

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 clasificador 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 $\tilde{\eta} : \mathcal{X} \rightarrow [0, 1]$, y:

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

Demuestre que

$$\mathbb{P}(g(X) \neq Y) - L^* \leq 2\mathbb{E} |\eta(X) - \tilde{\eta}(X)|$$

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
$\mathbb{P}(X = x)$	0.3	0.3	0.4

- 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

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

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

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:

$$\begin{aligned} (2) \quad L^* &= \mathbb{P}(g^*(X) \neq Y) \\ (3) \quad &= \mathbb{E} [\mathbb{P}(Y = 1|X = x) \mathbb{1}_{g^*(X)=0} + \mathbb{P}(Y = 0|X = x) \mathbb{1}_{g^*(X)=1} | X = x] \\ (4) \quad &= \mathbb{E} [\eta(X) \mathbb{1}_{g^*(X)=0} + (1 - \eta(X)) \mathbb{1}_{g^*(X)=1} | X = x] \end{aligned}$$

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

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

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:

$$\begin{aligned}
(6) \quad \mathbb{P}(g(X) \neq Y) - L^* &= \mathbb{E}_X [\eta(X)\mathbb{1}_{g^*(X)=0} + (1 - \eta(X))\mathbb{1}_{g^*(X)=1}] - \mathbb{E}_X [\eta(X)\mathbb{1}_{g(X)=0} + (1 - \eta(X))\mathbb{1}_{g(X)=1}] \\
(7) \quad &= \mathbb{E}_X [\eta(X)\mathbb{1}_{g^*(X)=0} + (1 - \eta(X))\mathbb{1}_{g^*(X)=1} - \eta(X)\mathbb{1}_{g(X)=0} - (1 - \eta(X))\mathbb{1}_{g(X)=1}] \\
(8) \quad &= \mathbb{E}_X [\eta(X)\mathbb{1}_{g^*(X)=0} - (1 - \eta(X))\mathbb{1}_{g(X)=1} + (1 - \eta(X))\mathbb{1}_{g^*(X)=1} - \eta(X)\mathbb{1}_{g(X)=0}] \\
(9) \quad &= \mathbb{E}_X [2\eta(X) - 1 | \mathbb{1}_{g^*(X) \neq g(X)}] \\
(10) \quad &= 2\mathbb{E}_X \left[\left| \eta(X) - \frac{1}{2} \right| \mathbb{1}_{g^*(X) \neq g(X)} \right]
\end{aligned}$$

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. ■

Segundo ejercicio. Para el segundo ejercicio, asumamos que $\tilde{\eta}$ es meramente un estimado de la probabilidad de que Y sea 1 dado un valor de X , donde $X \in \mathcal{X}$. Dado el clasificador g , entonces debe considerarse no solamente el error discutido en el ejercicio anterior, sino que debido a que $\tilde{\eta}(X) \neq \eta(X)$. Sustituyendo el resultado anterior en la desigualdad:

$$(11) \quad 2\mathbb{E} \left[\left| \eta(X) - \frac{1}{2} \right| \mathbb{1}_{g^*(X) \neq g(X)} \right] \leq 2\mathbb{E} [\eta(X) - \tilde{\eta}(X)]$$

Si $g(X) = g^*(X)$ entonces el lado izquierdo de la ecuación es 0, por lo que la desigualdad siempre se cumple. Entonces, solo debe revisarse en el caso que sean distintos. Si esta desigualdad se cumple para todo $X \in \{\mathcal{X} : g(X) \neq g^*(X)\}$, entonces se cumple para sus valores esperados. Para que $g(X) \neq g^*(X)$, debe cumplirse que $g(X) = 0$, $g^*(X) = 1$ o $g(X) = 1$, $g^*(X) = 0$. Es decir, exclusivamente debe ser uno u otro; en una notación más estándar: $\mathbb{1}_{g^*(X) \neq g(X)} = g^*(X) \oplus g(X)$, donde \oplus se entiende como una disyunción exclusiva. Para que esto ocurra, entonces necesariamente:

$$(12) \quad \left(\eta(x) \leq \frac{1}{2} \wedge \tilde{\eta}(x) > \frac{1}{2} \right) \oplus \left(\eta(x) > \frac{1}{2} \wedge \tilde{\eta}(x) \leq \frac{1}{2} \right)$$

