Pajaritos - El patito feo

Las aves parásitas de cría ponen huevos en nidos de otras especies para que críen al pichón parásito. Hay dos especies hospederas que son indistinguibles a simple vista, la principal diferencia entre estas especies es el grado de discriminación y remoción de huevos parásitos de sus nidos.

La especie “aceptadora“ de huevos parásitos, remueve del nido sólo el 30% de los huevos parásitos. Y la especie “rechazadora“ remueve el 80% de los huevos parásitos.

Pajaritos - El patito feo

Las aves parásitas de cría ponen huevos en nidos de otras especies para que crían al pichón parásito. Hay dos especies hospederas que son indistinguibles a simple vista, la principal diferencia entre estas especies es el grado de discriminación y remoción de huevos parásitos de sus nidos.

La especie "aceptadora" de huevos parásitos, remueve del nido sólo el 30% de los huevos parásitos. Y la especie "rechazadora" remueve el 80% de los huevos parásitos. Sea

$$Y = \begin{cases} R & \text{rechazador} \\ A & \text{aceptador} \end{cases}.$$

Pajaritos - El patito feo

Las aves parásitas de cría ponen huevos en nidos de otras especies para que crían al pichón parásito. Hay dos especies hospederas que son indistinguibles a simple vista, la principal diferencia entre estas especies es el grado de discriminación y remoción de huevos parásitos de sus nidos.

La especie "aceptadora" de huevos parásitos, remueve del nido sólo el 30% de los huevos parásitos. Y la especie "rechazadora" remueve el 80% de los huevos parásitos. Sea

$$Y = \begin{cases} R & \text{rechazador} \\ A & \text{aceptador} . \end{cases}$$

El 90% de los nidos son de la especie "aceptadora" y el 10% a la especie "rechazadora".

$$P(Y = A) = 0.9$$

$$P(Y = R) = 0.1$$

Clasificación - El patito feo

El objetivo es conociendo el número de huevos removidos predecir si la especie es rechazadora o aceptadora.

Clasificación - El patito feo

El objetivo es conociendo el número de huevos removidos predecir si la especie es rechazadora o aceptadora.

Supongamos que en un nido se colocan $n = 8$ huevos parasitarios.

Sea $X =$ número de huevos removidos.

Clasificación - El patito feo

El objetivo es conociendo el número de huevos removidos predecir si la especie es rechazadora o aceptadora.

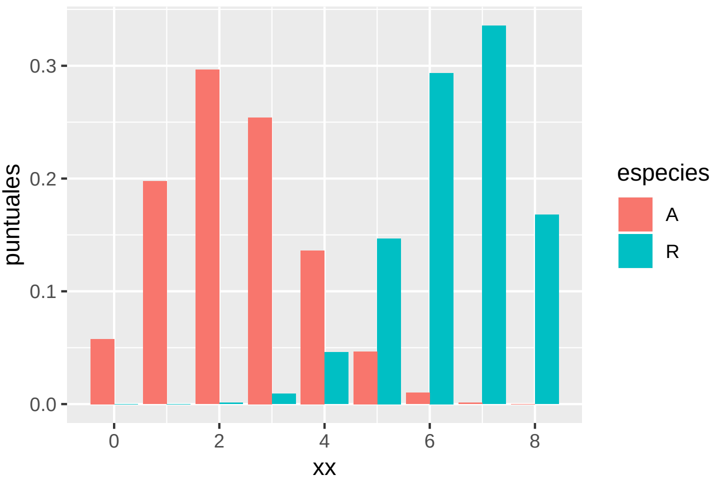
Supongamos que en un nido se colocan $n = 8$ huevos parasitarios.

Sea $X =$ número de huevos removidos. Sabemos,

$$X|_{Y=A} \sim Bi(8, 0.3), \text{ es decir } p_{X|_{Y=A}}(x) = \binom{8}{x} 0.3^x 0.7^{8-x}$$

$$X|_{Y=R} \sim Bi(8, 0.8), \text{ es decir } p_{X|_{Y=R}}(x) = \binom{8}{x} 0.8^x 0.2^{8-x}$$

Distribuciones condicionales



Distribuciones condicionales

Cuál es la distribución de X??

