UNIVERSIDAD DE SANTIAGO DE C
FACULTAD DE INGENIERÍA
DEPTO. DE INGENIERÍA INFORMA

224
322
93

PEP 2
Análisis de Datos

Prof: Max Chacón
junio 2022

1. El profesor de electro-fisiología cognitiva de la U de Mancheter Mastrit Merek, estudia los trastornos bipolares mediante experimentos de juegos de azar. Para modelar estos trastornos plantea un modelo de toma de decisiones Bayesiano. Aquí las probabilidades a priori corresponden a las expectativas que tiene los sujetos de ganar y las probabilidades a posteriori es lo que determina finalmente si el sujeto apostará o no. Para probar su hipótesis genera un experimento con dos grupos de sujetos (bipolares y sanos), los que realizan apuestas en un juego al azar, según el nivel de dinero apostado. Del experimento se obtienen las siguientes probabilidades a priori en función del dinero apostado para sujetos sanos como: $Ps = \left(1 - \frac{d}{1020}\right)$, las probabilidades a priori para los bipolares son: $Pb = \left(1 - \frac{d}{1990}\right)$. Además se tiene la siguiente tabla de verosimilitudes para ambos grupos:

Nivel de la apuesta (*d*)	Verosimilitud Bipolares P(d/b)	Verosimilitud Sanos P(d/s)
£ 1000	0.9	0,94
£ 900	0.93	0,85
£ 800	0,87	0,89
£ 700	0,86	0,86
£ 500	0,95	0,95
£ 400	0,91	0,88
£ 300	0,89	0,87
£ 200	0,87	0,89
£ 100	0,92	0,91

Deduzca primero, en literales, las probabilidades que se requieren para determinar la decisión de apostar para cada nivel de apuesta.
Usando éstas fórmulas, determine si hay diferencias entre bipolares y sanos. Si estas diferencias existen, indique a su juicio cuales son las causas de estas diferencias, en términos de las expectativas de los sujetos. (1.8)

2. Para evaluar la diabetes Mellitus existen dos variables de interés, el nivel sobrepeso y el nivel de glucosa en la sangre del paciente. De un hospital se tiene la siguiente tabla dada a continuación.

Usando un clasificador Bayesiano ingenuo, determine si un paciente obeso con 90 mg/ml de glucosa en sangre, será diabético o no.

Peso	Glucosa [mg/ml]	Diabético
Sobrepeso	90	No
Normal	110	No
Obeso	70	No
Sobrepeso	120	Si
Sobrepeso	130	Si
Obeso	80	No
Normal	140	No
Obeso	100	Si
Normal	150	Si
Normal	160	Si

(2)

3. La clasificación general de las bacterias es de acuerdo a su Gram, Gram positivo G(+) y Gram negativo G(-). Para determinar la influencia de cada tipo de Gram en la infección (*I*) se cuenta con las siguientes probabilidades. La probabilidad de tener infección es de 40%, la probabilidad de que el Gram sea positivo es de 30% y la probabilidad de que el Gram sea negativo es de 40%. Además, para el análisis se calcularon las siguientes probabilidades conjuntas:

$P(I \cap \overline{G(+)}) = 2/5$; $P(I \cap G(+)) = 1/5$; $P(I \cap \overline{G(-)}) = 6/10$; $P(I \cap G(-)) = 0$

Usando el método de árboles de decisión, determine cuál es el Gram más adecuado para detectar infección.

(2,2)

Formulas

- Ingenuo: $p(c_i / a_i) = p(c_i)\prod_j p(a_j / c_i)$; para a_i=x continua $p(x / c_i) = e^{-\frac{(x-\bar{x})^2}{2\sigma^2}} \Big/ \sigma\sqrt{2\pi}$

- Árboles de decisión: $Ganancia(V) = -\sum_i p(c_i) ldp(c_i) + \sum_j p(v_j) \sum_i p(c_i / v_j) ldp(c_i / v_j)$;

$RazonGanancia(V) = -Ganancia(V) \Big/ \sum_j p(v_j) ldp(v_j)$.

1) a priori = expectativas de ganar → aumento
a posteriori = Determina finalmente = verosimilitud × a priori

Nivel Apuesta	A priori Bipolares P	A priori sanos P	A posterior Bipolares P	A posterior sanos P
1000	0,50	0,02	0,45	0,02
900	0,55	0,12	0,51	0,10
800	0,60	0,22	0,52	0,20
700	0,65	0,31	0,56	0,27
500	0,75	0,51	0,71	0,49
400	0,80	0,61	0,73	0,54
300	0,85	0,71	0,77	0,62
200	0,90	0,80	0,78	0,71
100	0,95	0,90	0,87	0,82

0,1 0,1 0,2 y que hace con 0,2 0,2

a) p(Decision Apostar / Bipolares) = p(decisión apostar) × p(bipolares/ decisión apostar)
↳ a posteriori ↳ A priori ↳ verosimilitud 0,1

p(Decision Apostar / sanos) = p(decisión apostar) × p(sanos / decisión apostar)
↳ aposteriori 0,2

por lo tanto para cada nivel de apuesta se debe sacar la probabilidad a posteriori a partir de la multiplicación entre su verosimilitud y su valor a priori 0,2

b) ~~Las diferencias existentes recaen en las expectativas de cada grupo de sujeto, donde los bipolares tienen mayor expectativa que los sanos en todos los niveles, además los bipolares determinan finalmente en apostar en todos los niveles de apuesta.~~

No lo decía en el nivel si fue apuesta o no

tot 1.1

2) paciente obeso con 90 mg/mL

A priori

$p(\text{Diabético} = \text{sí}) = \frac{5}{10} = \frac{1}{2}$

$p(\text{Diabético} = \text{No}) = \frac{5}{10} = \frac{1}{2}$ 0,2

verosimilitud

$p(\text{Peso} = \text{Obeso} \mid \text{Diabético} = \text{sí}) = \frac{1}{5}$ 0,2

