TAREA 3: MACHINE LEARNING

JOSÉ LUIS AGUILAR CHARFÉN

Para esta tarea suponga el caso de clasificación binaria donde $Y \in \{0,1\}$, sea $\eta(x) = \mathbb{P}(Y = 1|X = x)$ y g^* el clasificador bayesiano óptimo. El error bayesiano óptimo se define como el error de clasificación del clasficador bayesiano óptimo, denotado como:

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

1. Sea g cualquier otro clasificador. Demuestre que:
$$\mathbb{P}(g(X) \neq Y) - L^* = 2\mathbb{E}\left[\left|\eta(X) - \frac{1}{2}\right| \mathbb{1}_{g^*(X) \neq g(X)}\right]$$
2. Sea $\tilde{\eta} : \mathcal{X} \to [0,1]$, y:
$$g(x) = \begin{cases} 1 & \text{si } \tilde{\eta}(x) < \frac{1}{2}, \\ 0 & \text{en otro caso.} \end{cases}$$
   Demuestre que
$$\mathbb{P}(g(X) \neq Y) - L^* \leq 2\mathbb{E}\left|\eta(X) - \tilde{\eta}(X)\right|$$
3. Supongamos que (X,Y) tiene la siguiente distribución conjunta:

Tabla 1. *Distribución conjunta de (X,Y) y distribución marginal de X.*

	$X=0$	$X=1$	$X=2$
$Y=0$	0.15	0.1	0.3
$Y=1$	0.15	0.2	0.1
$\mathbb{P}(X=x)$	0.15	0.1	0.3

- Calcula el clasificador Bayesiano óptimo para Y si la función de costo es la indicadora (falso positivo tiene el mismo costo que un falso negativo)
- Calcula la probabilidad de cometer un error
- Calcula el clasificador Bayesiano óptimo para Y si predecir mal un 0 cuesta el doble que predecir mal un 1.

Solución

Para el primer inciso, sea $g(x)$ definido de manera idéntica que para el segundo inciso; recordando $g^*(x)$:

$$g^*(x) = \begin{cases} 1 & \text{si } \eta(x) < \frac{1}{2}, \\ 0 & \text{en otro caso.} \end{cases} \tag{1}$$

Debido a que $Y \in \{0,1\}$, entonces meramente deben sumarse las probabilidades de cuando $g(X) = 0$ y $Y = 1$ en un primer caso, o que $g(X) = 1$ y $Y = 0$. Debido a que $\eta(x)$ es meramente la probabilidad de que Y sea 1 dado que X sea x, entonces, si se toma el valor esperado:

$$L^* = \mathbb{P}(g^*(X) \neq Y) \tag{2}$$
$$= \mathbb{E}\left[\mathbb{P}(Y = 1|X = x)\mathbb{1}_{g^*(X)=0} + \mathbb{P}(Y = 0|X = x)\mathbb{1}_{g^*(X)=1}|X = x\right] \tag{3}$$
$$= \mathbb{E}\left[\eta(X)\mathbb{1}_{g^*(X)=0} + (1 - \eta(X))\mathbb{1}_{g^*(X)=1}|X = x\right] \tag{4}$$

De manera análoga, para un clasificador g cualquiera, entonces la función de pérdida es la siguiente:

$$\mathbb{P}(g(X) \neq Y) = \mathbb{E}\left[\eta(X)\mathbb{1}_{g(X)=0} + (1 - \eta(X))\mathbb{1}_{g(X)=1}|X = x\right] \tag{5}$$

Date: 2 de abril de 2022.

Se decide simplificar la notación para que $\mathbb{E}_X[]\cdot] = \mathbb{E}[\cdot|X = x]$.Restando las dos cantidades entonces se obtiene la siguiente ecuación:

$$\mathbb{P}(g(X) \neq Y) - L^* = \mathbb{E}_X\left[\eta(X)\mathbb{1}_{g^*(X)=0} + (1-\eta(X))\mathbb{1}_{g^*(X)=1}\right] - \mathbb{E}_X\left[\eta(X)\mathbb{1}_{g(X)=0} + (1-\eta(X))\mathbb{1}_{g(X)=1}\right] \tag{6}$$

$$= \mathbb{E}_X\left[\eta(X)\mathbb{1}_{g^*(X)=0} + (1-\eta(X))\mathbb{1}_{g^*(X)=1} - \eta(X)\mathbb{1}_{g(X)=0} - (1-\eta(X))\mathbb{1}_{g(X)=1}\right] \tag{7}$$

$$= \mathbb{E}_X\left[\eta(X)\mathbb{1}_{g^*(X)=0} - (1-\eta(X))\mathbb{1}_{g(X)=1} + (1-\eta(X))\mathbb{1}_{g^*(X)=1} - \eta(X)\mathbb{1}_{g(X)=0}\right] \tag{8}$$

$$= \mathbb{E}_X\left[|2\eta(X) - 1|\mathbb{1}_{g^*(X)\neq g(X)}\right] \tag{9}$$

$$= 2\mathbb{E}_X\left[\left|\eta(X) - \frac{1}{2}\right|\mathbb{1}_{g^*(X)\neq g(X)}\right] \tag{10}$$

Puede llegarse a la ecuación 7 por linealidad del valor esperado. Para llegar a la ecuación 8, vale la pena observar que se está agrupando cuando $g(X) = 0$ y $g^*(X) = 1$, y cuando $g(X) = 1$ y $g^*(X) = 0$ en la indicadora que $g^*(X) \neq g(X)$. Se toma el valor absoluto debido a que es irrespectivo de cuál clasificador sea, solamente la probabilidad es el converso. Por último, se llega a la ecuación 10 nuevamente por linealidad del valor esperado, y con esto se llega al resultado esperado, y concluye la primera demostración. ■