Conceptualizing Curse of Dimensionality with Parallel Coordinates

G Devi , Charu Chauhan and Sutanu Chakraborti
Department of Computer Science and Engineering
Indian Institute of Technology Madras,
Chennai-600036, Tamil Nadu, India
{devsjee, charu.chauhan.8888}@gmail.com, sutanuc@cse.iitm.ac.in

Abstract

We report on a novel use of parallel coordinates as a pedagogical tool for illustrating the non-intuitive properties of high dimensional spaces with special emphasis on the phenomenon of Curse of Dimensionality. Also, we have collated what we believe to be a representative sample of diverse approaches that exist in literature to conceptualize the Curse of Dimensionality. We envisage that the paper will have pedagogical value in structuring the way Curse of Dimensionality is presented in classrooms and associated lab sessions.

1 Introduction

Humans find it hard to imagine high dimensional spaces since they are exclusively adapted to a world in three dimensions (3D). In practice, though immersed in a 3D world, humans are more often experienced to living in two dimensional (2D) world, and this can be appreciated from the way we rely on 2D maps to localize ourselves. This is unlike fishes which experience the 3D space of ocean completely. Techniques like Flatland trick (Abbott 2006), require us to construct a four dimensional (4D) picture from many 3D projections by imagining how we are able to construct a 3D image in brain from multiple 2D projections. In spite of this being a great introduction to imagining hyperspace, it is quite challenging for anyone to extrapolate the ideas to realistic high dimensional spaces. Students in Artificial Intelligence (AI) and Machine Learning (ML) frequently encounter high dimensional design spaces in the form of multivariate data and complex models with large number of parameters. The phenomenon called Curse of Dimensionality (COD)(Bellman 1961) which is unique to high dimensional spaces could have profound effect on the design of search algorithms in AI. It is also one of the major challenges for parameter estimation of learning algorithms in ML.

In this paper, we have attempted to collate what we believe to be a representative sample of diverse perspectives that exist in literature for illustrating COD. These perspectives are founded on either mathematical conclusions or statistical comparisons. We will see in the following section that while these approaches provide analytically useful tools, they still leave a lot to imagination. This is because of our fundamental impairment in conceiving higher dimensional spaces in a way that perceptually grounds its abstractions.

Inselberg's novel contribution of Parallel Coordinates (Inselberg 1985) makes some significant headway in overcoming this limitation. We have created a novel use of Parallel Coordinates as a pedagogical tool for visualizing the properties of high dimensional spaces and found it to be more convincing in conveying the non-intuitive phenomena of COD to students.

2 Perspectives on High Dimensional Spaces

In this section, we have presented the different perspectives on COD in a sequence that will be of pedagogical value. Firstly, there is an exponential increase in the volume of space spanned by a hypercube as we go to higher dimensions. Secondly, the volume of hypersphere approaches zero as the number of dimensions increase. Another apparently non-intuitive observation is that as we go to higher dimensions, the hypercubes become spiky in shape. These observations have been elaborated upon in the following discussion.

Observation 1 Richard Hamming gives a beautiful introduction to n dimensional spaces in his book (Hamming 2003).

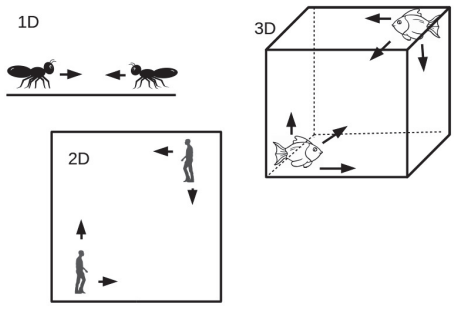

Figure 1: Ants moving in a line could meet each other more frequently than two men walking on the ground. Two fishes swimming in an ocean or fish tank have even fewer chances of meeting each other due to more number of degrees of freedom available to them than men or ants.

Figure 1 is in line with Hamming's introduction and highlights the fact that the chance of two living beings meeting each other decreases with increase in the degrees of freedom available to them for movement in their living space. This effect is due to an exponential increase in the space (or volume) enclosed by a hypercube as the number of dimensions increase.

