Tarea 06 - Reconocimiento de patrones
Giovanni Gamaliel López Padilla

Problema 01

Supongamos que (X,Y) son variables aleatorias discretas con la siguiente distribución conjunta:

	X=1	X=2	X=3	X=4
Y=0	0.1	0.05	0.05	0.15
Y=1	0.12	0.1	0.25	0.18

Queremos predecir Y en base del valor observado para X.

- **Calcula el clasificador Bayesiano Optimo si equivocarse de categoría tiene costo 1 y no equivocarse tiene costo 0. ¿Cuál es el costo (error) promedio para este clasificador?**
- **Calcula el clasificador Bayesiano Optimo si clasificar una observación mal cuando el verdadero valor es Y = 1 tiene un costo 3 y en el otro caso tiene costo 2.**

Para una x fija buscamos la asignación $\hat{Y}(x)$ que minimiza el error mostrado en la ecuación 1.

$$E_{Y|X=x}[L(Y,\hat{Y}(x))] \tag{1}$$

Para este problema se tiene que

$$E_{Y|X=x}[L(Y,\hat{Y}(x))] = L(0,\hat{Y}(x))P(Y=0|X=x) + L(1,\hat{Y}(x))P(Y=1|X=x)$$

con $L(Y,Y)=0$.

$$\text{Si } \hat{Y}(x)=0 \Rightarrow E_{Y|X=x}[L(Y,\hat{Y}(x))] = L(1,0)P(Y=1|X=x)$$
$$\text{Si } \hat{Y}(x)=1 \Rightarrow E_{Y|X=x}[L(Y,\hat{Y}(x))] = L(0,1)P(Y=1|X=x)$$

Si

$$\frac{L(1,0)P(Y=1|X=x)}{L(0,1)P(Y=1|X=x)} > 1$$

entonces $\hat{Y}(x)=1$ minimiza el error. De lo contrario $\hat{Y}(x)=0$ obtiene el mínimo. Si el cociente es igual a 1 entonces las dos opciones minimizan. En particular elegimos $\hat{Y}(x)=0$.

Por lo tanto el clasificador tiene la siguiente forma:

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

por el teorema de bayes, se obtiene lo siguiente:

$$\begin{aligned}
\hat{Y}(x) &= \mathbb{I}\left[\frac{P(Y=1|X=x)}{P(Y=1|X=x)} > \frac{L(0,1)}{L(1,0)}\right] \\
&= \mathbb{I}\left[\frac{\frac{P(X=x|Y=1)P(Y=1)}{P(X=x)}}{\frac{P(X=x|Y=0)P(Y=0)}{P(X=x)}} > \frac{L(0,1)}{L(1,0)}\right] \\
&= \mathbb{I}\left[\frac{P(X=x|Y=1)}{P(X=x|Y=0)} > \frac{L(0,1)P(Y=0)}{L(1,0)P(Y=1)}\right]
\end{aligned}$$

Usando la ley de la probabilidad total, se tiene el siguiente resultado

$$\begin{aligned}
P(Y=0) &= \sum P(Y=0|X=x)P(X=x) = \sum P(Y=0, X=x) = 0.35 \\
P(Y=1) &= \sum P(Y=1|X=x)P(X=x) = \sum P(Y=1, X=x) = 0.65
\end{aligned}$$

Calculando las probabilidades condicionales se tiene lo siguiente

$$\begin{aligned}
P(X=1|Y=1) &= \frac{12}{65} \qquad & P(X=1|Y=0) &= \frac{10}{35} \\
P(X=2|Y=1) &= \frac{10}{65} \qquad & P(X=2|Y=0) &= \frac{5}{35} \\
P(X=3|Y=1) &= \frac{25}{65} \qquad & P(X=3|Y=0) &= \frac{5}{35} \\
P(X=4|Y=1) &= \frac{18}{65} \qquad & P(X=4|Y=0) &= \frac{15}{35}
\end{aligned}$$

por lo tanto, los coeficientes son

$$\begin{aligned}
\frac{P(X=1|Y=1)}{P(X=1|Y=0)} &= \frac{420}{650} \qquad & \frac{P(X=1|Y=1)}{P(X=1|Y=0)} &= \frac{875}{325} \\
\frac{P(X=2|Y=1)}{P(X=2|Y=0)} &= \frac{350}{325} \qquad & \frac{P(X=1|Y=1)}{P(X=1|Y=0)} &= \frac{630}{975}
\end{aligned}$$