Distribuciones condicionales

Cuál es la distribución de X?? $p_X(x)$?

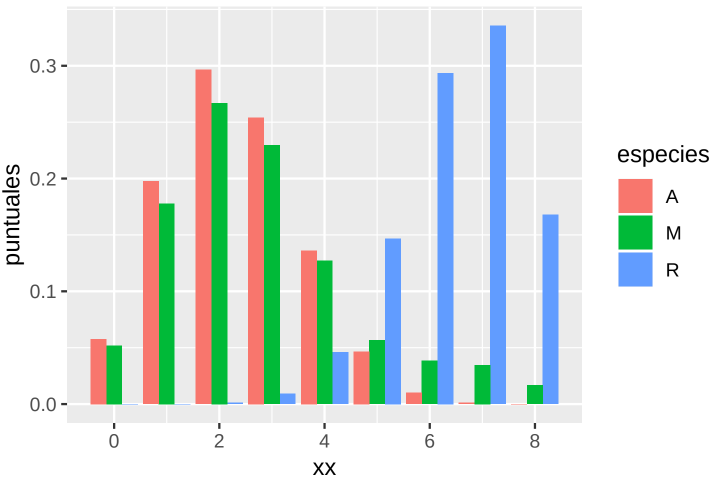
Distribuciones condicionales

Cuál es la distribución de X?? $p_X(x)$?

$$\begin{aligned} p_X(x) &= p_{X|Y=R}(x)P(Y=R) + p_{X|Y=A}(x)P(Y=A) \\ &= 0.1\binom{8}{x}0.8^x 0.2^{8-x} + 0.9\binom{8}{x}0.3^x 0.7^{8-x} \end{aligned}$$

0	1	2	3	4	5	6	7	8
0.052	0.178	0.267	0.230	0.127	0.057	0.038	0.035	0.017

Distribuciones condicionales y puntual



Distribuciones conjunta

$$p_{XY}(x,y) = p_{X|Y=y}(x)P(Y=y)$$

$Y/\ X$	0	1	2	3	4	5	6	7	8
R	0.000	0.000	0.000	0.001	0.005	0.015	0.029	0.034	0.017
A	0.052	0.178	0.267	0.229	0.123	0.042	0.009	0.001	0.000

Pajaritos - El patito feo

Cómo hacemos para decidir mirando un nido con 8 huevos parasitarios y dependiendo cuantos huevos sean removidos si la especie es rechazadora o aceptadora.

Si se remueven 5 huevos; es decir si $X = 5$

Pajaritos - El patito feo

Cómo hacemos para decidir mirando un nido con 8 huevos parasitarios y dependiendo cuantos huevos sean removidos si la especie es rechazadora o aceptadora.

- Si se remueven 5 huevos; es decir si $X = 5$ ¿de qué clase de nido diría que se trata?
- Si se remueven 3 huevos; (X=3) ¿de qué clase de nido diría que se trata?

Queremos construir un Clasificador

Queremos construir un Clasificador

Una Regla (de clasificación) asigna a $x \in \{0, 1, \ldots, 8\}$ un tipo de hospedador: $\{A, R\}$ es decir, buscamos

$$h : \{0, 1, \ldots, 8\} \to \{A, R\}$$

Por ejemplo, si remueve 5 huevos o más es rechazadora

x	0	1	2	3	4	5	6	7	8
clasificador h_1	A	A	A	A	A	R	R	R	R

Queremos construir un Clasificador

Una Regla (de clasificación) asigna a $x \in \{0, 1, \ldots, 8\}$ un tipo de hospedador: $\{A, R\}$ es decir, buscamos