Volume of a hypercube of edge length $2r$ in n dimensions, V_{cn}, is given by

$$V_{cn}(2r) \ = \ (2r)^n \tag{1}$$

It can be seen from the above equation that the volume enclosed by a hypercube increases exponentially with increasing values of n.

Observation 2 Unlike hypercubes, hyperspheres exhibit a curious behavior. Volume of a hypersphere in n dimensions, $V_{sn}(r)$, is given by :

$$\frac{\pi^{n/2} \ r^n}{\Gamma(\frac{n}{2}+1))} \tag{2}$$

Using the relation $\Gamma(n) = (n-1)!$, we get

$$V_{sn}(r) = \frac{\pi^{n/2} \ r^n}{(\frac{n}{2})!} \tag{3}$$

In the above equation, the $n!$ term in the denominator increases rapidly in comparison to the numerator. This is so because factorial function outgrows the exponential function after a particular value of n, depending on the value of base that we use in exponential functions. An unit hypersphere grows in volume up to five dimensions and then begins to shrink.

One motivation to study the behaviour of hyperspheres is its application in range queries. Range queries involve identifying the objects located within a particular distance from the given query. In general, a bounding rectangle around the query is used as an approximation for range queries. As we move to higher dimensions, for a given distance threshold, the volume of the bounding rectangle keeps increasing exponentially whereas the volume of the hypersphere approaches zero.

Another point to note is that most of the volume of a hypersphere is in a narrow annulus as explained below. Consider a circle inscribed within a square as in Figure 2(a). The area of the inscribed circle will be a constant fraction of that of square. Similarly, for a sphere, the volume will be a constant fraction of that of the circumscribing hypercube (see Figure 2(b)). Therefore, the volume of a hypersphere in n dimensions will be a constant fraction of the volume of the circumscribing hypercube. This constant depends on n and equals $\frac{\pi^{n/2}}{(n/2)!}$ which can be derived from Equation 3. Let C_n denote this constant that is dependent on n.

Volume of hypersphere of radius r in n dimensions

$$= C_n \ r^n \tag{4}$$

Volume of hypersphere of radius $r \ (1-\epsilon)$ in n dimensions

$$= C_n \ [r \ (1-\epsilon)]^n \tag{5}$$

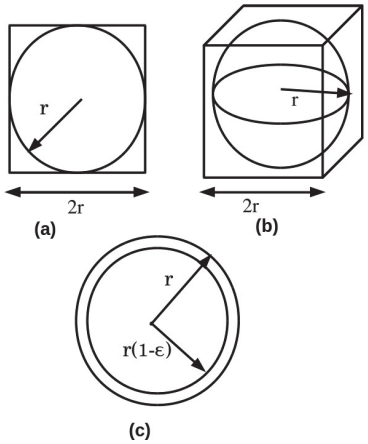

Figure 2: Most of the volume of a sphere is in a narrow annulus

where $0 < \epsilon < 1$.

The difference in volume between these two hyperspheres given by Equation 6 will give the volume concentrated in the region outer to the hypersphere of radius $r \ (1-\epsilon)$ and inner to the hypersphere of radius r (see Figure 2(c)).

$$C_n \ r^n - C_n \ [r \ (1-\epsilon)]^n \ = \ C_n \ r^n \ [1-(1-\epsilon)^n] \tag{6}$$

$$C_n \ r^n - C_n \ [r \ (1-\epsilon)]^n \ = \ C_n \ r^n \tag{7}$$

Even for very small values of ϵ, the term $(1-\epsilon)^n$ in Equation 6 tends to 0 as n tends to ∞. This means that in higher dimensions, almost all of the volume of the hypersphere is near the surface and there is negligible volume in the interior.

