Por favor seguir las instrucciones indicadas por Johan (se debe entregar todo integrado en un solo pdf.) (-1) por no seguir esta indicación.

Nota: 9 - 1 = 8 / 10.

Reconocimiento Estadístico de Patrones
Tarea 6

Randy Osbaldo Ibarra Cayo

B. Preguntas Cortas

Pregunta 4 **Nota: 2 / 2**

Verifica la parte faltante de la ecuación de Cauchy-Schwartz generalizada: Mostrar que en $x \propto B^{-1}d$ se alcanza la cota superior. (video clase 25/3 minuto 22)

Solución

Buscamos $x \in \mathbb{R}^d$ que maximice

$$\frac{(x^T d)^2}{x^T B x}$$

donde $d \in \mathbb{R}^d$ y $B \in \mathbb{R}^{d\times d}$ es simétrica y definida positiva. Por la desigualdad de Cauchy-Schwartz generalizada

$$\frac{(x^T d)^2}{x^T B x} \leq \frac{(x^T B x)(d^T B^{-1} d)}{x^T B x} = d^T B^{-1} d$$ Ok

y notemos que $d^T B^{-1} d$ no depende de x, por lo que encontramos una cota superior para toda x. Por otro lado, notemos que si $x = B^{-1}d$ se tiene que

$$\frac{(x^T d)^2}{x^T B x} = \frac{[(B^{-1}d)^T d]^2}{(B^{-1}d)^T B (B^{-1}d)} = \frac{(d^T B^{-T} d)^2}{d^T B^{-T} B B^{-1} d} = \frac{(d^T B^{-T} d)^2}{d^T B^{-T} d} = d^T B^{-T} d$$ Ok

es decir, la función

$$\frac{(x^T d)^2}{x^T B x}$$

alcanza su máximo en $x = B^{-1}d$. Ok

Pregunta 5 **Nota: 1 / 2**

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

	$X = 1$	$X = 2$	$X = 3$	$X = 4$
$Y = 0$	0, 10	0, 05	0,05	0,15
$Y = 1$	0, 12	0, 10	0,25	0,18

Queremos predecir Y en base del valor observado para X.

- Calcula el Clasificador Bayesiano Óptimo 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 Óptimo si clasificar una observación mal cuando el verdadero valor es $Y = 1$ tiene un costo 3 y en el otro caso tiene costo 2.

Solución

El Clasificador Bayesiano Óptimo resulta de encontrar $\hat{y}$ que minimiza el error para cada x, es decir, para x fija buscamos la asignación $\hat{y}(x)$ que minimiza el error

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

Para este problema Y toma solo dos valores, por lo que podemos expresar el error de la siguiente forma

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

Además, en este problema $L(Y, Y) = 0$, de modo que

- Si $\hat{y}(x) = 0$, entonces $E_{Y|X=x}\left[L(Y, \hat{y}(x))\right] = L(1, 0))P(Y = 1|X = x)$.
- Si $\hat{y}(x) = 1$, entonces $E_{Y|X=x}\left[L(Y, \hat{y}(x))\right] = L(0, 1))P(Y = 0|X = x)$.

Ok

Ahora bien, buscamos minimizar el error y para esto basta comparar $L(1, 0))P(Y = 1|X = x)$ y $L(0, 1))P(Y = 0|X = x)$, esto lo hacemos comparando el cociente de ambos valores con 1

- Si $[L(1, 0)P(Y = 1|X = x)]/[L(0, 1)P(Y = 0|X = x)]$ es mayor a 1 significa que el error se minimiza cuando $\hat{y}(x) = 1$.
- Si $[L(1, 0)P(Y = 1|X = x)]/[L(0, 1)P(Y = 0|X = x)]$ es menor a 1 significa que el error se minimiza cuando $\hat{y}(x) = 0$.
- Si $[L(1, 0)P(Y = 1|X = x)]/[L(0, 1)P(Y = 0|X = x)]$ es igual 1 significa que el error es igual para cualqquier asignación, por lo que elegimos en particular $\hat{y}(x) = 0$.