$$h : \{0, 1, \ldots, 8\} \to \{A, R\}$$

Por ejemplo, si remueve 5 huevos o más es rechazadora

x	0	1	2	3	4	5	6	7	8
clasificador h_1	A	A	A	A	A	R	R	R	R

Otro clasificador, si remueve un número par de huevos es rechazadora

x	0	1	2	3	4	5	6	7	8
clasificador h_2	R	A	R	A	R	A	R	A	R

Queremos construir un Clasificador

Queremos construir un Clasificador

Otra manera de escribirlo

x	0	1	2	3	4	5	6	7	8
clasificador h_1	A	A	A	A	A	R	R	R	R

$$h_1(x) = \begin{cases} A & x \in \{0,1,2,3,4\} \\ R & x \in \{5,6,7,8\} \end{cases}$$

Queremos construir un Clasificador

Otra manera de escribirlo

x	0	1	2	3	4	5	6	7	8
clasificador h_1	A	A	A	A	A	R	R	R	R

$$h_1(x) = \begin{cases} A & x \in \{0,1,2,3,4\} \\ R & x \in \{5,6,7,8\} \end{cases}$$

x	0	1	2	3	4	5	6	7	8
clasificador h_2	R	A	R	A	R	A	R	A	R

$$h_2(x) = \begin{cases} A & x \in \{1,3,5,7\} \\ R & x \in \{0,2,4,6,8\} \end{cases}$$

Queremos construir un Clasificador

Otra manera de escribirlo

x	0	1	2	3	4	5	6	7	8
clasificador h_1	A	A	A	A	A	R	R	R	R

$$h_1(x) = \begin{cases} A & x \in \{0,1,2,3,4\} \\ R & x \in \{5,6,7,8\} \end{cases}$$

x	0	1	2	3	4	5	6	7	8
clasificador h_2	R	A	R	A	R	A	R	A	R

$$h_2(x) = \begin{cases} A & x \in \{1,3,5,7\} \\ R & x \in \{0,2,4,6,8\} \end{cases}$$

Cuál es más razonable??

Estimación del error de clasificación de h

Error de Clasificación Medio: $L(h) = \mathbb{P}(h(X) \neq Y)$

Estimación del error de clasificación de h

Error de Clasificación Medio: $L(h) = \mathbb{P}(h(X) \neq Y)$
Por ejemplo con el clasificador h_1,

$$h_1(x) = \begin{cases} A & x \in \{0, 1, 2, 3, 4\} \\ R & x \in \{5, 6, 7, 8\} \end{cases}$$

$P(h_1(X) \neq Y)$

Estimación del error de clasificación de h

Error de Clasificación Medio: $L(h) = \mathbb{P}(h(X) \neq Y)$
Por ejemplo con el clasificador h_1,

$$h_1(x) = \begin{cases} A & x \in \{0, 1, 2, 3, 4\} \\ R & x \in \{5, 6, 7, 8\} \end{cases}$$

$$P(h_1(X) \neq Y) =$$

$$= P(h_1(X) \neq Y | Y = R) P(Y = R) + P(h_1(X) \neq Y | Y = A) P(Y = A)$$

Estimación del error de clasificación de h

Error de Clasificación Medio: $L(h) = \mathbb{P}(h(X) \neq Y)$
Por ejemplo con el clasificador h_1,

$$h_1(x) = \begin{cases} A & x \in \{0,1,2,3,4\} \\ R & x \in \{5,6,7,8\} \end{cases}$$

$$P(h_1(X) \neq Y) =$$
$$= P(h_1(X) \neq Y | Y = R) P(Y = R) + P(h_1(X) \neq Y | Y = A) P(Y = A)$$
$$= P(X \in \{0,1,2,3,4\} | Y = R) P(Y = R) + P(X \in \{5,6,7,8\} | Y = A)$$

