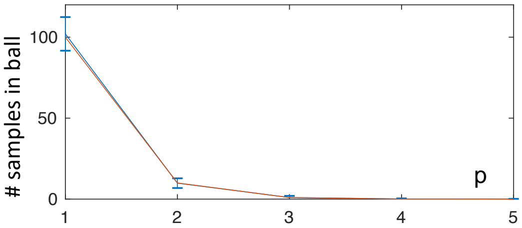
Section 2D. Parametric Models

Statistics for Data Science

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

Curse of Dimensionality: Recap

CoD: When the number of inputs p in your problem is large, the CoD does not allow you to use local averaging around a point $\mathbf{x}$ to estimate the regression function $f(\mathbf{x})$



- How can we beat the CoD? Parametric models...

Parametric Models

To overcome the curse of dimensionality, we use parametric models. For example, *linear models* of the form:

$$f_L(\mathbf{x}; \beta) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$$

Reminder: $\mathbf{x} = (x_1, \ldots, x_p)^\intercal$; hence x_i are not samples, but the components of the vector $\mathbf{x}$

- We can estimate the parameters β_i from a dataset of observations $\mathcal{D}$. We will denote the estimated parameters as $\widehat{\beta}_i$ and the estimated linear function as

$$\widehat{f}_L(\mathbf{x}; \beta) = \widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \widehat{\beta}_2 x_2 + \cdots + \widehat{\beta}_p x_p$$

- The main reason why this model does not suffer from CoD is that to estimate the parameters β_i and to make predictions we make use of the whole dataset $\mathcal{D}$, instead of a small number of samples.

Parametric Models (cont.)

- During this course, we will see a variety of parametric models: Linear models, polynomial models, regression trees, support vector machines, etc.
- For example, for $p = 1$, a polynomial model takes the form

$$\widehat{f}_L(X; \alpha) = \widehat{\alpha}_0 + \widehat{\alpha}_1 X + \widehat{\alpha}_2 X^2 + \ldots + \widehat{\alpha}_d X^d$$

where d is the degree of the polynomial model. The larger the d, the more *flexible* our model is, since it can learn functions with more and more oscillations.