Problemas con el uso de la inferencia frecuentista para el clasificador bayes ingenuo

Tenemos la siguiente tabla dónde anteriormente hemos estimado las probabilidades de ocurrencia de cada palabra con frecuencias en la muestra.

Palabra	P(palabra|basura)	P(palabra|normal)
mándanos	1/3	1/2
tu	2/3	2/2
clave	2/3	0/2
coche	1/3	2/2
revisión	2/3	1/2
de	1/3	1/2

Supongamos que recibamos el siguiente mensaje: ``mándanos tu clave de coche''. ¿Cómo lo clasificaríamos?

P(mándanos tu clave de coche|normal) = 0 y luego tenemos que P(normal|mándanos tu clave de coche) = 0 también. Cualquier mensaje que contiene la palabra ``clave'' sería clasificado como correo basura automáticamente.

Clasificación bayesiana

- Usando los métodos bayesianos, las probabilidades desconocidas son variables:

$$\Theta_b = P(\text{basura}),\ \Theta_n = 1 - \Theta_b = P(\text{normal}).$$

$$\Theta_{mb} = P(\text{mándanos}|\text{basura}),\ \Theta_{mn} = P(\text{mándanos}|\text{normal}),\ \ldots$$

- Cada parámetro tiene su distribución a priori, por ejemplo, una distribución uniforme:

$$f(\Theta_b) = 1,\ f(\Theta_{mb}) = 1,\ f(\Theta_{mn}) = 1.$$

- Dados los datos, usamos el teorema de Bayes para modificar las probabilidades.

  En el ejemplo había 3 mensajes basura y 2 mensajes normales. Luego:

$$f(\Theta_b|\text{datos}) \propto f(\Theta_b)\, f(\text{datos}|\Theta_b) \propto \Theta_b^3(1-\Theta_b)^2$$

  que es una distribución beta: Beta(4,3).

  De manera parecida, tenemos $\Theta_{mb}|\text{datos} \sim \text{Beta}(2,3)$ y $\Theta_{mn}|\text{datos} \sim \text{Beta}(2,2)$.

- Luego podemos calcular estimaciones puntuales bayesianas utilizando las medias a posteriori:

P(basura|datos) = E[Θ_b|datos] = 4/7, P(normal|datos) = 3/7,
P(mándanos|basura, datos) = 2/5,
P(mándanos|normal, datos) = 2/4 = 1/2.

Se muestran las estimaciones de las probabilidades marginales para todas las palabras en la siguiente tabla.

Palabra	P(palabra|basura)	P(palabra|normal)
mándanos	$2/5$	$1/2$
tu	$3/7$	$3/4$
clave	$3/5$	$1/4$
coche	$2/5$	$3/4$
revisión	$3/5$	$1/2$
de	$2/5$	$1/2$

- Por último, se puede utilizar el teorema de Bayes para clasificar el nuevo mensaje como basura o no:

P(basura|mensaje) = P(mensaje|basura)P(basura)/P(mensaje)

donde el denominador es

P(mensaje) = P(mensaje|basura)P(basura)+P(mensaje|normal)P(normal)

$$= \frac{2}{5}\frac{3}{7}\frac{3}{5}\frac{2}{5}\left(1-\frac{3}{5}\right)\frac{2}{5} * \frac{4}{7} + \frac{1}{2}\frac{3}{4}\frac{1}{4}\frac{3}{4}\left(1-\frac{1}{2}\right)\frac{1}{2} * \frac{3}{7} \approx 0.0169$$

Luego P(basura|mensaje) = $\frac{\frac{2}{5}\frac{3}{7}\frac{3}{5}\frac{2}{5}\left(1-\frac{3}{5}\right)\frac{2}{5} * \frac{4}{7}}{0.0169}$ = 0.5552.