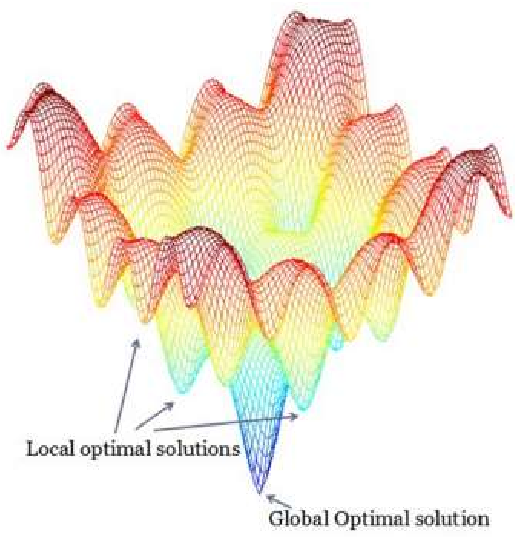
Along with, say, ten thousand other images, each image exists as one point in the feature space, with one value for each of the 67,500 dimensions. Because of the sheer enormousness of the space and because distance is so *small* in high-dimensional spaces, the data will be incredibly sparse. This causes models to overfit to the data, spaces.

This is why convolutional neural networks have become so popular with images — instead of relying on a point-in-space model so many other algorithms use, it uses simple data-altering layers to transform series of matrices. However, a huge part of neural networks is *gradient descent* — the updating of the parameters (weights, biases, etc.) in a way that will optimize the network's path to a reduction in error. This requires mapping out what errors form for a certain combination of the millions of trainable parameters.

Yet this is one example of the *blessing of dimensionality*. In fact, what was a weakness with the curse of dimensionality can be repurposed as a blessing in other contexts.

> The blessing of dimensionality and the curse of dimensionality are two sides of the same coin.

The goal of a neural network is to reach the *global minima* of the error space; that is, to find the perfect set of parameters that yields the lowest error possible. In order to reach that global minima, it starts out at some point on the error space and incrementally moves with each training step in what it believes is the right direction. It turns out that in the error space, there are many local minima — dips in the error space that are minima but not the lowest in the entire error space.



Source. Image free to share.

neural network training. If it doesn't prevent the network from finding a global minima, it can, at the very least, make training a lot longer.

In high dimensional space, however, distance is warped and not proportionately large, as demonstrated earlier. This turns out to be a huge advantage with local minima, whose once seemingly-global dips are rendered small bumps in the surface any reasonable optimizer could overcome. On the other hand, the global minima would be noticeable enough for the network to converge at that point. So, in certain cases, *increasing the dimensionality* can be a blessing with better performance and faster training time. Of course, this may be cancelled out by overfitting of parameters, so decisions need to be carefully crafted based on experimentation and knowledge.

There are other applications of the Blessing of Dimensionality in other areas of machine learning and in other fields, from mathematics to physics. What's important to understand is that in machine learning, *everything is a trade-off.* The Curse of Dimensionality may be less cursed than most people would believe.

If you enjoyed,

You may also enjoy some of my other articles on mathematics and machine learning.

Not All Infinite Sets Are the Same Size
Cantor's Famous Diagonal Argument
medium.com

All the Distributions You Need to Know
Binomial, Bimodal, Poisson, + More
medium.com

By Towards Data Science

Every Thursday, the Variable delivers the very best of Towards Data Science: from hands-on tutorials and cutting-edge research to original features you don't want to miss. Take a look.

Get this newsletter You'll need to sign in or create an account to receive this newsletter.

Machine Learning AI Data Science Mathematics Artificial Intelligence