

CSE 455/555 Spring 2013 Homework: Dimensionality and Component Analysis Solutions

Jason J. Corso

Computer Science and Engineering

SUNY at Buffalo

jcorso@buffalo.edu

Dimensionality

As we discussed in the class, nearest neighbor algorithms are very intuitive for data in low-dimensions. They assume we should be able to find a good number of examples close to a given point x and use the examples to estimate the class of x (these are its neighbors).

However, in high-dimensions, this assumption breaks down. This makes learning hard and is an instance of the curse of dimensionality. Consider inputs distributed in a d -dimensional unit hypercube $[0, 1]^d$. Suppose we define a hypercubical neighborhood around a point x to capture a fraction r of the unit volume. The point and the hypercubical neighborhood are assumed to be fully inside the unit hypercube without loss of generality. Denote the length of each side of the hypercubic neighborhood as l .

1. Derive l for $d = 1$ and $d = 2$.
2. Show that $l = r^{1/d}$. What is l with $d = 100$ for $r = 0.1$? In other words, to capture 10 percent of the data in the 100-dimensional space, what is that length of a side of the hypercubic neighborhood?
3. For $d = 10100$, how big on average do you need to make l in order to capture 1% of the data? How big for 10%?

Solutions

1. For $d = 1$, our neighborhood is just a line of length 1, so to capture ratio r of that neighborhood we need a "cube" that is a line of length r , so $l = r$.
For $d = 2$, our neighborhood is a 1 by 1 square, so to capture some ratio r of the area we need to solve $rA = l^2$, where A is our area (specifically $A = 1 \cdot 1 = 1$). This gives us $l = \sqrt{r}$.
2. We can generalize our equation from the $d = 2$ case, to yield $rA = l^d$. Since A always equals $1^d = 1$, this resolves simply into $l = r^{1/d}$.
To solve for the particular case mentioned, we just plug in our r and d values, yielding $l = 0.1^{1/100} \approx 0.977$.
3. Again we just plug in some numbers and solve here. For $r = 0.01$, we get $l = 0.01^{1/10100} \approx 0.9995$. For $r = 0.1$ we get $l = 0.1^{1/10100} \approx 0.9998$.