Section 4B. Bayes Classifier and Local Averaging

Statistics for Data Science

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

Classification Problem: Bayes Classifier

- **Bayes optimal classifier**:
  - Assuming that we knew the conditional class probabilities, the *Bayes Optimal Classifier* (BOC) $C(\mathbf{x})$ is given by

    $$C(\mathbf{x}) = \arg\max_k p_k(\mathbf{x})$$

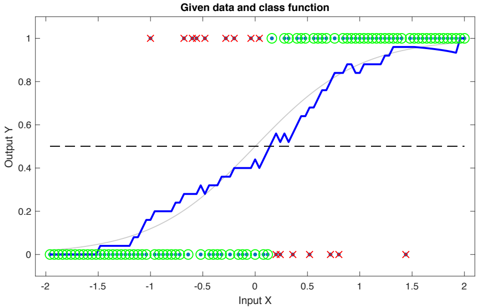
    In other words, given an input $\mathbf{x}$, the BOC assigns the label k that maximizes the conditional class probability. We will show later on that the BOC minimizes the classification error rate...
  - *Example*: In the previous slide, since $p_1(x) \geq p_0(x)$ for $x \geq 0$, the Bayes optimal classifier is

    $$C(x) = \begin{cases} 1 & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}$$

  - In practice, *we do not know the conditional class probabilities*; in this situation, how can we find the Bayes optimal classifier?...

Classification Problem: Local Averaging

- **Numerical results**:
  - We can numerically estimate the conditional class probability $\widehat{p}_1(x)$ from data using local averaging (see blue curve in figure below)
  - Using the estimate $\widehat{p}_1(x)$, we can find an approximation of the Bayes optimal classifier $\widehat{C}(x) = 1$ for $\widehat{p}_1(x) \geq \widehat{p}_0(x) \iff \widehat{p}_1(x) \geq 1/2$; $\widehat{C}(x) = 0$ otherwise



Curse of Dimensionality

- ***Curse of dimensionality***: However, when $\mathbf{x}$ is a high-dimensional vector, local averaging does not provide a reliable estimate of the conditional class probabilities. To overcome this issue, we will use structured models to represent $p_k(\mathbf{x})$...

Penn Engineering