Section 4J. Classification Errors

Statistics for Data Science

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

Classification Errors

- Consider a classification problem with these ingredients:
  - Two classes $\mathcal{C} = \{0, 1\}$
  - A training dataset $\mathcal{D}_{\mathsf{Tr}} = \{(\mathbf{x}_i, y_i)\}$ with $y_i \in \{0, 1\}$
  - A classifier $C : \mathbb{R}^p \to \{0, 1\}$

- The output of a classifier can be categorized as follows:
  - If $C(\mathbf{x}_i) = y_i = 1$, we say that point i is a *true positive* (TP)
  - If $C(\mathbf{x}_i) = y_i = 0$, we say that point i is a *true negative* (TN)
  - If $C(\mathbf{x}_i) = 1$ and $y_i = 0$, we say that point i is a *false positive* (FP)
  - If $C(\mathbf{x}_i) = 0$ and $y_i = 1$, we say that point i is a *false negative* (FN)

Classification Errors (cont.)

- The total number of errors made by your classifier is equal to the sum of FPs and FNs, which can be counted as:

$$|\mathrm{FP}| = \sum_{i=1}^{N} \delta\left(C\left(\mathbf{x}_i\right) - 1\right) \delta\left(y_i\right)$$

$$|\mathrm{FN}| = \sum_{i=1}^{N} \delta\left(C\left(\mathbf{x}_i\right)\right) \delta\left(y_i - 1\right)$$

  where $\delta\left(x\right)$ is called a Dirac's delta function, defined as $\delta\left(x\right) = 1$ for $x = 0$ and $\delta\left(x\right) = 0$ for $x \neq 0$.
- The *classification error rate* is defined as

$$\mathrm{Err}_{\mathrm{Tr}} = \frac{|\mathrm{FP}| + |\mathrm{FN}|}{N}$$

Classification Errors: LDA Theory

Analysis of errors in LDA with one input and two classes:

- $\Pr(Y = 0) = \pi_0$ and $\Pr(Y = 1) = \pi_1$
- $\Pr(X = x | Y = k) = f_k(x)$ with $k \in \{0, 1\}$

How much is the *expected* classification error rate of the Bayes optimal classifier $C(x)$? Remember,

$$C(x) = \begin{cases} 0 & \text{if } \pi_0 f_0(x) > \pi_1 f_1(x) \iff x < \gamma \\ 1 & \text{if } \pi_1 f_1(x) \geq \pi_0 f_0(x) \iff x \geq \gamma \end{cases}$$