Sistemas Inteligentes II

Tema 2. Redes bayesianas

Repaso de probabilidad (I)

Ejercicios curiosos de probabilidad

Probabilidad y oposiciones

En un examen de oposición de 100 temas, el candidato debe desarrollar un tema a elegir entre cuatro sorteados al azar. ¿Cuál es el número mínimo de temas que tiene que estudiar para garantizar que la probabilidad de que se sepa al menos uno de los cuatro temas sorteados sea de al menos un 75%?

Opción A: Más de 75 temas

Opción B: Entre 50 y 75 temas

Opción C: Entre 25 y 50 temas

Opción D: Menos de 25 temas

Probabilidad y oposiciones

En un examen de oposición de 100 temas, el candidato debe desarrollar un tema que elige entre cuatro sorteados al azar. ¿Cuál es el número mínimo de temas que tiene que estudiar para que la probabilidad de que se sepa al menos uno de los cuatro temas sorteados sea de al menos un 75%?

Sea x el número de temas que estudia.

P(saber al menos un tema) = 1- P (no saber ningún tema)

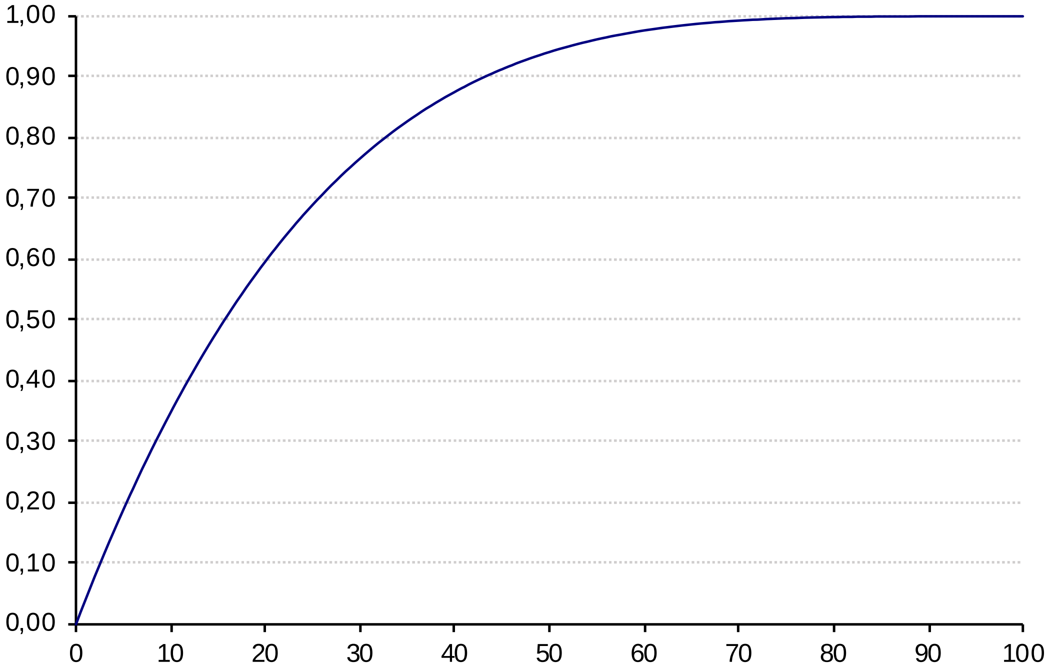
P (no saber ningún tema) = $\frac{100-x}{100} \frac{99-x}{99} \frac{98-x}{98} \frac{97-x}{97}$

Hay que encontrar un x entero tal que

p(saber al menos un tema) >= 0.75

En este caso, **x= 29 temas**.

Gráficamente



Simulador para saber el número de temas que estudiar

El problema del cumpleaños

Calcular la probabilidad de que en una clase de 40 personas al menos dos de ellas cumplan el mismo día.

Opción A: Mayor que 0.8

Opción B: Entre 0.5 y 0.8

Opción C: Entre 0.3 y 0.5

Opción D: Menor que 0.3

El problema del cumpleaños

Calcular la probabilidad de que en una clase de 40 personas al menos dos de ellas cumplan el mismo día.