Observation 3 It is a hard to imagine fact that hypercubes become spiky in their shape in high dimensions. Consider a square of unit edge length placed at same origin as a circle of unit radius shown in Figure 3. The maximum distance between any points within the unit square is equal to the length of the diagonal which is $\sqrt{2}$. In the unit circle, the maximum distance between any two points is the diameter which is 2 units. Clearly, the corners of the square lie within the circle. When $n \ = \ 4$, the diagonal of hypercube is of length $\sqrt{4 * (1)^2} \ = \ 2$ units. The diameter of the hypersphere also remains 2 units which means that the corners of the hypercube touch the surface of hypersphere in four dimensions. In higher dimensions, the corners of the hypercube extend outside the hypersphere and hence becomes spiky in shape as can be seen in Figure 3. A hypercube has almost no volume at the centre. Entire volume is contained in the corners of the hypercube in higher dimensions.

Observation 4 We can derive the ratio of volume of an inscribed hypersphere of radius r to the volume of the circumscribing hypercube of edge length $2r$ from Equations 2 and 1.

$$\frac{V_{sn}}{V_{cn}} = \frac{\pi^{(n/2)}}{2^n \ (n/2)!} \tag{8}$$

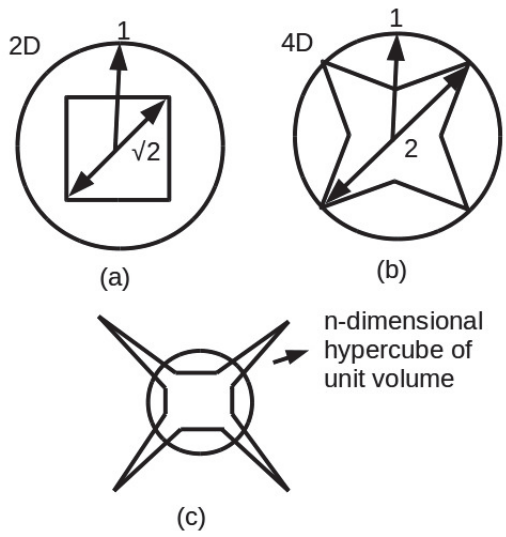

Figure 3: Adapted from (Hopcroft and Kannan 2014) illustrating a sphere enclosing a cube in 2,4 and n dimensions

As n goes to infinity, the volume of the hypersphere becomes insignificant relative to that of the hypercube. This implies the fact that almost the entire high dimensional volume is far away from the center or in other words, near the corners of the hypercube.

2.1 Manifestations of COD

We now discuss the manifestation of COD on nearest neighbour search and training sample size required by learning algorithms. These manifestations are directly related to the observations made in the previous section.

On Nearest Neighbor Search The exponential increase in volume of hypercube affects the nearest neighbour search algorithms in higher dimensions. The manifestation of COD on nearest neighbor search can also be illustrated using the idea of a hypercubical neighborhood (Hastie et al. 2005) as shown in Figure 4. Given a query point, the expected edge length of the hypercubical neighborhood containing it such that it covers a fraction r of the total observations is $r^{\frac{1}{p}}$ where p is the number of dimensions . As we go to higher dimensions, this hypercubical neighborhood becomes very large. For example, in ten dimensions, to cover 1% of the observations, it is necessary to search 63% of the range of each input variable which is a very large search space for nearest neighbor search. This drives home the fact that it is no longer possible to limit the number of candidates for distance calculation by pruning the search space to a smaller bounding box around the query point

On Sample Size The same authors (Hastie et al. 2005) have also used sampling density as a measure to explain the exponential increase in the number of training samples needed for machine learning algorithms as we move from lower to higher dimensions. The sampling density is proportional to $N^{\frac{1}{p}}$ where N is the sample size. Thus, if $N_1 = 100$ represents a dense sample for a single parameter model, then $N_{100} = 100^{10}$ is the sample size required for the same sampling density with 10 inputs which is very large. Also, since more data points move towards the surface of the sphere, extrapolation is needed instead of interpolation making the task of prediction more difficult.