En consecuencia obtenemos un clasificador de la siguiente forma

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

Notemos que el Teorema de Bayes nos dice

$$\begin{aligned}P(Y = 1|X = x) &= \frac{P(Y = 1|X = x)P(Y = 1)}{P(X = x)} \\ P(Y = 0|X = x) &= \frac{P(Y = 0|X = x)P(Y = 0)}{P(X = x)}\end{aligned}$$

Por lo cual podemos reescribir nuestro clasificador como

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

Usando la Ley de Probabilidad Total podemos obtener la función de probabilidad de Y

$$P(Y=0)=\sum_x P(Y=0|X=x)P(X=x)=\sum_x P(Y=0,X=x)=0.10+0.05+0.05+0.15=0.35$$
$$P(Y=1)=\sum_x P(Y=1|X=x)P(X=x)=\sum_x P(Y=1,X=x)=0.12+0.10+0.25+0.18=0.65$$

ahora, calculamos las probabilidades condicionales

$$P(X=1|Y=1)=\frac{P(X=1,Y=1)}{P(Y=1)}=\frac{0.12}{0.65}=\frac{12}{65}$$
$$P(X=2|Y=1)=\frac{P(X=2,Y=1)}{P(Y=1)}=\frac{0.10}{0.65}=\frac{10}{65}$$
$$P(X=3|Y=1)=\frac{P(X=3,Y=1)}{P(Y=1)}=\frac{0.25}{0.65}=\frac{25}{65}$$
$$P(X=4|Y=1)=\frac{P(X=4,Y=1)}{P(Y=1)}=\frac{0.18}{0.65}=\frac{18}{65}$$

Ok.
Basta compara los numeradores
esto es, las probabilidades que
aparecen en la tabla.

$$P(X=1|Y=0)=\frac{P(X=1,Y=0)}{P(Y=0)}=\frac{0.10}{0.65}=\frac{10}{35}$$
$$P(X=2|Y=0)=\frac{P(X=2,Y=0)}{P(Y=0)}=\frac{0.05}{0.65}=\frac{5}{35}$$
$$P(X=3|Y=0)=\frac{P(X=3,Y=0)}{P(Y=0)}=\frac{0.05}{0.65}=\frac{5}{35}$$
$$P(X=4|Y=0)=\frac{P(X=4,Y=0)}{P(Y=0)}=\frac{0.15}{0.65}=\frac{15}{35}$$

Ok

y por último cálculamos los cocientes

$$\frac{P(X=1|Y=1)}{P(X=1|Y=0)}=\frac{12/65}{10/35}=\frac{12\times 35}{65\times 10}=\frac{420}{650}$$
$$\frac{P(X=2|Y=1)}{P(X=2|Y=0)}=\frac{10/65}{5/35}=\frac{10\times 35}{65\times 5}=\frac{350}{325}$$
$$\frac{P(X=3|Y=1)}{P(X=3|Y=0)}=\frac{25/65}{5/35}=\frac{25\times 35}{65\times 5}=\frac{875}{325}$$
$$\frac{P(X=4|Y=1)}{P(X=4|Y=0)}=\frac{18/65}{15/35}=\frac{18\times 35}{65\times 15}=\frac{630}{975}$$
$$\frac{P(Y=0)}{P(Y=1)}=\frac{0.35}{0.65}=\frac{35}{65}$$

Ok

- Suponemos que $L(Y,\hat{y}(X))=I(Y\neq\hat{y}(X))$, es decir, equivocarse de categoría tiene costo 1. Entonces
$$\frac{L(0,1)}{L(1,0)}=\frac{1}{1}=1$$
y el Clasificador Bayesiano Óptimo está dado por
  - Si $x=1$, entonces
$$\begin{aligned}\hat{y}(1)&=\mathbb{I}\left[\frac{P(X=1|Y=1)}{P(X=1|Y=0)}>\frac{L(0,1)}{L(1,0)}\frac{P(Y=0)}{P(Y=1)}\right]\\&=\mathbb{I}\left[\frac{420}{650}>\frac{35}{65}\right]\\&=\mathbb{I}\left[\frac{42}{65}>\frac{35}{65}\right]\\&=1\end{aligned}$$

