The Curse of Dimensionality

Robert B. Heckendorn
University of Idaho

February 6, 2022

1 The Curse of Dimensionality

There are a bunch of effects attributed to higher dimensional space. Here are the two most important ones that come immediately to mind:

1. The space gets really big as dimensions go up and not in the way we think it does. Part of this is even the units of volume change. If you were a physicist you would remark that the volume of the 2D square was in meters squared and in 3D space was meters cubed, which is not the same units. Making it a bit hard to do an apples to apples comparison. But we can do comparisons like ask how many k dim unit spheres (balls) can we pack into a k dim unit cube, which for most dimensions is actually greater than 1. For example, we can get 22 unit spheres in a 9 dimensional unit cube. How does this vast size affect us? It means it gets HARDER, VERY FAST, TO SAMPLE FROM ALL OF K DIMENSIONAL SPACE for large k. In order to get a meaningful number of samples might take an exponential number of samples. Without enough samples you wonder what key examples are you missing what patterns will you not discover? Sampling also tends to have bias, like what is easy to sample.

2. It turns out that as the number of dimensions get large many irrelevant dimensions are included in any one distance measurement. It means most of the signal tends to be wiped out by noise.

3. Then there is the "concentration effect of distance". For many measures such as L_i, where L_2 is Euclidean distance, the distances between points in your data tend to average out for large numbers of dimensions. Meaning all points look very similar (equally distant from each other). So much so that for many cases the points are statistically indistinguishable. This is a property of having many dimensions of data clouding your distance measure. The result is it is hard to conclude anything by similarity. There are things you can do to reduce this effect, but they don't always apply and only reduce it and not eliminate it.

I dont claim to be an expert but that is my take on it.