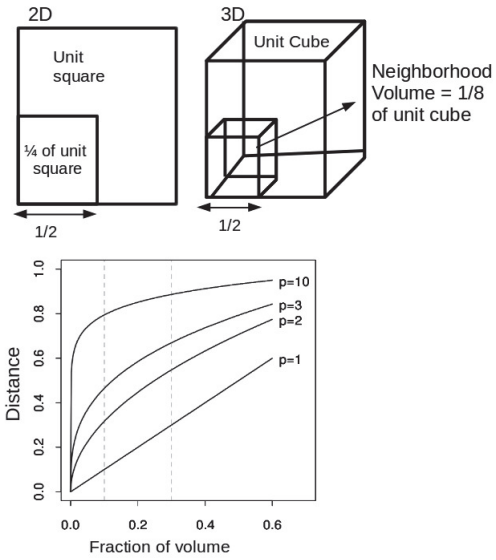

Figure 4: Adapted from (Hastie et al. 2005) to illustrate the exponential increase of search space in high dimensions. The graph shows the edge length of the subcube needed to capture a fraction of the volume of the data for different dimensions p. In ten dimensions we need to cover 80% of the range of each coordinate to capture 10% of the data.

3 Parallel Coordinates as a Pedagogical Tool for Conceptualizing COD

In order to make the phenomenon of COD easier to grasp we have created visualization for some of the discussed perspectives through a novel use of parallel coordinates. Parallel coordinates is an interesting topic of research in itself. However, for understanding the concepts explained in this paper, it would suffice to know only the fundamentals discussed in Section 3.1. We have used version 2.2 of XDAT, a free Parallel coordinates software, for our illustrations.

3.1 Parallel Coordinates (PC)

Parallel Coordinates(PC) is a novel contribution by Alfred Inselberg (Inselberg 1985) for visualizing

1. Multivariate data
2. High dimensional geometry

To visualize a dataset of n dimensions, n parallel lines are drawn on the plane which are typically vertical and equally spaced as shown in Figure 5.

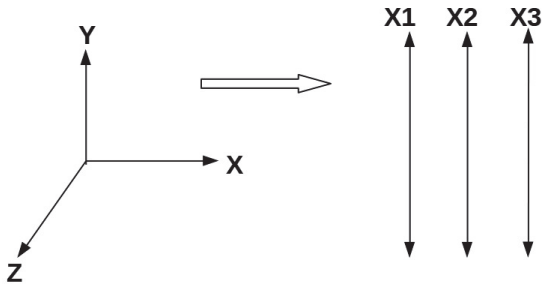

Figure 5: The three orthogonal axes in Cartesian coordinates become three parallel lines in Parallel coordinates.

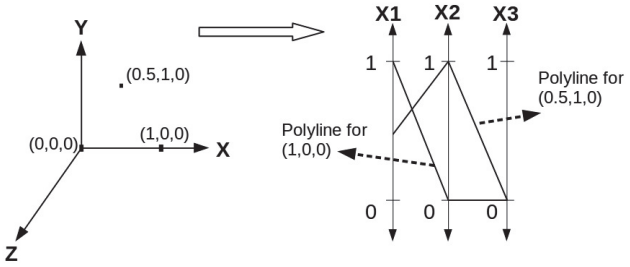

Figure 6: A point in Cartesian coordinates becomes a polyline in Parallel coordinates.

Point to Line Duality in Parallel Coordinates A point in n dimensional space is represented as a polyline with i^{th} coordinate of the point on the i^{th} parallel axis. An example is shown in Figure 6.

A line in an n dimensional space is represented by the point of intersection of the set of all polylines corresponding to the infinite number of points on the line in n dimensional space. An example is shown in Figure 7.

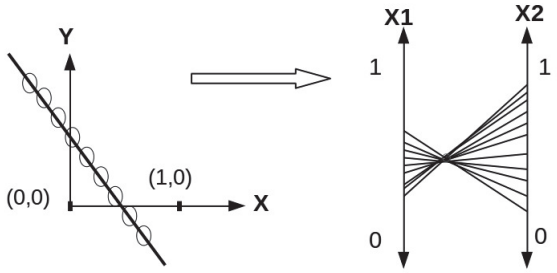

Figure 7: A line in Cartesian coordinates becomes a point (the point of intersection of all polylines) in Parallel coordinates.