– Si $x = 2$, entonces

$$\begin{aligned}
\hat{y}(2) &= \mathbb{I}\left[\frac{P(X=2|Y=1)}{P(X=2|Y=0)} > \frac{L(0,1)}{L(1,0)}\frac{P(Y=0)}{P(Y=1)}\right] \\
&= \mathbb{I}\left[\frac{350}{325} > \frac{35}{65}\right] \\
&= \mathbb{I}\left[\frac{700}{650} > \frac{35}{65}\right] \\
&= \mathbb{I}\left[\frac{70}{65} > \frac{35}{65}\right] \\
&= 1
\end{aligned}$$

– Si $x = 3$, entonces

$$\begin{aligned}
\hat{y}(3) &= \mathbb{I}\left[\frac{P(X=3|Y=1)}{P(X=3|Y=0)} > \frac{L(0,1)}{L(1,0)}\frac{P(Y=0)}{P(Y=1)}\right] \\
&= \mathbb{I}\left[\frac{875}{325} > \frac{35}{65}\right] \\
&= \mathbb{I}\left[\frac{1750}{650} > \frac{35}{65}\right] \\
&= \mathbb{I}\left[\frac{175}{65} > \frac{35}{65}\right] \\
&= 1
\end{aligned}$$

– Si $x = 4$, entonces

$$\begin{aligned}
\hat{y}(4) &= \mathbb{I}\left[\frac{P(X=4|Y=1)}{P(X=4|Y=0)} > \frac{L(0,1)}{L(1,0)}\frac{P(Y=0)}{P(Y=1)}\right] \\
&= \mathbb{I}\left[\frac{630}{975} > \frac{35}{65}\right] \\
&= \mathbb{I}\left[\frac{210}{325} > \frac{35}{65}\right] \\
&= \mathbb{I}\left[\frac{420}{650} > \frac{35}{65}\right] \\
&= \mathbb{I}\left[\frac{42}{65} > \frac{35}{65}\right] \\
&= 1
\end{aligned}$$

Ok

Por último, calculamos el error

– Si $\hat{y}(x) = 0$, entonces el error es $E\left[L(Y, \hat{y}(x))\right] = L(1,0)P(Y=1) = P(Y=1) = 0.65$.

– Si $\hat{y}(x) = 1$, entonces el error es $E\left[L(Y, \hat{y}(x))\right] = L(0,1)P(Y=0) = P(Y=0) = 0.35$.

Sólo hay un error.
En este caso, como y=1
siempre, el error es **0.35**.

- Suponemos que $L(Y, \hat{y}(x))$ está dada por

$$\begin{aligned}
L(1,0) &= 3 \\
L(0,1) &= 2
\end{aligned}$$

es decir, clasificar una observación mal cuando el verdadero valor es $Y = 1$ tiene un costo 3 y en el otro caso tiene costo 2. Entonces

$$\frac{L(0,1)}{L(1,0)} = \frac{2}{3} < 1$$

Ok

y el Clasificador Bayesiano Óptimo está dado por

- Si $x = 1$, entonces

$$\begin{aligned}
\hat{y}(1) &= \mathbb{I}\left[\frac{P(X=1|Y=1)}{P(X=1|Y=0)} > \frac{L(0,1)}{L(1,0)}\frac{P(Y=0)}{P(Y=1)}\right] \\
&= \mathbb{I}\left[\frac{420}{650} > \frac{2}{3}\frac{35}{65}\right] \\
&= \mathbb{I}\left[\frac{42}{65} > \frac{2}{3}\frac{35}{65}\right] \\
&= 1
\end{aligned}$$

- Si $x = 2$, entonces

$$\begin{aligned}
\hat{y}(2) &= \mathbb{I}\left[\frac{P(X=2|Y=1)}{P(X=2|Y=0)} > \frac{L(0,1)}{L(1,0)}\frac{P(Y=0)}{P(Y=1)}\right] \\
&= \mathbb{I}\left[\frac{350}{325} > \frac{2}{3}\frac{35}{65}\right] \\
&= \mathbb{I}\left[\frac{700}{650} > \frac{2}{3}\frac{35}{65}\right] \\
&= \mathbb{I}\left[\frac{70}{65} > \frac{2}{3}\frac{35}{65}\right] \\
&= 1
\end{aligned}$$