Supongamos que el costo de equivocarse es 1, entonces

$$\frac{L(0,1)}{L(1,0)} = \frac{1}{1} = 1$$

Con estos resultados obtener que el clasificador esta definido por:
si $x = 1$

$$\begin{aligned}\hat{y}(x) &= \mathbb{I}\left[\frac{P(X=1|Y=1)}{P(X=1|Y=0)} > \frac{L(0,1)P(Y=0)}{L(1,0)P(Y=1)}\right]\\ &= \mathbb{I}\left[\frac{420}{650} > \frac{35}{65}\right]\\ &= 1\end{aligned}$$

si $x = 2$

$$\begin{aligned}\hat{y}(x) &= \mathbb{I}\left[\frac{P(X=2|Y=1)}{P(X=2|Y=0)} > \frac{L(0,1)P(Y=0)}{L(1,0)P(Y=1)}\right]\\ &= \mathbb{I}\left[\frac{70}{65} > \frac{35}{65}\right]\\ &= 1\end{aligned}$$

si $x = 3$

$$\begin{aligned}\hat{y}(x) &= \mathbb{I}\left[\frac{P(X=3|Y=1)}{P(X=3|Y=0)} > \frac{L(0,1)P(Y=0)}{L(1,0)P(Y=1)}\right]\\ &= \mathbb{I}\left[\frac{175}{65} > \frac{35}{65}\right]\\ &= 1\end{aligned}$$

si $x = 4$

$$\begin{aligned}\hat{y}(x) &= \mathbb{I}\left[\frac{P(X=4|Y=1)}{P(X=4|Y=0)} > \frac{L(0,1)P(Y=0)}{L(1,0)P(Y=1)}\right]\\ &= \mathbb{I}\left[\frac{42}{65} > \frac{35}{65}\right]\\ &= 1\end{aligned}$$

Calculando el error se tiene lo siguiente:

Si $\hat{y}(x) = 1$, entonces

$$E_{Y|X=x}[L(y, \hat{y}(x))] = L(0,1)P(Y=0) = P(Y=0) = 0.35$$

Supongamos que $L(y, \hat{y}(x))$ esta dada por

$$L(1,0) = 3 \qquad L(0,1) = 2$$

es decir, clasificar mal cuando el verdadero valor de y es 1 tiene un costo de 3, en otro caso tiene un costo de 2.

por lo tanto el clasificador esta dado por si $x = 1$

$$\begin{aligned}\hat{y}(x) &= \mathbb{I}\left[\frac{P(X=1|Y=1)}{P(X=1|Y=0)} > \frac{L(0,1)P(Y=0)}{L(1,0)P(Y=1)}\right]\\ &= \mathbb{I}\left[\frac{420}{650} > \frac{70}{195}\right]\\ &= 1\end{aligned}$$

si $x = 2$

$$\begin{aligned}\hat{y}(x) &= \mathbb{I}\left[\frac{P(X=2|Y=1)}{P(X=2|Y=0)} > \frac{L(0,1)P(Y=0)}{L(1,0)P(Y=1)}\right]\\ &= \mathbb{I}\left[\frac{70}{65} > \frac{70}{195}\right]\\ &= 1\end{aligned}$$

si $x = 3$

$$\begin{aligned}\hat{y}(x) &= \mathbb{I}\left[\frac{P(X=3|Y=1)}{P(X=3|Y=0)} > \frac{L(0,1)P(Y=0)}{L(1,0)P(Y=1)}\right]\\ &= \mathbb{I}\left[\frac{175}{65} > \frac{70}{195}\right]\\ &= 1\end{aligned}$$

si $x = 4$

$$\begin{aligned}\hat{y}(x) &= \mathbb{I}\left[\frac{P(X=4|Y=1)}{P(X=4|Y=0)} > \frac{L(0,1)P(Y=0)}{L(1,0)P(Y=1)}\right]\\ &= \mathbb{I}\left[\frac{42}{65} > \frac{70}{195}\right]\\ &= 1\end{aligned}$$

