Teorema: *La función de decisión*

$$g^*(\mathbf{x}) = \begin{cases} 1 & si\ \eta(\mathbf{x}) > 1/2 \\ 0 & si\ no \end{cases}$$

es el clasificador Bayes, es decir, para cualquier clasificador $g : \mathbb{R}^p \to \{0, 1\}$, *se cumple que*

$$L^* := \mathbb{P}\{g^*(\mathbf{X}) \neq Y\} \leq \mathbb{P}\{g(\mathbf{X}) \neq Y\}.$$

Dem: Para cualquier clasificador g y todo $\mathbf{x} \in \mathbb{R}^d$ se cumple que

$$\begin{aligned}
&\mathbb{P}\{g(\mathbf{X}) \neq Y | \mathbf{X} = \mathbf{x}\} \\
&\quad = 1 - \mathbb{P}\{g(\mathbf{X}) = Y | \mathbf{X} = \mathbf{x}\} \\
&\quad = 1 - (\mathbb{P}\{Y = 1, g(\mathbf{X}) = 1 | \mathbf{X} = \mathbf{x}\} + \mathbb{P}\{Y = 0, g(\mathbf{X}) = 0 | \mathbf{X} = \mathbf{x}\}) \\
&\quad = 1 - \left(\mathbb{1}_{\{g(\mathbf{x})=1\}} \mathbb{P}\{Y = 1 | \mathbf{X} = \mathbf{x}\} + \mathbb{1}_{\{g(\mathbf{x})=0\}} \mathbb{P}\{Y = 0 | \mathbf{X} = \mathbf{x}\}\right) \\
&\quad = 1 - \left(\mathbb{1}_{\{g(\mathbf{x})=1\}} \eta(\mathbf{x}) + \mathbb{1}_{\{g(\mathbf{x})=0\}} (1 - \eta(\mathbf{x}))\right).
\end{aligned}$$

Por tanto,

$$\begin{aligned}
&\mathbb{P}\{g(\mathbf{X}) \neq Y | \mathbf{X} = \mathbf{x}\} - \mathbb{P}\{g^*(\mathbf{X}) \neq Y | \mathbf{X} = \mathbf{x}\} \\
&\quad = \eta(\mathbf{x})[\mathbb{1}_{\{g^*(\mathbf{x})=1\}} - \mathbb{1}_{\{g(\mathbf{x})=1\}}] + (1 - \eta(\mathbf{x}))[\mathbb{1}_{\{g^*(\mathbf{x})=0\}} - \mathbb{1}_{\{g(\mathbf{x})=0\}}] \\
&\quad = (2\eta(\mathbf{x}) - 1)[\mathbb{1}_{\{g^*(\mathbf{x})=1\}} - \mathbb{1}_{\{g(\mathbf{x})=1\}}] \geq 0.
\end{aligned}$$

$\square$

Teorema: *El vector* $\mathbf{a}$ *proporcional a* $\mathbf{S}_{\mathbf{X}}^{-1}(\bar{\mathbf{x}}_0 - \bar{\mathbf{x}}_1)$ *maximiza la separación* $\mathcal{S}$.

Dem: Basta maximizar en $\mathbf{a}$ la expresión

$$\mathcal{S}^2(\mathbf{a}) = \frac{(\mathbf{a}'\bar{\mathbf{x}}_0 - \mathbf{a}'\bar{\mathbf{x}}_1)^2}{\mathbf{a}'\mathbf{S}_{\mathbf{X}}\mathbf{a}}.$$

Se aplica el siguiente resultado (ver Johnson y Wichern 2007):

> Lema de maximización: *Sea* $\mathbf{B}$ *una matriz definida positiva y* $\mathbf{d}$ *un vector* $d \times 1$*. Entonces, para un vector* $\mathbf{x} \neq 0$ *arbitrario, se cumple que*
>
> $$\max_{\mathbf{x}\neq 0} \frac{(\mathbf{x}'\mathbf{d})^2}{\mathbf{x}'\mathbf{B}\mathbf{x}} = \mathbf{d}'\mathbf{B}^{-1}\mathbf{d}$$
>
> *y este máximo se alcanza en* $\mathbf{x} = c\,\mathbf{B}^{-1}\mathbf{d}$ *para cualquier constante* $c \neq 0$.

Tomando $\mathbf{d} = \bar{\mathbf{x}}_0 - \bar{\mathbf{x}}_1$, tenemos que el máximo de $\mathcal{S}^2$ se alcanza en $c\,\mathbf{S}_{\mathbf{X}}^{-1}(\bar{\mathbf{x}}_0 - \bar{\mathbf{x}}_1)$ para cualquier $c \neq 0$. El máximo es precisamente el cuadrado de la distancia de Mahalanobis $(\bar{\mathbf{x}}_0 - \bar{\mathbf{x}}_1)'\mathbf{S}_X^{-1}(\bar{\mathbf{x}}_0 - \bar{\mathbf{x}}_1)$.

$\square$