- Si $x = 3$, entonces

$$\begin{aligned}
\hat{y}(3) &= \mathbb{I}\left[\frac{P(X=3|Y=1)}{P(X=3|Y=0)} > \frac{L(0,1)}{L(1,0)}\frac{P(Y=0)}{P(Y=1)}\right] \\
&= \mathbb{I}\left[\frac{875}{325} > \frac{2}{3}\frac{35}{65}\right] \\
&= \mathbb{I}\left[\frac{1750}{650} > \frac{2}{3}\frac{35}{65}\right] \\
&= \mathbb{I}\left[\frac{175}{65} > \frac{2}{3}\frac{35}{65}\right] \\
&= 1
\end{aligned}$$

- Si $x = 4$, entonces

$$\begin{aligned}
\hat{y}(4) &= \mathbb{I}\left[\frac{P(X=4|Y=1)}{P(X=4|Y=0)} > \frac{L(0,1)}{L(1,0)}\frac{P(Y=0)}{P(Y=1)}\right] \\
&= \mathbb{I}\left[\frac{630}{975} > \frac{2}{3}\frac{35}{65}\right] \\
&= \mathbb{I}\left[\frac{210}{325} > \frac{2}{3}\frac{35}{65}\right] \\
&= \mathbb{I}\left[\frac{420}{650} > \frac{2}{3}\frac{35}{65}\right] \\
&= \mathbb{I}\left[\frac{42}{65} > \frac{2}{3}\frac{35}{65}\right] \\
&= 1
\end{aligned}$$

Ok

- Suponemos que $L(Y, \hat{y}(x))$ está dada por

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

es decir, clasificar una observación mal cuando el verdadero valor es $Y = 1$ tiene un costo 3 y en el otro caso tiene costo 2. Entonces

$$\frac{L(0,1)}{L(1,0)} = \frac{3}{2}$$

No, está al revés. Es 2 / 3.

y el Clasificador Bayesiano Óptimo está dado por

- Si $x = 1$, entonces

$$\begin{aligned}
\hat{y}(1) &= \mathbb{I}\left[\frac{P(X=1|Y=1)}{P(X=1|Y=0)} > \frac{L(0,1)}{L(1,0)}\frac{P(Y=0)}{P(Y=1)}\right] \\
&= \mathbb{I}\left[\frac{420}{650} > \frac{3}{2}\frac{35}{65}\right] \\
&= \mathbb{I}\left[\frac{42}{65} > \frac{105}{130}\right] \\
&= \mathbb{I}\left[\frac{84}{130} > \frac{105}{130}\right] \\
&= 0
\end{aligned}$$

Incorrecto. Debe ser 1.

- Si $x = 2$, entonces

$$\begin{aligned}
\hat{y}(2) &= \mathbb{I}\left[\frac{P(X=2|Y=1)}{P(X=2|Y=0)} > \frac{L(0,1)}{L(1,0)}\frac{P(Y=0)}{P(Y=1)}\right] \\
&= \mathbb{I}\left[\frac{350}{325} > \frac{3}{2}\frac{35}{65}\right] \\
&= \mathbb{I}\left[\frac{700}{650} > \frac{105}{130}\right] \\
&= \mathbb{I}\left[\frac{70}{65} > \frac{105}{130}\right] \\
&= \mathbb{I}\left[\frac{140}{130} > \frac{105}{130}\right] \\
&= 1
\end{aligned}$$

- Si $x = 3$, entonces

$$\begin{aligned}
\hat{y}(3) &= \mathbb{I}\left[\frac{P(X=3|Y=1)}{P(X=3|Y=0)} > \frac{L(0,1)}{L(1,0)}\frac{P(Y=0)}{P(Y=1)}\right] \\
&= \mathbb{I}\left[\frac{875}{325} > \frac{3}{2}\frac{35}{65}\right] \\
&= \mathbb{I}\left[\frac{1750}{650} > \frac{105}{130}\right] \\
&= \mathbb{I}\left[\frac{175}{65} > \frac{105}{130}\right] \\
&= \mathbb{I}\left[\frac{350}{130} > \frac{105}{130}\right] \\
&= 1
\end{aligned}$$