Problema 02

Deriva el clasificador Bayesiano óptimo para el caso de tres clases y una función de costo simétrica cuando:

$$X|Y=1 \sim \mathcal{N}(\mu_1, \Sigma) \qquad X|Y=2 \sim \mathcal{N}(\mu_2, \Sigma) \qquad X|Y=3 \sim \mathcal{N}(\mu_3, \Sigma)$$

y

$$P(Y=1) = 2P(Y=2) = P(Y=3)$$

Para una x fija, buscamos la asgianción $\hat{y}(x)$ que minimiza el error

$$E_{Y|X=x}[L(y,\hat{y}(x))]$$

En este caso

$$\begin{aligned} E_{Y|X=x}[L(y,\hat{y}(x))] =& L(1,\hat{y}(x))P(Y=1|X=x)+ \\ & L(2,\hat{y}(x))P(Y=2|X=x)+ \\ & L(3,\hat{y}(x))P(Y=3|X=x) \end{aligned}$$

Si $\hat{y}(x) = 1$, entonces

$$\begin{aligned} E_{Y|X=x}[L(y,\hat{y}(x))] &= L(2,1)P(Y=2|X=x) + L(3,1)P(Y=3|X=x) \\ &= P(Y=2|X=x) + P(Y=3|X=x) \end{aligned}$$

Si $\hat{y}(x) = 2$, entonces

$$\begin{aligned} E_{Y|X=x}[L(y,\hat{y}(x))] &= L(1,2)P(Y=1|X=x) + L(3,2)P(Y=3|X=x) \\ &= P(Y=1|X=x) + P(Y=3|X=x) \end{aligned}$$

Si $\hat{y}(x) = 3$, entonces

$$\begin{aligned} E_{Y|X=x}[L(y,\hat{y}(x))] &= L(1,3)P(Y=1|X=x) + L(2,3)P(Y=2|X=x) \\ &= P(Y=1|X=x) + P(Y=2|X=x) \end{aligned}$$

Usando el teorema de bayes se obtiene que

$$\begin{aligned} P(Y=2|X=x) + P(Y=3|X=x) &= \frac{P(X=x|Y=2)P(Y=2)}{\sum_j P(X=x|Y=j)P(Y=j)} + \frac{P(X=x|Y=3)P(Y=3)}{\sum_j P(X=x|Y=j)P(Y=j)} \\ P(Y=1|X=x) + P(Y=3|X=x) &= \frac{P(X=x|Y=1)P(Y=1)}{\sum_j P(X=x|Y=j)P(Y=j)} + \frac{P(X=x|Y=3)P(Y=3)}{\sum_j P(X=x|Y=j)P(Y=j)} \\ P(Y=1|X=x) + P(Y=2|X=x) &= \frac{P(X=x|Y=1)P(Y=1)}{\sum_j P(X=x|Y=j)P(Y=j)} + \frac{P(X=x|Y=2)P(Y=2)}{\sum_j P(X=x|Y=j)P(Y=j)} \end{aligned}$$

si se cumpla la condición

$$P(X=x|Y=1)P(Y=1) \leq P(X=x|Y=2)P(Y=2)$$
$$P(X=x|Y=1)P(Y=1) \leq P(X=x|Y=3)P(Y=3)$$

entonces se cumple que

$$E_{Y|X=x}[L(y,1)] \leq E_{Y|X=x}[L(y,2)]$$
$$E_{Y|X=x}[L(y,1)] \leq E_{Y|X=x}[L(y,3)]$$

por lo que elegimos $\hat{y}(x) = 1$.

