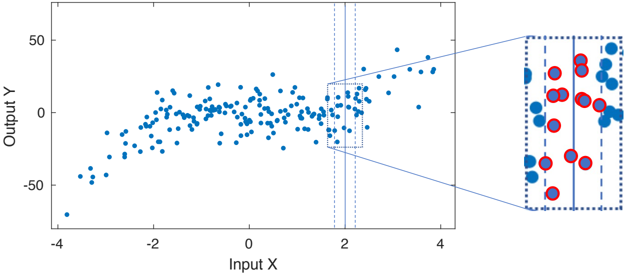
Section 2C. Curse of Dimensionality

Statistics for Data Science

Victor M. Preciado, PhD MIT EECS
Dept of Electrical & Systems Engineering
University of Pennsylvania
preciado@seas.upenn.edu

Recap: Local averaging

Local averaging: In principle, we can estimate the regression function using a local empirical average...



Are we done with this course?

Curse of Dimensionality

Basic Idea: We can use local averaging with a small r to estimate the regression function f. However, for learning problems where $p \gg 1$, this approach is unfeasible because of the *curse of dimensionality* (CoD), which is the fact that, for p large, points in the dataset tend to be far away from each other

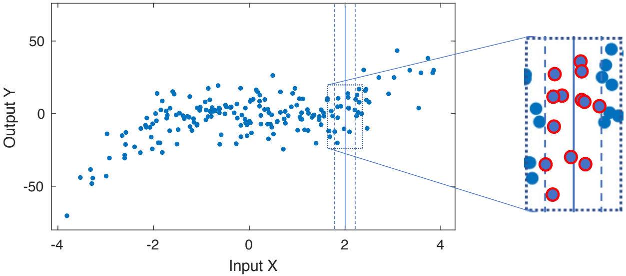


Figure: Several samples are within the margin $X = 2 \pm r$.

Curse of Dimensionality (cont.)

We illustrate the CoD in the bullets below:

- For $p = 1$, we consider N points randomly distributed over $S_1 = [-1, 1] \subset \mathbb{R}^1$ and a value $\mathbf{x} \in S_1$. How many points are in the set $\mathcal{D}_r(\mathbf{x})$ for r small? The number of points $|\mathcal{D}_r(\mathbf{x})|$ is a random variable with mean

$$\mathbb{E}\left[|\mathcal{D}_r(\mathbf{x})|\right] = \frac{N}{2} 2r = Nr$$

- For $p = 2$, we consider N points in the circle of radius one, denoted by $S_2 \subset \mathbb{R}^2$, and a value $\mathbf{x} \in S_2$. What is the expectation of the number of points in the set $\mathcal{D}_r(\mathbf{x})$ for r small?

$$\mathbb{E}\left[|\mathcal{D}_r(\mathbf{x})|\right] = \frac{N}{\pi 1^2} \pi r^2 = Nr^2$$

Curse of Dimensionality (cont.)

We can keep asking the same question for $p \gg 1$. We define the p-dimensional hypersphere of radius r centered at zero as

$$S_p(r) = \{\mathbf{x} \in \mathbb{R}^p : \ \|\mathbf{x}\| \leq r\}$$

The volume of $S_p(r)$ is given by $\mathrm{Vol}(S_p(r)) = k_p r^p$, where k_p is a constant that depends on p

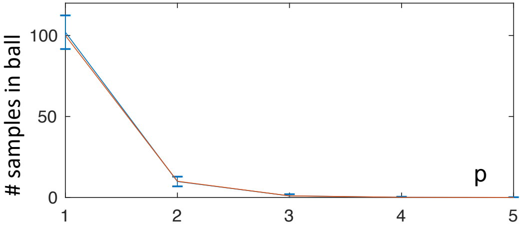
- For a generic p, we consider N points in $S_p(1)$ and a particular $\mathbf{x} \in S_p$. What is the expectated number of points in the ball $\mathcal{D}_r(\mathbf{x})$ for r small?

$$\mathbb{E}\left[|\mathcal{D}_r(\mathbf{x})|\right] = \frac{N}{k_p 1^p} k_p r^p = N r^p$$

  In other words, if we have N random points in a hypersphere $S_p(1)$ of volume $k_p 1^p$, we have $N r^p$ points in a small hypersphere of volume $k_p \varepsilon^p$.

Curse of Dimensionality: Simulations

Simulations: We consider $N = 1,000$ random points in $S_p(1)$. We then count the number of points within a distance $r = 0.1$ from an $\mathbf{x} \in \mathbb{R}^p$ when $p = 1, 2, \ldots$ Repeat 100 times for each value of p and compute the empirical mean and standard deviation for each value of p



- The number of points close to $\mathbf{x}$ decay exponentially as p increases; therefore, local averaging with a small r is not a reliable method to estimate $f(\mathbf{x})$ for large p.