Plotting of Hypercube and Hypersphere in PC A n dimensional hypercube on Parallel coordinates is represented by plotting the polylines corresponding to the corners of the hypercube. In Figure 8, the square is represented by the four polylines representing the four corners. A circle in Parallel coordinates is represented by plotting the polylines corresponding to the infinite number of points on the circumference of the circle and can be seen from Figure 8. A n dimensional hypersphere is represented in Parallel coordinates by $n-1$ copies of a circle having the same radius and center as the hypersphere. The Parallel coordinate plot for a sphere is shown in Figure 8.

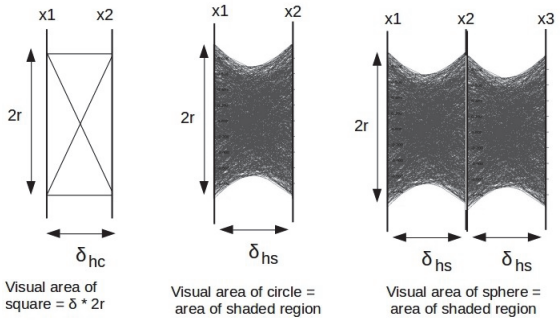

Figure 8: Parallel coordinate plots of Square, Circle and Sphere

3.2 Visualization of COD on Parallel Coordinates

Visual Area In order to illustrate the shrinking and exponentially increasing volumes of hypersphere and hypercube respectively, we introduce the novel concept of *Visual Area*. Visual area of a set of polylines is defined as the area of envelope of the set of polylines, i.e., area of the polygon formed by the maximum and minimum of the plotted coordinate values on each axis in the Parallel coordinates plot. The visual areas corresponding to a square, circle and sphere are as explained in Figure 8.

In the following discussion, we use δ to denote the inter-axis distance in Parallel coordinate plots. By placing a constraint on the distance between the parallel axes, it is possible to make the visual area proportional to the volume enclosed by the n dimensional object. The formulation is explained below.

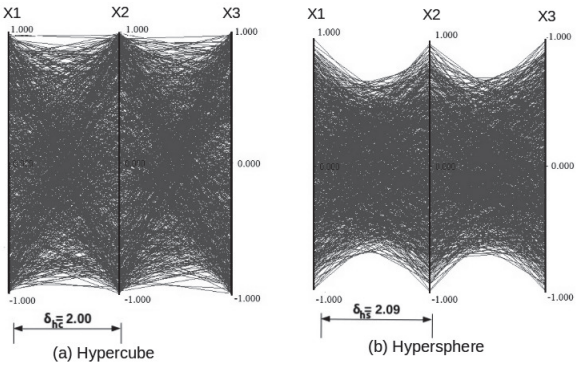

Figure 9: Visual areas of cube and sphere in 3 dimensions

Hypercube Let δ_{hc} denote the distance between axes for the parallel coordinate plot of hypercube of edge length $2r$.