Estimación del error de clasificación de h

Error de Clasificación Medio: $L(h) = \mathbb{P}(h(X) \neq Y)$
Por ejemplo con el clasificador h_1,

$$h_1(x) = \begin{cases} A & x \in \{0, 1, 2, 3, 4\} \\ R & x \in \{5, 6, 7, 8\} \end{cases}$$

$$\begin{aligned}
&P(h_1(X) \neq Y) = \\
&= P(h_1(X) \neq Y | Y = R) P(Y = R) + P(h_1(X) \neq Y | Y = A) P(Y = A) \\
&= P(X \in \{0, 1, 2, 3, 4\} | Y = R) P(Y = R) + P(X \in \{5, 6, 7, 8\} | Y = A) \\
&= \sum_{x=0}^{4} \binom{8}{x} 0.8^x 0.2^{8-x} 0.1 + \sum_{x=5}^{8} \binom{8}{x} 0.3^x 0.7^{8-x} 0.9 \\
&= 0.05779905
\end{aligned}$$

Estimación del error de clasificación de h

Error de Clasificación Medio: $L(h) = \mathbb{P}(h(X) \neq Y)$
Por ejemplo con el clasificador h_1,

$$h_1(x) = \begin{cases} A & x \in \{0,1,2,3,4\} \\ R & x \in \{5,6,7,8\} \end{cases}$$

$$\begin{aligned}
&P(h_1(X) \neq Y) = \\
&= P(h_1(X) \neq Y | Y = R) P(Y = R) + P(h_1(X) \neq Y | Y = A) P(Y = A) \\
&= P(X \in \{0,1,2,3,4\} | Y = R) P(Y = R) + P(X \in \{5,6,7,8\} | Y = A) \\
&= \sum_{x=0}^{4} \binom{8}{x} 0.8^x 0.2^{8-x} 0.1 + \sum_{x=5}^{8} \binom{8}{x} 0.3^x 0.7^{8-x} 0.9 \\
&= 0.05779905
\end{aligned}$$

Estimación del error de clasificación de h

Error de Clasificación Medio: $L(h) = \mathbb{P}(h(X) \neq Y)$
Con el clasificador h_2,

$$h_2(x) = \begin{cases} A & x \in \{1,3,5,7\} \\ R & x \in \{0,2,4,6,8\} \end{cases}$$

Estimación del error de clasificación de h

Error de Clasificación Medio: $L(h) = \mathbb{P}(h(X) \neq Y)$
Con el clasificador h_2,

$$h_2(x) = \begin{cases} A & x \in \{1,3,5,7\} \\ R & x \in \{0,2,4,6,8\} \end{cases}$$

$$P(h_2(X) \neq Y) =$$
$$= P(h_2(X) \neq Y | Y = R) P(Y = R) + P(h_2(X) \neq Y | Y = A) P(Y = A)$$

Estimación del error de clasificación de h

Error de Clasificación Medio: $L(h) = \mathbb{P}(h(X) \neq Y)$
Con el clasificador h_2,

$$h_2(x) = \begin{cases} A & x \in \{1,3,5,7\} \\ R & x \in \{0,2,4,6,8\} \end{cases}$$

$$P(h_2(X) \neq Y) =$$
$$= P(h_2(X) \neq Y | Y = R) P(Y = R) + P(h_2(X) \neq Y | Y = A) P(Y = A)$$
$$= P(X \in \{1,3,5,7\} | Y = R) P(Y = R) + P(X \in \{0,2,4,6,8\} | Y = A)$$

Estimación del error de clasificación de h

Error de Clasificación Medio: $L(h) = \mathbb{P}(h(X) \neq Y)$
Con el clasificador h_2,

$$h_2(x) = \begin{cases} A & x \in \{1, 3, 5, 7\} \\ R & x \in \{0, 2, 4, 6, 8\} \end{cases}$$

$$\begin{aligned}
&P(h_2(X) \neq Y) = \\
&= P(h_2(X) \neq Y | Y = R) P(Y = R) + P(h_2(X) \neq Y | Y = A) P(Y = A) \\
&= P(X \in \{1, 3, 5, 7\} | Y = R) P(Y = R) + P(X \in \{0, 2, 4, 6, 8\} | Y = A) \\
&= \sum_{x=1,3,5,7} \binom{8}{x} 0.8^x 0.2^{8-x} 0.1 + \sum_{x=0,2,4,6,8} \binom{8}{x} 0.3^x 0.7^{8-x} 0.9 \\
&= 0.4994551
\end{aligned}$$

