Tarea IV

Y. Sarahi García González

CIMAT

Problema 1

Supongamos que para un problema de clasificación binaria, se construye un clasificador donde se permite, además de regresar como predición 0 y 1, también abstenerse.

El costo de predecir 1 si la verdadera categoria es 0, es 1 peso. El costo de predecir 0 si la verdadera categoria es 1, también es 1 peso. El costo de abstenerse es θ, una constante dada de antemano: $0 < \theta < \frac{1}{2}$.

Calcula el clasificador Bayesiano óptimo en función de θ y $P(Y = 1|X = x)$.

Sol.

Definimos la función de costo, donde ***abstenerse*** lo denotaremos con la etiqueta 2:

$$L(Y = y, \hat{Y} = \hat{y}) = \begin{cases} 0 & \hat{y} = y, y \in \{0,1\} \\ 1 & \hat{y} \neq y, y \in \{0,1\} \\ \theta & \hat{y} = 2, y \in \{0,1\} \end{cases} \tag{1}$$

Buscamos resolver:

$$\min_{\hat{Y}(x)} \left\{ E_{Y|X=x} L\left(Y, \hat{Y}(x)\right) \right\}$$

Como nuestro caso es discreto, lo anterior es:

$$\min_{\hat{Y}(x)} \left\{ \sum_{y} L\left[Y, \hat{Y}(x)\right] p\left(Y = y|X = x\right) \right\}$$

A continuación, vamos a concentrarmos únicamente el la suma, a la que denotareos S.

$$S = L\left[Y = 0, \hat{Y}(x)\right] p\left(Y = 0|X = x\right) + L\left[Y = 1, \hat{Y}(x)\right] p\left(Y = 1|X = x\right)$$

Ahora, según tomemos $\hat{Y}$ **fijo**, el valor de esta suma cambia,

- $\hat{y}(x) = 0 \implies S = L[Y = 1, \hat{Y} = 0] * P(Y = 1|X = x)$
- $\hat{y}(x) = 1 \implies S = L[Y = 0, \hat{Y} = 1] * P(Y = 0|X = x)$
- $\hat{y}(x) = 2 \implies S = L[Y = 0, \hat{Y} = 2] * P(Y = 0|X = x) + L[Y = 1, \hat{Y} = 2] * P(Y = 1|X = x)$

Donde en cada caso tomando en cuenta de la función de costo (1) que $L\left[Y, \hat{Y}(x)\right] = 0$ cuando $\hat{y} = y$. Ahora, sustituyedo los demás valores de L:

- $\hat{y}(x) = 0 \implies S = P(Y = 1|X = x)$
- $\hat{y}(x) = 1 \implies S = P(Y = 0|X = x)$
- $\hat{y}(x) = 2 \implies S = \theta * P(Y = 0|X = x) + \theta * P(Y = 1|X = x)$

Y usando la propiedad de probabilidad conjunta $P(A|B) = 1 - P(A^c|B)$

- $\hat{y}(x) = 0 \implies S = P(Y = 1|X = x) \equiv P$
- $\hat{y}(x) = 1 \implies S = 1 - P(Y = 1|X = x) \equiv 1 - P$
- $\hat{y}(x) = 2 \implies S = \theta$

Finalmente, veamos que las dos primeras posibilidades están directamente relacionadas, que sólo nos abstenemos si θ es menor que el mínimo de P y P^c, y que $P \in [0, 1]$ y $\theta \in (0, \frac{1}{2})$, de manera que el clasicador Bayesiano optimo en términos de $theta$ y de $P(Y = 1|X = x) \equiv P$ es:

$$\hat{Y}(x) = \begin{cases} 0 & P < 0{,}5, P < \theta \\ 1 & P > 0{,}5, P > \theta \\ 2 & \theta < min\{P, 1 - P\} \end{cases} \tag{2}$$

Problema 2

Considera un problema de clasificación binaria con predictores X. Supongamos que $P(Y = 1) = P(Y = 0)$ y que $P(X|Y = i)$ sigue una distribución Poisson con parámetro λ_i.

Derive el clasificador Bayesiano óptimo si el costo de un falso positivo es dos veces el costo de un falso negativo.

Sol.

Si el costo de un falso positivo es dos veces el costo de un falso negativo, la matriz de confusión es:

	$\hat{Y} = 1$	$\hat{Y} = 0$
Y=0	0	L_{FN}
Y=1	$2L_{FN}$	0

Y el clasificdor Bayesiano en el caso binario está dado por:

$$\hat{Y}(x) = I\left[\frac{P(X = x|Y = 1)}{P(X = x|Y = 0)} > \frac{L(0, 1)P(Y = 0)}{L(1, 0)P(Y = 1)}\right] \tag{3}$$

Pero sabemos que

$$P(X = x|Y = i) = \frac{\lambda_i^x exp(\lambda_i)}{x!}$$

Además $P(Y=1)=P(Y=0)$ y $L(0,1)=2L(1,0)$, por lo que el clasificador Bayesiano óptimo binario es:

$$\hat{Y} = I\left[\frac{\lambda_1^x exp(\lambda_1)}{\lambda_0^x exp(\lambda_0)} > 2\right] \tag{4}$$

Problema 3

Supongamos que X, Y sean v.a. discretas:

	X=0	X=1	X=2
Y=0	0.1	0.3	0.25
Y=1	0.25	0.05	0.05

1. Si $L(0,1)=L(1,0)$, calcula el clasificador Bayesiano óptimo de Y usando X.
2. Si $L(0,1)=2L(1,0)$ calcula el clasificador Bayesiano óptimo y su error (promedio) correspondiente.