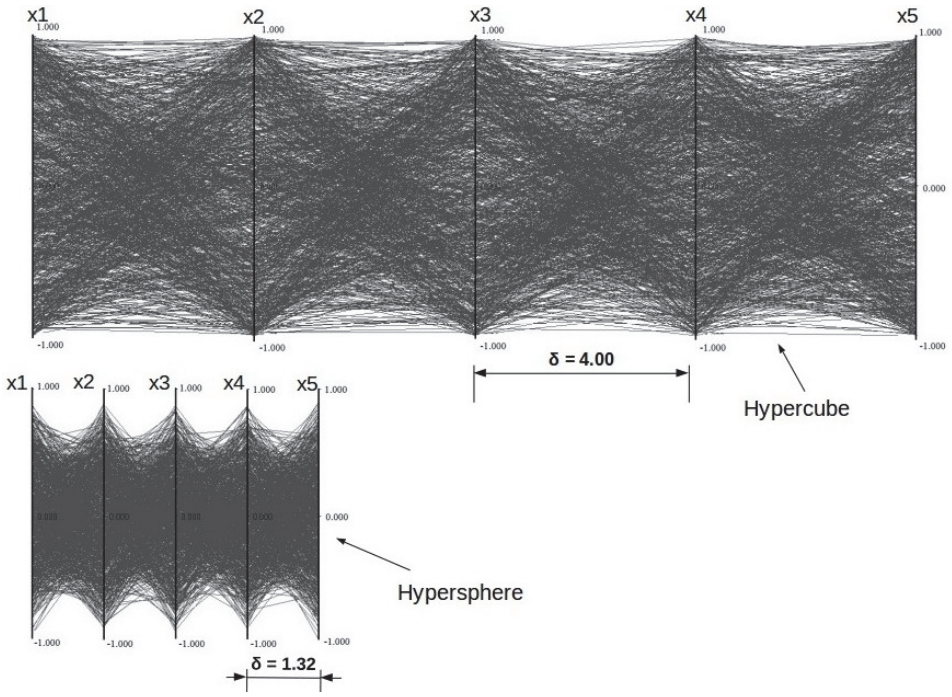

Figure 10: Visual areas of cube and sphere in 5 dimensions

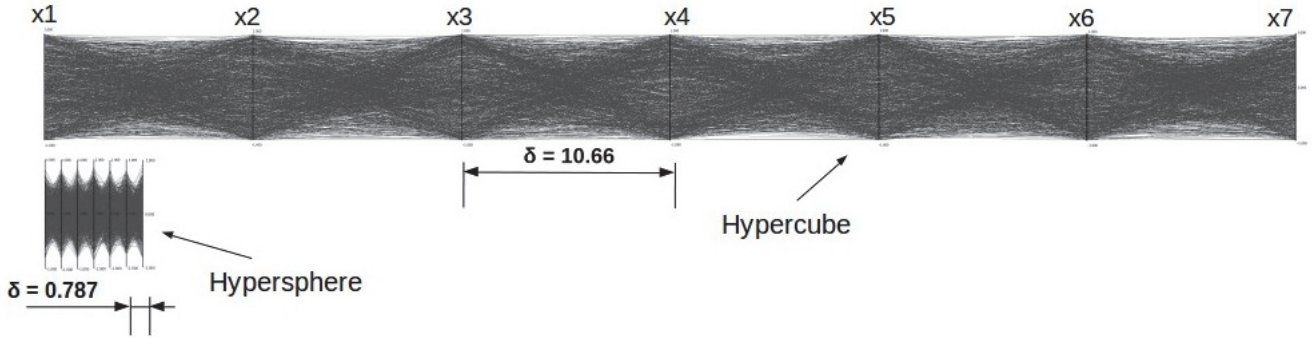

Figure 11: Visual areas of cube and sphere in 7 dimensions

In 2 dimensions, we need the visual area to be proportional to the area of the square as shown in below equation.

$$2r * \delta_{hc} \propto (2r)^2 \tag{9}$$

In 3 dimensions, we need the visual area to be proportional to the volume of the cube.

$$2r * 2\delta_{hc} \propto (2r)^3 \tag{10}$$

Generalizing the above idea to n dimensions,

$$2r * (n-1)\, \delta_{hc} \propto (2r)^n \tag{11}$$

$$\delta_{hc} \propto \frac{(2r)^{n-1}}{(n-1)} \tag{12}$$

Taking the proportionality constant as 1, δ_{hc} becomes equal to $\frac{(2r)^{n-1}}{n-1}$. The rapid increase in the visual areas in Figures 9 to 11 is analogous to the exponential increase in the volume of hypercubes.

Hypersphere Let δ_{hs} denote the distance between axes for the parallel coordinates plot of hypersphere. The visual area of the hypersphere of n dimensions is $n-1$ times the visual area of a circle of the same radius. From Figure 8 we can say that the visual area of a circle of radius r is ρ ($0<\rho<1$) times the visual area of a square of edge length $2r$. Proceeding in the same manner and using Equation 3, we get

$$\delta_{hs} \propto \frac{\pi^{n/2}\, r^{n-1}}{2\rho\, \Gamma(\frac{n}{2}+1)\, (n-1)} \tag{13}$$

Taking the proportionality constant as 2ρ, and using the relation $\Gamma(n) = (n-1)!$, we get δ_{hs} equal to $\frac{\pi^{n/2}}{(n/2)!}\ \frac{r^{n-1}}{n-1}$. As $n \to \infty$, it can be seen from Equation 13 that $\delta_{hs} \to 0$. This is in line with the shrinking of volume of a hypersphere and can be appreciated from the shrinking of inter-axis distances in Figures 9 to 11.

