5. Decisores de Máxima Verosimilitud (ML) y de Máximo a Posteriori (MAP)

February 21, 2019

I.5 Decisores ML y MAP

Supongamos ahora que se establece una partición sobre el espacio muestral Ω, $B_i \in \mathscr{P}(\Omega)\, i = 1 \dots K$. Consideremos que **acontence la observación** del suceso A, para el que conocemos las **probabilidades de transición** $P((A/B_i)$.

Nos preguntamos qué suceso condicionante ha causado la observación:

- **Principio de Máxima Verosimilitud (ML)**: en ausencia de más información, elegimos el suceso B_i maximiza la verosimilitud $L(B_i) = P(A/B_i)$.
- **Principio de Máximo *a Posteriori* (MAP)**: si, además de las probabilidades de transición, se conocen las probabilidades *a priori*, pueden calcularse las probabilidasdes *a posteri* con el Teorema de Bayes y el de la Probabilidad Total. En tal caso, elegimos el suceso B_i que maximiza la probabilidad a posteriori.

Consideremos de nuevo el caso del lanzamiento de un dado, y que el suceso observado es el $A = \{1, 2, 4, 6\}$. Si el dado no está trucado, ya hemos visto cuáles son las probabilidades de transición de la observación por que el lanzamiento sea divisible o no por tres, obteniéndose directamente las verosimilitudes de ambas posibilidades:

- $L(div3) \equiv P(A/div3) = \frac{1}{2}$
- $L(nodiv3) \equiv P(A/nodiv3) = \frac{3}{4}$

Como $L(nodiv3) > L(div3)$, el **principio de máxima verosimilitud (ML)** conduce a seleccionar el suceso *nodiv*3 como **causante más verosímil** de la observación. Adviértase que $L(nodiv3) + L(div3) \neq 1$.

Si conocemos las probabilidades *a priori*, $P(div3) = {}^1/_3$ y $P(nodiv3) = {}^2/_3$, ya hemos visto que el Teorema de la Probabilidad Total permite calcular $P(A) = {}^2/_3$. Ahora **podemos calcular las probabilidades *a posteriori* haciendo uso del Teorema de Bayes**:

$$P(div3/A) = \frac{P(A/div3)P(div3)}{P(A)} = \frac{{}^1/_2\,{}^1/_3}{{}^2/_3} = \frac{1}{4}$$

$$P(nodiv3/A) = \frac{P(A/nodiv3)P(nodiv3)}{P(A)} = \frac{{}^3/_4\,{}^2/_3}{{}^2/_3} = \frac{3}{4}$$

Como $P(nodiv3/A) > P(div3/A)$, el **principio de máximo *a posteriori* (MAP)** conduce a seleccionar el suceso *nodiv*3 como **causante más probable** de la observación. Adviértase que $P(nodiv3/A) + P(div3/A) = 1$.

En el ejemplo que acabamos de ver, el decisor de máxima verosimilitud (MAP) y el de máximo a posteriori (ML) proporcionan dan lugar a la misma toma de decisión: seleccionar como causa el suceso no divisible por tres.

Sin embargo, es posible que ambos decisores den resultados diferentes. **El decisor MAP incorpora información *a priori*, de la que no dispone el decisor ML**, que puede resultar de suma utilidad. Veámoslo con un nuevo ejemplo:

Consideremos el lanzamiento del dado bueno, pero ahora la observación es el suceso $A = \{3, 4, 5\}$

Las verosimilitudes ahora resultan:

$$L(div3) \equiv P(A/div3) = \frac{P(\{3,4,5\} \cap \{3,6\})}{P(\{3,6\})} = \frac{^1/_6}{^1/_3} = \frac{1}{2}$$

$$L(nodiv3) \equiv P(A/nodiv3) = \frac{P(\{3,4,5\} \cap \{1,2,4,5\})}{P(\{1,2,4,5\})} = \frac{^1/_3}{^2/_3} = \frac{1}{2}$$

Dado que $L(div3) = L(nodiv3)$ el decisor ML no nos sirve.

Si conocemos las probabilidades *a priori*, $P(div3) = {}^1/_3$ y $P(nodiv3) = {}^2/_3$, el Teorema de la Probabilidad Total permite calcular $P(A)$:

$$P(A) = P(A/div3)P(div3) + P(A/nodiv3)P(nodiv3) = \frac{1}{2}\frac{1}{3} + \frac{1}{2}\frac{2}{3} = \frac{1}{2}$$

Y calculamos las probabilidades *a posteriori* con el Teorema de Bayes:

$$P(div3/A) = \frac{P(A/div3)P(div3)}{P(A)} = \frac{^1/_2\, ^1/_3}{^1/_2} = \frac{1}{3}$$

$$P(nodiv3/A) = \frac{P(A/nodiv3)P(nodiv3)}{P(A)} = \frac{^1/_2\, ^2/_3}{^1/_2} = \frac{2}{3}$$

Resulta que $P(nodiv3/A) > P(div3/A)$, por lo que es más probable que la causa de la observación haya sido el suceso no divisible por tres.

£Qué ha pasado? * Si el suceso condicionante es divisible por tres, la observación A se satisface la mitad de las veces, esto es, siempre que salga 3 (1 caso favorable de 2 posibles). Igualmente, si el suceso condicionante no es divisible por tres, la observación A se satisface la mitad de las veces, esto es, siempre que salga 4 ó 5 (2 casos favorables de 4 posibles). Por tanto, **las verosimilitudes son idénticas**.

- Sin embargo, es más probable que resultado sea no divisible por tres (4 casos favorables de 6 posibles) a que sea divisible por tres (2 casos favorables de 6 posibles). **El decisor MAP considera las diferentes probabilidades *a priori*, ponderando aquellos resultados en que éstas son mayores. Ello lo realiza a través del Teorema de Bayes al multiplicar las verosimilitudes por las probabilidades *a priori***. La división por $P(A)$ es una normalización, para que resulte una probabilidad válida.

Por ello, si las probabilidades *a priori* son iguales, el decisor MAP no puede mejorar el resultado del ML. Pero si son distintas, aprovecha la información disponible *a priori*. Podemos por ello considerar que el decisor ML es un decisor MAP en el que no tenemos información *a priori*, situación que se modela asignando idénticas probabilidades *a priori* a todos los sucesos condicionantes.