Suponiendo el primer caso, entonces se tiene la desigualdad:

$$(13) \quad \left| \eta(x) - \frac{1}{2} \right| \leq |\eta(X) - \tilde{\eta}(X)|$$

$$(14) \quad \frac{1}{2} - \eta(x) \leq \tilde{\eta}(X) - \eta(X)$$

$$(15) \quad \frac{1}{2} \leq \tilde{\eta}(X)$$

Que es verdadero para todo $X \in \{\mathcal{X} : g(X) \neq g^*(X)\}$ dado el primer caso. Resolviendo la desigualdad en el segundo caso:

$$(16) \quad \left| \eta(x) - \frac{1}{2} \right| \leq |\eta(X) - \tilde{\eta}(X)|$$

$$(17) \quad \eta(x) - \frac{1}{2} \leq \eta(X) - \tilde{\eta}(X)$$

$$(18) \quad \frac{1}{2} \geq \tilde{\eta}(X)$$

Que también es verdadero para todo $X \in \{\mathcal{X} : g(X) \neq g^*(X)\}$ dado el segundo caso. Debido a que la desigualdad se cumple, entonces se llega al resultado:

$$(19) \quad 2\mathbb{E} \left[\left| \eta(X) - \frac{1}{2} \right| \mathbb{1}_{g^*(X) \neq g(X)} \right] \leq 2\mathbb{E} [\eta(X) - \tilde{\eta}(X)]$$

Y con esto concluye la demostración. ■

Tercer ejercicio. Para el tercer ejercicio, se divide en incisos la solución. En su primer inciso, considerando en primer lugar a $\eta(x)$, partiendo de la tabla 1:

$$(20) \quad \eta(x) = \begin{cases} \frac{0.15}{0.3}, & x = 0 \\ \frac{0.2}{0.3}, & x = 1 \\ \frac{0.1}{0.4}, & x = 2 \end{cases}$$

$$(21) \quad = \begin{cases} \frac{1}{2}, & x = 0 \\ \frac{2}{3}, & x = 1 \\ \frac{1}{4}, & x = 2 \end{cases}$$

Si un falso positivo tiene el mismo costo que un falso negativo, entonces la función de costo puede ser definida de la siguiente manera:

$$(22) \quad L^* = \mathbb{P}(Y = 1 \cap g^*(X) = 0) + \mathbb{P}(Y = 0 \cap g^*(X) = 1)$$

Que es equivalente a cómo se ha descrito desde el comienzo, por lo que $g^*(x)$ sigue siendo igual. Por lo tanto, el clasificador óptimo de Bayes es el siguiente:

$$(23) \quad g^*(x) = \begin{cases} 0, & x = 0 \\ 1, & x = 1 \\ 0, & x = 2 \end{cases}$$

Que es la respuesta al primer inciso. Calculando la probabilidad de cometer un error:

$$(24) \quad \mathbb{P}(Y \neq g(X)) = \mathbb{P}(Y = 1 \cap g^*(X) = 0) + \mathbb{P}(Y = 0 \cap g^*(X) = 1)$$

Por ley de probabilidad total puede escribirse de la siguiente manera:

$$\begin{aligned} \mathbb{P}(Y \neq g(X)) &= \sum_{x=0}^2 \mathbb{P}(Y = 1 \cap g^*(x) = 0) + \mathbb{P}(Y = 0 \cap g^*(x) = 1) \\ &= 0 + 0.15 + 0.1 + 0 + 0 + 0.1 \\ &= 0.35 \end{aligned}$$

Y con esto concluye el segundo inciso del tercer ejercicio. Para el tercer inciso, la función de costo puede describirse de la siguiente manera:

$$(25) \quad L^* = \mathbb{P}(Y = 1 \cap g^*(X) = 0) + 2\mathbb{P}(Y = 0 \cap g^*(X) = 1)$$

$$(26) \quad = \mathbb{E}_X [P(Y = 1|X = x)\mathbb{1}_{g^*(X=0)} + 2P(Y = 0|X = x)\mathbb{1}_{g^*(X=1)}]$$

$$(27) \quad = \mathbb{E}_X [\eta(X)\mathbb{1}_{g^*(X=0)} + 2(1 - \eta(X))\mathbb{1}_{g^*(X=1)}]$$

Entonces, $g^*(x)$ puede ser minimizado tomando para todo x :

$$(28) \quad g^*(x) = \begin{cases} 1, & \eta(x) > 2 - 2\eta(x), \\ 0 & \text{en otro caso.} \end{cases}$$

$$(29) \quad = \begin{cases} 1, & \eta(x) > \frac{2}{3}, \\ 0 & \text{en otro caso.} \end{cases}$$

Por lo tanto, el clasificador de Bayes óptimo es el siguiente:

$$(30) \quad g^*(x) = \begin{cases} 0, & x = 0 \\ 0, & x = 1 \\ 0, & x = 2 \end{cases}$$

Y su costo, en este caso, sería de 0.45^1 . Si por el otro lado, en vez de tomar como límite que $\eta(x) > \frac{2}{3}$ para que lo clasifique como 1, fuera una desigualdad no estricta ($\eta(x) \geq \frac{2}{3}$), el clasificador de Bayes sería igual al mostrado en la ecuación 23, con un costo de 0.45 igualmente.

¹No es muy complicado de mostrar; se sigue la misma lógica que en el inciso anterior. Se omite el procedimiento debido a que queda fuera del alcance de la pregunta y se menciona meramente como comentario.