Estimación del error de clasificación de h

Error de Clasificación Medio: $L(h) = \mathbb{P}(h(X) \neq Y)$
Con el clasificador h_2,

$$h_2(x) = \begin{cases} A & x \in \{1,3,5,7\} \\ R & x \in \{0,2,4,6,8\} \end{cases}$$

$$\begin{aligned}
&P(h_2(X) \neq Y) = \\
&= P(h_2(X) \neq Y | Y = R) P(Y = R) + P(h_2(X) \neq Y | Y = A) P(Y = A) \\
&= P(X \in \{1,3,5,7\} | Y = R) P(Y = R) + P(X \in \{0,2,4,6,8\} | Y = A) \\
&= \sum_{x=1,3,5,7} \binom{8}{x} 0.8^x 0.2^{8-x} 0.1 + \sum_{x=0,2,4,6,8} \binom{8}{x} 0.3^x 0.7^{8-x} 0.9 \\
&= 0.4994551
\end{aligned}$$

Error de clasificación de h

Error de Clasificación Medio: $L(h) = \mathbb{P}(h(X) \neq Y)$

$$h_1(x) = \begin{cases} A & x \in \{0,1,2,3,4\} \\ R & x \in \{5,6,7,8\} \end{cases}$$

$$P(h_1(X) \neq Y) = .05779905$$

,

$$h_2(x) = \begin{cases} A & x \in \{1,3,5,7\} \\ R & x \in \{0,2,4,6,8\} \end{cases}$$

$$P(h_2(X) \neq Y) = 0.4994551$$

Estimación del error de clasificación de h

Error de Clasificación Medio: $L(h) = \mathbb{P}(h(X) \neq Y)$

Estimación del error de clasificación de h

Error de Clasificación Medio: $L(h) = \mathbb{P}(h(X) \neq Y)$
Por ejemplo con el clasificador h_t,

$$h_t(x) = \begin{cases} A & x \in \{0, \dots t-1\} \\ R & x \in \{t, \dots, 8\} \end{cases}$$

$$P(h_t(X) \neq Y)$$

Estimación del error de clasificación de h

Error de Clasificación Medio: $L(h) = \mathbb{P}(h(X) \neq Y)$
Por ejemplo con el clasificador h_t,

$$h_t(x) = \begin{cases} A & x \in \{0, \ldots t-1\} \\ R & x \in \{t, \ldots, 8\} \end{cases}$$

$$P(h_t(X) \neq Y) =$$
$$= P(h_t(X) \neq Y | Y = R) P(Y = R) + P(h_t(X) \neq Y | Y = A) P(Y = A)$$

Estimación del error de clasificación de h

Error de Clasificación Medio: $L(h) = \mathbb{P}(h(X) \neq Y)$
Por ejemplo con el clasificador h_t,

$$h_t(x) = \begin{cases} A & x \in \{0, \ldots t-1\} \\ R & x \in \{t, \ldots, 8\} \end{cases}$$

$$
\begin{aligned}
&P(h_t(X) \neq Y) = \\
&= P(h_t(X) \neq Y | Y = R) P(Y = R) + P(h_t(X) \neq Y | Y = A) P(Y = A) \\
&= P(X \in \{0, \ldots t-1\} | Y = R) P(Y = R) + P(X \in \{t, \ldots, 8\} | Y = A)
\end{aligned}
$$

Estimación del error de clasificación de h

Error de Clasificación Medio: $L(h) = \mathbb{P}(h(X) \neq Y)$
Por ejemplo con el clasificador h_t,

$$h_t(x) = \begin{cases} A & x \in \{0, \dots t-1\} \\ R & x \in \{t, \dots, 8\} \end{cases}$$

$$\begin{aligned}
&P(h_t(X) \neq Y) = \\
&= P(h_t(X) \neq Y | Y = R) P(Y = R) + P(h_t(X) \neq Y | Y = A) P(Y = A) \\
&= P(X \in \{0, \dots t-1\} | Y = R) P(Y = R) + P(X \in \{t, \dots, 8\} | Y = A) \\
&= \sum_{x=0}^{t-1} \binom{8}{x} 0.8^x 0.2^{8-x} 0.1 + \sum_{x=t}^{8} \binom{8}{x} 0.3^x 0.7^{8-x} 0.9
\end{aligned}$$