Sol.

Buscamos:

$$\hat{Y}(x) = I\left[\frac{P(Y=1|X=x)}{P(Y=0|X=x)} > \frac{L(0,1)}{L(1,0)}\right] \tag{5}$$

Por lo que necesitamos las probabilidades mariginales y conjuntas de X y Y. Primero, a partir de la tabla, obtenemos la distribución marginal de X, $P(X)=\sum_y P(Y=y, X=x)$,

$$P[X=x] = \begin{cases} 0{,}35 & \text{si } X=0 \\ 0{,}35 & \text{si } X=1 \\ 0{,}3 & \text{si } X=2 \end{cases}$$

Y lo mismo para la distribución marginal de Y

$$P[Y=y] = \begin{cases} 0{,}65 & \text{si } Y=0 \\ 0{,}35 & \text{si } Y=1 \end{cases}$$

Usando el Teorema de Bayes 6 para la probabilidad condicional $Y \mid X$:

$$P(Y=y \mid X=x) = \frac{P(Y=y)P(X=x \mid Y=y)}{P(X=x)} = \frac{P(X=x, Y=x)}{P(X=x)} \tag{6}$$

	$Y=0$	$Y=1$
$Y \mid X=0$	$\frac{2}{7}$	$\frac{5}{7}$
$Y \mid X=1$	$\frac{6}{7}$	$\frac{1}{7}$
$Y \mid X=2$	$\frac{5}{6}$	$\frac{1}{6}$

$L(0,1) = L(1,0)$

El estimador está dado por:

$$\widehat{Y}(x) = I\left(\frac{P(Y=1 \mid X=x)}{P(Y=0 \mid X=x)} > 1\right) \tag{7}$$

Usando la tabla de probabilidades conjuntas:

$$\begin{aligned}
\frac{P(Y=1 \mid X=0)}{P(Y=0 \mid X=0)} &= \frac{5/7}{2/7} = \frac{5}{2} > 1\\
\frac{P(Y=1 \mid X=1)}{P(Y=0 \mid X=1)} &= \frac{1/7}{6/7} = \frac{1}{6} < 1\\
\frac{P(Y=1 \mid X=2)}{P(Y=0 \mid X=2)} &= \frac{1/6}{5/6} = \frac{1}{5} < 1
\end{aligned}$$

Por lo que el clsificador Bayesiano optimo es:

$$\hat{Y}(x) = \begin{cases} 1 & x = 0 \\ 0 & x \in \{1,2\} \end{cases} \tag{8}$$

$L(0,1) = 2L(1,0)$

En este caso el estimador está dado por:

$$\widehat{Y}(x) = I\left(\frac{P(Y=1 \mid X=x)}{P(Y=0 \mid X=x)} > 2\right) \tag{9}$$

Usando la tabla de probabilidades conjuntas:

$$\begin{aligned}
\frac{P(Y=1 \mid X=0)}{P(Y=0 \mid X=0)} &= \frac{5/7}{2/7} = \frac{5}{2} > 2\\
\frac{P(Y=1 \mid X=1)}{P(Y=0 \mid X=1)} &= \frac{1/7}{6/7} = \frac{1}{6} < 2\\
\frac{P(Y=1 \mid X=2)}{P(Y=0 \mid X=2)} &= \frac{1/6}{5/6} = \frac{1}{5} < 2
\end{aligned}$$

Por lo que el clsificador Bayesiano optimo es:

$$\hat{Y}(x) = \begin{cases} 1 & x = 0 \\ 0 & x \in \{1,2\} \end{cases} \tag{10}$$

Ahora el error, la suma de todos los casos donde no se asigna la etiqueta correcta:

$$\begin{aligned}
\mathbb{E}[L(Y,\widehat{Y}(x))] = L(1,0)&P(Y=0 \mid X=0)P(X=0)\\
+ L(0,1)&P(Y=1 \mid X=1)P(X=1)\\
+ L(0,1)&P(Y=1 \mid X=2)P(X=2)
\end{aligned}$$

$$
\begin{aligned}
\implies \mathbb{E}[L(Y, \widehat{Y}(x))] &= \frac{2}{7}0{,}35L(1,0) + \frac{1}{7}0{,}35L(0,1) + \frac{1}{6}0{,}3L(0,1) \\
&= L(1,0)\frac{1}{10} + L(1,0)\frac{1}{10} \\
&= \frac{3}{10}L(1,0)
\end{aligned}
$$