Ratio of Inter-axis Distances Explaining COD Due to the constraint placed on inter-axis distances, the visual areas obtained will be proportional to the volumes. Hence, we can use the ratio of inter-axis distances or ratio of visual areas instead of the ratio of volumes to conceptualize COD.

$$\frac{\delta_{hs}}{\delta_{hc}} = \frac{\pi^{(n/2)}}{2^{n-1}\ (n/2)!} \tag{14}$$

In Equation 14, for large values of n, the numerator is very small compared to the denominator indicative of the already discussed fact that the proportion of volume occupied by a hypersphere inscribed within a hypercube becomes insignificant at high dimensions.

From the Figures 9 to 11, we can observe that the visual area for hypercubes increases very rapidly as the number of dimensions increase. On the other hand, the visual area of hyperspheres shrinks and approaches zero in higher dimensions which can be appreciated easily from Figures 9 to 11. Hence, by placing appropriate constraints on the inter-axis distances, visual area can be made a surrogate for volume thereby helping to visualize the phenomenon of COD.

4 Topic Structuring

We propose the following ordering of observations for presenting COD in class.

1. Motivate the need for studying the phenomena of COD (Section 2.1 on Manifestations of COD).
2. Introduce the exponential increase in volume of a hypercube with increase in dimensions. (Observation 1 in Section 2).
3. Introduce the shrinking volume of hypersphere with increasing dimensions (Observation 2 in Section 2).
4. Discuss the moving of points towards the corners of hypercube (Observation 3 in Section 2).
5. Compare and contrast the behavior of the volume of a hypersphere to that of a hypercube (Observation 4 in Section 2).
6. Illustrate the above observation on the ratio of volume of hypersphere to hypercube using Visual Area in Parallel Coordinates (Section 3.2).

5 Conclusion

The diverse perspectives on the Curse of Dimensionality is indicative of the effects that this phenomenon can have while working with high dimensional data. For example, we have seen from the discussion on COD that the data sample becomes sparse in high dimensions. This has its effect in Machine Learning in the process of model choice. Choosing a model for the given data depends on the number of parameters to estimate for the model and the number of training examples available. Similarly, the perspective that almost all the volume is near the surface of the hypersphere, in other words, towards the corners of the hypercube, has an adverse effect on the search algorithms based on locality of search. More such effects of COD can be seen in practice for the other perspectives also.

Many works exist in literature on using Parallel coordinates for analyzing data. Our work is novel in that it provides a way to visualize the search space itself. By equating visual area to volume, we could effectively see the COD being manifested in high dimensions.We believe that our approach to introduce, explain and illustrate the phenomena of Curse of Dimensionality will have pedagogical value in conveying the idea to the readers.

6 Acknowledgments

The authors would like to thank Shubhranshu Shekhar and Abhijit Sahoo, Teaching Assistants for the course titled *Memory Based Reasoning in Artificial Intelligence*, for the useful discussions which helped in framing the idea of Visual Area.

References

Abbott, E. A. 2006. *Flatland: A romance of many dimensions*. Oxford University Press.

Bellman, R. 1961. Curse of dimensionality. *Adaptive control processes: a guided tour. Princeton, NJ*.

Hamming, R. R. 2003. *Art of Doing Science and Engineering: Learning to Learn*. CRC Press.

Hastie, T.; Tibshirani, R.; Friedman, J.; and Franklin, J. 2005. The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer* 27(2):83–85.

Hopcroft, J., and Kannan, R. 2014. Foundations of data science.

Inselberg, A. 1985. The plane with parallel coordinates. *The Visual Computer* 1(2):69–91.