Section 2B. Regression Function

Statistics for Data Science

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

Regression Function: Theory

- If we had access to $f_{Y|X}$ explicitly (which is typically impossible), one could compute the regression function f as follows

$$f(\mathbf{x}) = \mathbb{E}[Y|X = \mathbf{x}] = \int_{y=-\infty}^{y=+\infty} y\, f_{Y|X}(y|\mathbf{x})\, dy$$

  Notice that the result of the integral is a function of $\mathbf{x}$ alone.

- One can prove that the regression function is the solution to the following optimization

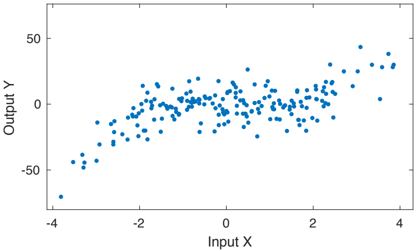
$$f(\cdot) = \arg\min_{g(\cdot)} \mathbb{E}\left[(Y - g(X))^2\right]$$

  where the minimization is over all functions g and the expectation is called the ***mean squared error (MSE)***.

Regression Function: Practice

Practical setup: In practice, we do not know explicitly the conditional PDF $f_{Y|X}$; however, we can sample data points from the additive model. Hence, the statistical learning problem can be posed as follows:

- **Given** a dataset $\mathcal{D} = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)\}$ where $(\mathbf{x}_i, y_i)$ are random samples drawn independently from the additive model (notice that $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^\mathsf{T} \in \mathbb{R}^p$ is a p-dimensional vector)
- **Find** an estimate of the regression function f. We will denote our estimate by $\widehat{f}$

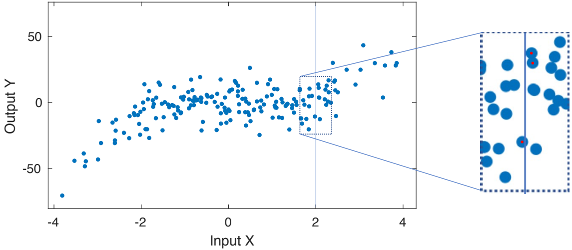


Regression Function: Practice (cont.)

- We might be tempted to approximate $f(\mathbf{x})$ using the *empirical* conditional mean, i.e.,

$$f(\mathbf{x}) = \mathbb{E}[Y|X = \mathbf{x}] \approx \frac{1}{|\mathcal{D}_{\mathbf{x}}|} \sum_{i \in \mathcal{D}_{\mathbf{x}}} y_i, \text{ where } \mathcal{D}_{\mathbf{x}} = \{i \in \{1, \ldots, N\} : \mathbf{x}_i = \mathbf{x}\}$$

  In other words, we should look for points in $\mathcal{D}$ for which the input $\mathbf{x}_i$ is *exactly* $\mathbf{x}$
- However, it is very unlikely we will find any samples in $\mathcal{D}$ for which $\mathbf{x}_i = \mathbf{x}$ exactly. See an example in the figure below for $\mathbf{x} = 2$



Regression Function: Practice (cont.)

- **Numerical solution**: Consider a window of width r and use the approximation

$$f(\mathbf{x}) = \mathbb{E}[Y|X = \mathbf{x}] \approx \frac{1}{|\mathcal{D}_r(\mathbf{x})|} \sum_{i \in \mathcal{D}_r(\mathbf{x})} y_i, \text{ where } \mathcal{D}_r(\mathbf{x}) = \{i \in \{1, \ldots, N\} : \|\mathbf{x}_i - \mathbf{x}\| \leq r\}$$

$\|\mathbf{x}_i - \mathbf{x}\|$ is the distance between points $\mathbf{x}_i$ and $\mathbf{x}$. In other words, we should look for points for which the input $\mathbf{x}_i$ is r-close a given vector $\mathbf{x}$