Minimality of Bayes classifier with respect to bayes risk

October 4, 2017

Let χ be an arbitrary state space and let X and Y be two random variables taking value in χ and $\{0,1\}$ respectively.

Given a function $f : \chi \mapsto \{0,1\}$ we define the risk function $R(f)$ as follows:

$$R(f) = \mathbb{P}(f(X) \neq Y)$$

And we define the bayes risk :

$$R^* = \inf\{R(f)\}$$

where the infinimum is taken over all the possible functions $f : \chi \mapsto \{0,1\}$.

Consider the regression function η defined by $\eta(x) = \mathbb{P}(Y = 1|X = x)$ for any x in χ, And define the Bayes classifier :

$$f^*(x) = 1_{\eta(x) > \frac{1}{2}}$$

Theorem : $R(f^*) = R^*$

Proof We want to prove that for any function f we have $R(f) - R(f^*) \geq 0$.

Observe that :

$$R(f) - R(f^*) = \int \mathbb{P}_X(dx)\Big(\mathbb{P}(f(x) \neq Y|X = x) - \mathbb{P}(f^*(x) \neq Y|X = x)\Big)$$

Therefore it's enough to prove that for any x in χ :

$$\mathbb{P}(f(x) \neq Y|X = x) - \mathbb{P}(f^*(x) \neq Y|X = x) \geq 0$$

Now :

$$\begin{aligned}\mathbb{P}(f(x) \neq Y|X = x) &= 1 - \mathbb{P}(f(x) = Y|X = x) \\ &= 1 - \Big(\mathbb{P}(f(x) = 1, Y = 1|X = x) + \mathbb{P}(f(x) = 0, Y = 0|X = x)\Big) \\ &= 1 - \Big(1_{f(x)=1}\mathbb{P}(Y = 1|X = x) + 1_{f(x)=0}\mathbb{P}(Y = 0|X = x)\Big) \\ &= 1 - \Big(1_{f(x)=1}\eta(x) + 1_{f(x)=0}(1 - \eta(x))\Big)\end{aligned}$$

As well as :

$$\mathbb{P}(f^*(x) \neq Y|X = x) = 1 - \Big(1_{f^*(x)=1}\eta(x) + 1_{f*(x)=0}(1 - \eta(x))\Big)$$

Using these two formulas :

$$\begin{aligned}&\mathbb{P}(f(x) \neq Y|X = x) - \mathbb{P}(f^*(x) \neq Y|X = x) \\ &= 1_{f^*(x)=1}\eta(x) + 1_{f*(x)=0}(1 - \eta(x)) - 1_{f(x)=1}\eta(x) - 1_{f(x)=0}(1 - \eta(x)) \\ &= 1_{f(x)=1}(1 - 2\eta(x)) - 1_{f^*(x)=1}(1 - 2\eta(x)) \\ &= (2\eta(x) - 1)(1_{f^*(x)=1} - 1_{f(x)=1})\end{aligned}$$

And it's clear the last expression is non-negative by checking the two cases $\eta(x) \geq \frac{1}{2}$ and $\eta(x) < \frac{1}{2}$. ■