Estimación del error de clasificación de h

$$h_t(x) = \begin{cases} A & x \in \{0, \dots t-1\} \\ R & x \in \{t, \dots, 8\} \end{cases}$$

$$P(h_t(X) \neq Y) = \sum_{x=0}^{t-1} \binom{8}{x} 0.8^x 0.2^{8-x} 0.1 + \sum_{x=t}^{8} \binom{8}{x} 0.3^x 0.7^{8-x} 0.9$$

X	0	1	2	3	4
$P(h_t(X) \neq Y)$	0.9000	0.8481	0.6702	0.4035	0.1757

X	5	6	7	8	todos A
$P(h_t(X) \neq Y)$	0.0578	0.0305	0.0508	0.0833	0.100

Estimación del error de clasificación de h

$$h_t(x) = \begin{cases} A & x \in \{0, \dots t-1\} \\ R & x \in \{t, \dots, 8\} \end{cases}$$

$$P(h_t(X) \neq Y) = \sum_{x=0}^{t-1} \binom{8}{x} 0.8^x 0.2^{8-x} 0.1 + \sum_{x=t}^{8} \binom{8}{x} 0.3^x 0.7^{8-x} 0.9$$

X	0	1	2	3	4
$P(h_t(X) \neq Y)$	0.9000	0.8481	0.6702	0.4035	0.1757

X	5	6	7	8	todos A
$P(h_t(X) \neq Y)$	0.0578	0.0305	0.0508	0.0833	0.100

Estimación del error de clasificación de h

$$h_6(x) = \begin{cases} A & x \in \{0, \dots 5\} \\ R & x \in \{6, \dots, 8\} \end{cases}$$

$$P(h_6(X) \neq Y) = \sum_{x=0}^{5} \binom{8}{x} 0.8^x 0.2^{8-x} 0.1 + \sum_{x=6}^{8} \binom{8}{x} 0.3^x 0.7^{8-x} 0.9$$

X	6
$P(h_t(X) \neq Y)$	0.0305

Clasificación: Marco Teórico

- $X \in \mathcal{X}$, $Y \in \mathcal{Y}$. (en nuestro caso $\mathcal{X} = \{0, 1, \ldots, 8\}$ y $\mathcal{Y} = \{A, R\}$)
- (X, Y) vector aleatorio, con puntual p_{XY}.
- Clasificador: Regla (de clasificación) que asigna a cada $x \in \mathcal{X}$ un elemento $y \in \mathcal{Y}$

$$\text{Clasificador} \quad h : \mathcal{X} \to \mathcal{Y}$$

- Error de Clasificación Medio del clasificador h

$$L(h) = \mathbb{P}(h(X) \neq Y)$$

- Objetivo (teórico): Encontrar h que minimice el error medio de clasificación.

Clasificador Óptimo

Teorema: El clasificador definido como

$$h_{opt}(x) = \begin{cases} A & x \in \{z : P(Y = A|X = z) \geq P(Y = R|X = z)\} \\ R & x \in \{z : P(Y = R|X = z) > P(Y = A|X = z)\} \end{cases}$$

es óptimo en el sentido que minimiza $L(h) = \mathbb{P}(h(X) \neq Y)$.

Clasificador Óptimo

Teorema: El clasificador definido como

$$h_{opt}(x) = \begin{cases} A & x \in \{z : P(Y = A|X = z) \geq P(Y = R|X = z)\} \\ R & x \in \{z : P(Y = R|X = z) > P(Y = A|X = z)\} \end{cases}$$

es óptimo en el sentido que minimiza $L(h) = \mathbb{P}(h(X) \neq Y)$.