Ok

– Si $x = 4$, entonces

$$\begin{aligned}
\hat{y}(4) &= \mathbb{I}\left[\frac{P(X=4|Y=1)}{P(X=4|Y=0)} > \frac{L(0,1)}{L(1,0)}\frac{P(Y=0)}{P(Y=1)}\right] \\
&= \mathbb{I}\left[\frac{630}{975} > \frac{3}{2}\frac{35}{65}\right] \\
&= \mathbb{I}\left[\frac{210}{325} > \frac{105}{130}\right] \\
&= \mathbb{I}\left[\frac{420}{650} > \frac{105}{130}\right] \\
&= \mathbb{I}\left[\frac{42}{65} > \frac{105}{130}\right] \\
&= \mathbb{I}\left[\frac{84}{130} > \frac{105}{130}\right] \\
&= 0
\end{aligned}$$

Incorrecto. Debe ser 1.

El clasificador bayesiano es el mismo del caso anterior: \hat{y}(x) = 1, para todo x.
Basta ver que la desigualdad del caso anterior P(X=x, y=1) > P(X=x, y=0) implica

3P(X=x, y=1) > 2P(X=x, y=1) > P(X=x, y=0).

Pregunta 6 Nota: 2 / 2

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

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

y

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

Solución

El Clasificador Bayesiano Óptimo resulta de encontrar $\hat{y}$ que minimiza el error para cada x, es decir, para x fija buscamos la asignación $\hat{y}(x)$ que minimiza el error

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

Para este problema Y toma tres valores, por lo que podemos expresar el error de la siguiente forma

$$E_{Y|X=x}\left[L(Y, \hat{y}(x))\right] = L(1, \hat{y}(x))P(Y = 1|X = x) + L(2, \hat{y}(x))P(Y = 2|X = x) + L(3, \hat{y}(x))P(Y = 3|X = x)$$

Luego $L(i, j) = \mathbb{I}(i = j)$, es decir, es una función de costo simétrica (Según el libro Duda Capítulo 2 página 9) de modo que

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

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

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

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

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

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

Entonces, para minimizar el error, dada una x fija debemos comparar los posibles errores que se obtendrán dependiendo de la clase a la que se asigne x. Los posibles errores son

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

Usando el Teorema de Bayes obtenemos las siguientes igualdades

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

por lo que basta comparar

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

y además, notemos que sólo se involucran tres valores: $P(X=x|Y=i)P(Y=i)$ para $i=1,2,3$. Podemos ordenar estos tres valores $(P(X=x|Y=i)P(Y=i))$ de menor a mayor, de donde se obtienen escencialmente tres casos

- Si $P(X=x|Y=1)P(Y=1) \leq P(X=x|Y=2)P(Y=2)$ y $P(X=x|Y=1)P(Y=1) \leq P(X=x|Y=3)P(Y=3)$, entonces se cumplen las siguientes dos desigualdades

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

  es decir

  $$
  \begin{aligned}
  E_{Y|X=x}\left[L(Y,1)\right] &\leq E_{Y|X=x}\left[L(Y,2)\right]\\
  E_{Y|X=x}\left[L(Y,1)\right] &\leq E_{Y|X=x}\left[L(Y,3)\right]
  \end{aligned}
  $$

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

- Si $P(X=x|Y=2)P(Y=2) \leq P(X=x|Y=1)P(Y=1)$ y $P(X=x|Y=2)P(Y=2) \leq P(X=x|Y=3)P(Y=3)$, entonces se cumplen las siguientes dos desigualdades

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

  es decir

  $$
  \begin{aligned}
  E_{Y|X=x}\left[L(Y,2)\right] &\leq E_{Y|X=x}\left[L(Y,1)\right]\\
  E_{Y|X=x}\left[L(Y,2)\right] &\leq E_{Y|X=x}\left[L(Y,3)\right]
  \end{aligned}
  $$

