Tarea n°5

En esta tarea, consideramos las notaciones siguientes.

— La muestra es una sucesión de variables $(X_1, Y_1), \ldots, (X_n, Y_n)$ donde $X_i \in \mathbb{R}_n$ y $Y_i \in \{0, 1\}$

— Un clasificador es una función $g : \mathbb{R}^d \to \mathbb{R}$.

— El calsificador óptimo se denota g^*.

— Denotamos L^*_X (o simplemente L^*) el error de Bayes : $L^* = \inf_g \mathbb{P}(g(X) = Y)$.

1. Sea $\phi : \mathbb{R}^d \to \mathbb{R}^k$ con $k \geq 0$, una función medible. Denotamos L^*_X el error de Bayes para (X, Y) y $L^*_{\phi(X)}$ el error de Bayes para $(\phi(X), Y)$. Mostrar que

$$L^*_{\phi(X)} \geq L^*_X$$

2. Sea $X' \in \mathbb{R}^k$ independiente de (X, Y). Mostar que $L^*_{(X,X')} = L^*_X$.

3. Decidimos permitir una tercera posibilidad para las funciones de decisión. El espacio de las posibilidades será entonces $\{0, 1, \text{"indeciso"}\}$. Se mide la utilidad de una tal función a través de $\mathbb{P}(g(X) = \text{"indeciso"})$ y $\mathbb{P}(g(X) \neq Y | g(X) \neq \text{"indeciso"})$. Para una $0 < c < 1/2$, definimos

$$g_c(x) = \begin{cases} 1 & \text{si } \eta(x) > 1/2 + c \\ 0 & \text{si } \eta(x) < 1/2 - c \\ \text{“}indeciso'' & \text{si no} \end{cases}$$

Demuestra que para cada función de decisión tal que

$$\mathbb{P}(g(X) = \text{"indeciso"}) \leq \mathbb{P}(g_c(X) = \text{"indeciso"})$$

entonces

$$\mathbb{P}(g(X) \neq Y | g(X) \neq \text{"indeciso"}) \geq \mathbb{P}(g_c(X) \neq Y | g_c(X) \neq \text{"indeciso"}).$$

Interpretar el resultado.