Demostración: Primero notemos que

$$\begin{aligned} h_{opt}(x) &= \begin{cases} A & x \in \{z : p_{Y|X=z}(A) \geq p_{Y|X=z}(R)\} \\ R & x \in \{z : p_{Y|X=z}(R) > p_{Y|X=z}(A)\} \end{cases} \\ &= \begin{cases} A & x \in \{z : \frac{p_{X|Y=A}(z)p_Y(A)}{p_X(z)} \geq \frac{p_{X|Y=R}(z)p_Y(R)}{p_X(z)}\} \\ R & x \in \{z : \frac{p_{X|Y=R}(z)p_Y(R)}{p_X(z)} > \frac{p_{X|Y=A}(z)p_Y(A)}{p_X(z)}\} \end{cases} \end{aligned}$$

Clasificador Óptimo

Luego

$$h_{opt}(x) \quad = \quad \begin{cases} A & x \in \{z : p_{X|Y=A}(z) p_Y(A) \geq p_{X|Y=R}(z) p_Y(R)\} \\ R & x \in \{z : p_{X|Y=R}(z) p_Y(R) > p_{X|Y=A}(z) p_Y(A)\} \end{cases}$$

Clasificador Óptimo

Luego

$$h_{opt}(x) = \begin{cases} A & x \in \{z : p_{X|Y=A}(z)p_Y(A) \geq p_{X|Y=R}(z)p_Y(R)\} \\ R & x \in \{z : p_{X|Y=R}(z)p_Y(R) > p_{X|Y=A}(z)p_Y(A)\} \end{cases}$$

Notemos que cualquier clasificador es de la forma

$$h(x) = \begin{cases} A & x \in C_A \\ R & x \in C_R \end{cases}$$

Entonces debemos probar que si tomamos
$C_A = \{z : p_{X|Y=A}(z)p_Y(A) \geq p_{X|Y=R}(z)p_Y(R)\}$ minimiza $L(h)$.

Clasificador Óptimo

Luego

$$h_{opt}(x) \;=\; \begin{cases} A & x \in \{z : p_{X|Y=A}(z)p_Y(A) \geq p_{X|Y=R}(z)p_Y(R)\} \\ R & x \in \{z : p_{X|Y=R}(z)p_Y(R) > p_{X|Y=A}(z)p_Y(A)\} \end{cases}$$

Notemos que cualquier clasificador es de la forma

$$h(x) = \begin{cases} A & x \in C_A \\ R & x \in C_R \end{cases}$$

Entonces debemos probar que si tomamos
$C_A = \{z : p_{X|Y=A}(z)p_Y(A) \geq p_{X|Y=R}(z)p_Y(R)\}$ minimiza $L(h)$.

$$L(h) \;=\; \mathbb{P}(h(X) \neq Y) = \sum_{x \in C_A} p_{X|Y=R}(z)p_Y(R) + \sum_{x \in C_R} p_{X|Y=A}(z)p_Y(A)$$

Clasificador Óptimo

Luego

$$h_{opt}(x) \quad = \quad \begin{cases} A & x \in \{z : p_{X|Y=A}(z)p_Y(A) \geq p_{X|Y=R}(z)p_Y(R)\} \\ R & x \in \{z : p_{X|Y=R}(z)p_Y(R) > p_{X|Y=A}(z)p_Y(A)\} \end{cases}$$

Notemos que cualquier clasificador es de la forma

$$h(x) = \begin{cases} A & x \in C_A \\ R & x \in C_R \end{cases}$$

Entonces debemos probar que si tomamos
$C_A = \{z : p_{X|Y=A}(z)p_Y(A) \geq p_{X|Y=R}(z)p_Y(R)\}$ minimiza $L(h)$.