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

- Si $P(X=x|Y=3)P(Y=3) \leq P(X=x|Y=1)P(Y=1)$ y $P(X=x|Y=3)P(Y=3) \leq P(X=x|Y=2)P(Y=2)$, entonces se cumplen las siguientes dos desigualdades

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

  es decir

  $$
  \begin{aligned}
  E_{Y|X=x}\left[L(Y,3)\right] &\leq E_{Y|X=x}\left[L(Y,1)\right]\\
  E_{Y|X=x}\left[L(Y,3)\right] &\leq E_{Y|X=x}\left[L(Y,2)\right]
  \end{aligned}
  $$

  Ok

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

Por lo tanto, para una x fija basta calcular tres cocientes y compararlos con 1 para poder determinar la clasificación correspondiente, o bien, aplicar la función log (Funcion no decreciente) al cociente y comparar con 0, pues esta comparación nos resume los tres puntos anteriores. Estos cocientes son

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

Ok

Dado que $X|Y = j \sim \mathcal{N}(\mu_j, \Sigma)$ vimos en clase

$$\begin{aligned}
\log\left[\frac{2P(X=x|Y=1)}{P(X=x|Y=2)}\right] &= \log(2) + (\mu_1 - \mu_2)^T\Sigma^{-1}x + \frac{1}{2}(\mu_2^T\Sigma^{-1}\mu_2 - \mu_1^T\Sigma^{-1}\mu_1) = l_{1,2}^T x + b_{1,2} + \log(2)\\
\log\left[\frac{P(X=x|Y=1)}{P(X=x|Y=3)}\right] &= (\mu_1 - \mu_3)^T\Sigma^{-1}x + \frac{1}{2}(\mu_3^T\Sigma^{-1}\mu_3 - \mu_1^T\Sigma^{-1}\mu_1) = l_{1,3}^T x + b_{1,3}\\
\log\left[\frac{P(X=x|Y=2)}{2P(X=x|Y=3)}\right] &= -\log(2) + (\mu_2 - \mu_3)^T\Sigma^{-1}x + \frac{1}{2}(\mu_3^T\Sigma^{-1}\mu_3 - \mu_2^T\Sigma^{-1}\mu_2) = l_{2,3}^T x + b_{2,3} - \log(2)
\end{aligned}$$

Ok

donde $l_{i,j} = \Sigma^{-1}(\mu_i - \mu_j)$ y $b_{i,j} = \frac{1}{2}(\mu_j^T\Sigma^{-1}\mu_j - \mu_i^T\Sigma^{-1}\mu_i)$. Por lo tanto, el clasificador bayesiano óptimo está dado por

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

Ok

Como último comentario, notamos que en este caso, nuestro clasificador hace uso los tres clasificadores de Bayes óptimos obtenidos de los tres problemas de clasificación binaria que se pueden definir y esencialmente estamos asignando la clase que indican estos tres clasificadores. Esto es posible gracias a la función de costo simétrica, pues esta función hace que se simplifiquen varias desigualdades importantes.

Estas ecuaciones pueden simplificarse a ecuaciones lineales de la forma

w_i^T x + C_i = 0.

B. Análisis de Datos

Ejercicio 1

Descripción

Trabajamos con los de datos fashion MNIST

`https://www.kaggle.com/zalando-research/fashionmnist`

Se trata de imágenes 28×28 de diez diferentes tipos de prendas. Trabajaremos con `fashion-mnist_train.csv`. Construye un clasificador $K - NN$. Divide los datos al azar en train y test. Usa el error de predicción sobre test para determinar el valor de la k óptima.

Solución

Haciendo algunas pruebas me doy cuenta que la función de R para construir un clasificador basado en $K - NN$ tarda demasiado, por lo que tomamos un subconjunto de tamaño 1000 aunque se debe tomar en cuenta que esta decisión nos deja con una cantidad menor de datos para construir el clasificador.

Ejercicio 2

Descripción

En este ejercicio trabajamos con los datos de Iris que Ronald Fisher usó para introducir LDA. Están en R como `iris`. Construye un clasificador *LDA* limitándose a los datos de la clase virginica y versicolor. Discute los resultados. Es interesante echar un ojo al articulo de Fisher:

`https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.1469-1809.1936.tb02137.x`

Solución