Si se cumpla la condición

$$P(X = x|Y = 2)P(Y = 2) \leq P(X = x|Y = 1)P(Y = 1)$$
$$P(X = x|Y = 2)P(Y = 2) \leq P(X = x|Y = 3)P(Y = 3)$$

entonces se cumple que

$$E_{Y|X=x}[L(y,2)] \leq E_{Y|X=x}[L(y,1)]$$
$$E_{Y|X=x}[L(y,2)] \leq E_{Y|X=x}[L(y,3)]$$

por lo que elegimos $\hat{y}(x) = 2$.

Si se cumpla la condición

$$P(X = x|Y = 3)P(Y = 3) \leq P(X = x|Y = 1)P(Y = 1)$$
$$P(X = x|Y = 3)P(Y = 3) \leq P(X = x|Y = 2)P(Y = 2)$$

entonces se cumple que

$$E_{Y|X=x}[L(y,3)] \leq E_{Y|X=x}[L(y,1)]$$
$$E_{Y|X=x}[L(y,3)] \leq E_{Y|X=x}[L(y,2)]$$

por lo que elegimos $\hat{y}(x) = 3$.

Por lo tanto, para una x fija debemos calcular tres coefiencientes y compararlos con 1 para determinar la clasificación correspondiente. Una alternativa es calcular el logaritmo de los coeficientes y compararlo con 0. Los coeficientes son los siguientes:

$$\frac{P(X = x|Y = 1)P(Y = 1)}{P(X = x|Y = 2)P(Y = 2)} = \frac{P(X = x|Y = 1)2P(Y = 2)}{P(X = x|Y = 2)P(Y = 2)} = \frac{2P(X = x|Y = 1)}{P(X = x|Y = 2)}$$
$$\frac{P(X = x|Y = 1)P(Y = 1)}{P(X = x|Y = 3)P(Y = 3)} = \frac{P(X = x|Y = 1)P(Y = 3)}{P(X = x|Y = 3)P(Y = 3)} = \frac{P(X = x|Y = 1)}{P(X = x|Y = 3)}$$
$$\frac{P(X = x|Y = 2)P(Y = 2)}{P(X = x|Y = 3)P(Y = 3)} = \frac{P(X = x|Y = 2)P(Y = 2)}{2P(X = x|Y = 3)P(Y = 2)} = \frac{P(X = x|Y = 2)}{2P(X = x|Y = 3)}$$

Dado que $X|Y = j \sim \mathcal{N}(\mu_j = \Sigma)$, entonces calculando los logaritmos se tiene que

$$\log\left(\frac{P(X=x|Y=1)}{P(X=x|Y=2)}\right) = \log(2) + (\mu_1-\mu_2)^T\Sigma^{-1}x + \frac{1}{2}(\mu_1^T\Sigma\mu_1 - \mu_2^T\Sigma\mu_2) = \ell_{1,2}^T x + b_{1,2}$$
$$\log\left(\frac{P(X=x|Y=1)}{P(X=x|Y=3)}\right) = (\mu_1-\mu_3)^T\Sigma^{-1}x + \frac{1}{2}(\mu_1^T\Sigma\mu_1 - \mu_3^T\Sigma\mu_3) = \ell_{1,3}^T x + b_{1,3}$$
$$\log\left(\frac{P(X=x|Y=2)}{P(X=x|Y=3)}\right) = (\mu_2-\mu_2)^T\Sigma^{-1}x + \frac{1}{2}(\mu_2^T\Sigma\mu_2 - \mu_3^T\Sigma\mu_3) - \log 2 = \ell_{2,3}^T x + b_{2,3}$$

por lo tanto

$$\hat{y}(x) = \begin{cases} 1 & \text{para } \ell_{1,2}^T x + b_{1,2} \geq 0 \text{ y } \ell_{1,3}^T x + b_{1,3} \geq 0 \\ 2 & \text{para } \ell_{1,3}^T x + b_{1,3} \leq 0 \text{ y } \ell_{2,3}^T x + b_{2,3} \geq 0 \\ 3 & \text{para } \ell_{2,3}^T x + b_{2,3} \leq 0 \text{ y } \ell_{2,3}^T x + b_{2,3} \leq 0 \end{cases}$$