$$\begin{aligned} L(h) &= \mathbb{P}(h(X) \neq Y) = \sum_{x \in C_A} p_{X|Y=R}(z)p_Y(R) + \sum_{x \in C_R} p_{X|Y=A}(z)p_Y(A) \\ &= \sum_{x \in C_A} p_{X|Y=R}(z)p_Y(R) + \sum_{x \in C_R} p_{X|Y=A}(z)p_Y(A) \\ &+ \sum_{x \in C_A} p_{X|Y=A}(z)p_Y(A) - \sum_{x \in C_A} p_{X|Y=A}(z)p_Y(A) \end{aligned}$$

Clasificador Óptimo

$$
\begin{aligned}
L(h) &= \sum_{x \in C_A} p_{X|Y=R}(z) p_Y(R) + \sum_{x \in C_R} p_{X|Y=A}(z) p_Y(A) \\
&+ \sum_{x \in C_A} p_{X|Y=A}(z) p_Y(A) - \sum_{x \in C_A} p_{X|Y=A}(z) p_Y(A)
\end{aligned}
$$

Clasificador Óptimo

$$
\begin{aligned}
L(h) &= \sum_{x \in C_A} p_{X|Y=R}(z) p_Y(R) + \sum_{x \in C_R} p_{X|Y=A}(z) p_Y(A) \\
&+ \sum_{x \in C_A} p_{X|Y=A}(z) p_Y(A) - \sum_{x \in C_A} p_{X|Y=A}(z) p_Y(A) \\
&= \sum_{x \in C_A} \left(p_{X|Y=R}(z) p_Y(R) - p_{X|Y=A}(z) p_Y(A)\right) \\
&+ \sum_{x \in C_R} p_{X|Y=A}(z) p_Y(A) + \sum_{x \in C_A} p_{X|Y=A}(z) p_Y(A) \\
&= \sum_{x \in C_A} \left(p_{X|Y=R}(z) p_Y(R) - p_{X|Y=A}(z) p_Y(A)\right) + p_Y(A)
\end{aligned}
$$

Entonces $L(h)$ es mínimo si para todo $x \in C_A$ se tiene que $(p_{X|Y=R}(z) p_Y(R) - p_{X|Y=A}(z) p_Y(A)) < 0$. Pero justamente

$$
h_{opt}(x) = \begin{cases} A & x \in \{z : p_{X|Y=A}(z) p_Y(A) \geq p_{X|Y=R}(z) p_Y(R)\} \\ R & x \in \{z : p_{X|Y=R}(z) p_Y(R) > p_{X|Y=A}(z) p_Y(A)\} \end{cases}
$$

Clasificador Óptimo: Pátito feo

$$h_{opt}(x) \quad = \quad \begin{cases} A & x \in \{z : p_{X|Y=A}(z)p_Y(A) \geq p_{X|Y=R}(z)p_Y(R)\} \\ R & x \in \{z : p_{X|Y=R}(z)p_Y(R) > p_{X|Y=A}(z)p_Y(A)\} \end{cases}$$

Clasificador Óptimo: Pátito feo

$$h_{opt}(x) = \begin{cases} A & x \in \{z : p_{X|Y=A}(z)p_Y(A) \geq p_{X|Y=R}(z)p_Y(R)\} \\ R & x \in \{z : p_{X|Y=R}(z)p_Y(R) > p_{X|Y=A}(z)p_Y(A)\} \end{cases}$$

Queremos comprar $p_{X|Y=A}(z)p_Y(A)$ vs $p_{X|Y=R}(z)p_Y(R)$. Que en este caso es como mirar la conjunta $p_{XY}(x,A)$ vs $p_{XY}(x,R)$

$Y/\ X$	0	1	2	3	4	5	6	7	8
R	0.000	0.000	0.000	0.001	0.005	0.015	0.029	0.034	0.017
A	0.052	0.178	0.267	0.229	0.123	0.042	0.009	0.001	0.000

$$h_{opt}(x) = \begin{cases} A & x \in \{0,1,2,3,4,5\} \\ R & x \in \{6,7,8\} \end{cases}$$

$$L(h_{opt}) = \mathbb{P}(h_{opt}(X) \neq Y) = 0.0305$$