

4-8-1'

BAYES OPTIMAL BINARY CLASSIFIER (0-1 loss function)

① Show that the Bayes optimal binary classifier, g^{**} defined by:

$$g^{**} = \underset{g}{\operatorname{argmin}} P[Y \neq g(x)] \quad \text{--- (I)}$$

is same as the function, h , defined by, $h(x) = \underset{z \in Y}{\operatorname{argmax}} P[Y = z | X = x]$.

IMP: DO NOT ASSUME "X" is a ~~continuous~~ DISCRETE R.V.

Proof: We start by re-writing $P[Y \neq g(x)]$ as $E[1_{Y \neq g(x)}]$,

where $1_{Y \in E} \equiv \begin{cases} 1 & \text{if } E \text{ happens} \\ 0 & \text{otherwise} \end{cases}$.

$$E[1_{Y \neq g(x)}] = E[E[1_{Y \neq g(x)} | X]] \quad \text{--- (II)} \quad (\because \text{Total expectation rule})$$

~~Now, $\underset{g}{\operatorname{argmin}}$ minimizing the RHS in the above is same as minimizing $E[1_{Y \neq g(x)}]$ for all x .~~

Now, the values the random variable $E[1_{Y \neq g(x)} | X]$ takes are $E[1_{Y \neq g(x)} | X = x]$. Since defining "g" is same as defining "g(x) for all x", if we choose g(x) such that $E[1_{Y \neq g(x)} | X = x]$ is minimized for at every x, then the given objective (I) will be minimized, since the RHS in (II) will be minimized.

$$\therefore g^{**}(x) = \underset{z \in Y}{\operatorname{argmin}} E[1_{Y \neq z} | X = x]$$

$$= \underset{z \in Y}{\operatorname{argmin}} P[Y \neq z | X = x]$$

$$= \underset{z \in Y}{\operatorname{argmax}} P[Y = z | X = x]$$

Hence Proved.

Remark: Unlike Karthik's approach, the above proof works with X as a conts. r.v. too! In fact, the proof is agnostic to conts/discrete nature of X. This was accomplished by using results about expectations.