$p(\text{Peso} = \text{obeso} \mid \text{Diabético} = \text{No}) = \frac{2}{5}$ 0,2

$p(\text{Glucosa} \mid \text{Diabético} = \text{sí}) =$

Donde glucosa $= \{100, 120, 130, 150, 160\}$

$\mu = 132$

$\sigma = 23{,}87$ 0,2

$\therefore \quad p(\text{Glucosa} \mid \text{Diabético} = \text{sí}) = \dfrac{e^{\frac{-(x-132)^2}{2 \times 23{,}87^2}}}{23{,}87 \times \sqrt{2\pi}}$

$p(\text{Glucosa} \mid \text{Diabético} = \text{No}) =$

Donde glucosa $= \{90, 110, 70, 80, 140\}$

$\mu = 98$

$\sigma = 27{,}75$ 0,2

$\therefore \quad p(\text{Glucosa} \mid \text{Diabético} = \text{No}) = \dfrac{e^{\frac{-(x-98)^2}{2 \times 27{,}75^2}}}{27{,}75 \times \sqrt{2\pi}}$

aposteriori

para diabético Sí.

$P(\text{Diabético} = \text{sí} \mid \text{peso} = \text{obeso}, \text{glucosa} = 90) = p(\text{Diabético} = \text{sí}) \times p(\text{glucosa} \mid \text{Diabético} = \text{sí}) \times p(\text{peso} = \text{obeso} \mid \text{Diabético} = \text{sí})$

$= \frac{1}{2} \times \dfrac{e^{\frac{-(90-132)^2}{2 \times 23{,}87^2}}}{23{,}87 \times \sqrt{2\pi}} \times \frac{1}{5} = 0{,}00036$ 0,2

0,2

Para diabético NO

$p(\text{Diabético} = \text{No} \mid \text{peso} = \text{obeso}, \text{glucosa} = 90) = p(\text{Diabético} = \text{No}) \times p(\text{glucosa} \mid \text{Diabético} = \text{No}) \times p(\text{peso} = \text{obeso} \mid \text{Diabético} = \text{No})$

$= \frac{1}{2} \times \dfrac{e^{\frac{-(90-98)^2}{2 \times 27{,}75^2}}}{27{,}75 \times \sqrt{2\pi}} \times \frac{2}{5} = 0{,}0028$ 0,4

0,2

como p(Diabético = Sí / glucosa=90, peso=obeso) es menor a
p(Diabético = No / glucosa=90, peso=obeso).
Se obtiene que para una persona obesa con 90 mg/mL de glucosa en la sangre no es diabético. 0,2

tot 2,0

3. Clase = Infección (I)

$p(\text{infección} = \text{sí}) = 0,4 \rightarrow p(\text{infección} = \text{NO}) = 0,6$

$p(\text{Gram}(+)) = 0,3$; $P(\overline{\text{Gram}(+)}) = 0,7$

$p(\text{Gram}(-)) = 0,4$; $P(\overline{\text{Gram}(-)}) = 0,6$ 0,3

$P(I \cap \overline{G}(+)) = \frac{2}{5}$

$P(I \cap G(+)) = \frac{1}{5}$

$P(I \cap \overline{G}(-)) = \frac{6}{10}$

$P(I \cap G(-)) = 0$

$p(B/A) = \frac{P(A \cap B)}{P(A)}$

$P(A/B) = \frac{P(A \cap B)}{P(B)}$

- $P(I / \overline{G}(+)) = \frac{P(I \cap \overline{G}(+))}{P(\overline{G}(+))} = \frac{\frac{2}{5}}{0,7} = \frac{4}{7}$ ✓
- $P(I / G(+)) = \frac{P(I \cap G(+))}{P(G(+))} = \frac{1/5}{0,3} = \frac{2}{3}$ ✓
- $P(I / G(-)) = 0$
- $P(I / \overline{G(-)})\ \frac{P(I \cap \overline{G(-)})}{P(\overline{G(-)})} = \frac{6/10}{0,6} = 1$
- $P(\overline{I} / G(+)) = 1 - P(I / G(+)) = 1/3$ ✓
- $P(\overline{I} / \overline{G(+)}) = 1 - P(I / \overline{G(+)}) = 3/7$
- $P(\overline{I} / G(-)) = 1 - 0 = 1$
- $P(\overline{I} / \overline{G(-)}) = 1 - P(I / \overline{G(-)}) = 0$

1,2

$\inf(\text{infección}) = -(0,4 \log_2(0,4) + 0,6 \log_2(0,6)) = 0,971$ 0,1

$\inf(G(+)) = -0,3\left(\frac{2}{3}\log_2\left(\frac{2}{3}\right) + \frac{1}{3}\log_2\left(\frac{1}{3}\right)\right) - 0,7\left(\frac{4}{7}\log_2\left(\frac{4}{7}\right) + \frac{3}{7}\log_2\left(\frac{3}{7}\right)\right)$

$= 0,28 + 0,69 = 0,97$ 0,3

$\text{Ganancia}(G(+)) = \inf(C) - \inf(G(+)) = 0,971 - 0,970 = 0,001$

$inf(G(-)) = -0{,}4\left(0\log_2(0) + 1\log_2(1)\right) - 0{,}6\left(0\log_2(0) + 1\log_2(1)\right)$

$= 0$

0,3

Ganancia $(G(-)) = inf(infeccion) - inf(G(-)) = 0{,}971$

∴ la más adecuada es $G(-)$.

just 2,2