P(al menos dos cumplan el mismo día) =

1 – P(ninguna cumpla el mismo día) =

$$1 - \frac{364}{365}\frac{363}{365}\ldots\frac{326}{365} \sim 0.9$$

Repaso de probabilidad (II)

Definición de **probabilidad condicionada**

$$P(X/Y) = \frac{P(X,Y)}{P(Y)}$$

Dos variables X e Y son **independientes** si se tiene que

$$P(X/Y) = P(X)$$

Caracterización de la independencia

X e Y son independientes sí y sólo sí $P(X,Y) = P(X) \cdot P(Y)$

Teorema de Bayes (teorema de inversión)

$$P(Y/X) = \frac{P(X/Y)P(Y)}{P(X)}$$

Ley de probabilidad total

Sean $Y_1, \ldots, Y_n$ un conjunto de valores exhaustivo y excluyente. Entonces:

$$P(X) = \sum_{i=1}^{n} P(X/Y_i)P(Y_i)$$

Teorema de Bayes y probabilidad

Un análisis de sangre que te has realizado ha dado positivo para una rara enfermedad, que se manifiesta sólo en 1 de cada 10.000 personas. Los resultados del test son fiables al 99%, tanto en el caso de que sean positivos como negativos. ¿Cuál es aproximadamente la probabilidad que tienes de padecer la enfermedad?

Opción A: Mayor que 0.9
Opción B: Entre 0.6 y 0.9
Opción C: Entre 0.2 y 0.6
Opción D: Menor que 0.2

Teorema de Bayes y probabilidad

Un análisis de sangre que te has realizado ha dado positivo para una rara enfermedad, que se manifiesta sólo en 1 de cada 10.000 personas. Los resultados del test son fiables al 99%, tanto en el caso de que sean positivos como negativos. ¿Cuál es aproximadamente la probabilidad que tienes de padecer la enfermedad?

Sea A = padecer la enfermedad y B= dar positivo en el test

Según el Teorema de Bayes:

Prevalencia

Tasa de verdaderos positivos

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

Tasa de falsos positivos

$$= \frac{0.0001\cdot 0.99}{0.0001\cdot 0.99 + 0.9999\cdot 0.1} = 0.0098$$

¿Cómo cambia la probabilidad de estar enfermo habiendo dado positivo en la prueba en función de la prevalencia y las tasas de verdaderos positivos y falsos positivos?

Prevalencia	0,0001
Verdaderos Positivos	0,99
Falsos positvos	0,01
Probabilidad	**0,0098**

Prevalencia	0,01
Verdaderos Positivos	0,99
Falsos positvos	0,01
Probabilidad	**0,5**

Prevalencia	0,1
Verdaderos Positivos	0,99
Falsos positvos	0,01
Probabilidad	**0,9167**

Prevalencia	0,0001
Verdaderos Positivos	1
Falsos positivos	0,0001
Probabilidad	**0,5**

Repaso de probabilidad (II)

Cálculo de distribución marginal a partir de la distribución conjunta

Dado un conjunto de variables aleatorias $X_1,\ldots,X_n$, sea $p(X_1,\ldots,X_n)$ su distribución de probabilidad conjunta. La distribución marginal de X_j se calcula mediante:

$$P\left(X_j = x_j\right) = \sum_{X_1,\ldots,X_j=x_j,\ldots X_n} P(X_1, \ldots, X_{j-1}, x_j, X_{j+1}, \ldots, X_n)$$

Método probabilístico clásico

Individuos	Edad	Obesidad	Hernia	Indigestión	Vómitos
Individuo 1	Mayor_50	no	no	no	no
Individuo 2	Mayor_50	no	no	no	no
Individuo 3	Mayor_50	no	no	no	no
Individuo 4	Mayor_50	no	no	no	no
Individuo 5	Mayor_50	no	sí	no	sí
Individuo 6	Mayor_50	sí	sí	no	sí
Individuo 7	Mayor_50	sí	sí	no	sí
Individuo 8	Mayor_50	sí	sí	no	si
Individuo 9	Menor_50	no	no	no	no
Individuo 10	Menor_50	no	no	no	no
Individuo 11	Menor_50	no	no	no	no
Individuo 12	Menor_50	no	no	no	no
Individuo 13	Menor_50	no	no	no	no
Individuo 14	Menor_50	no	no	no	no
Individuo 15	Menor_50	no	no	no	no
Individuo 16	Menor_50	no	no	no	no
Individuo 17	Menor_50	no	no	sí	sí
Individuo 18	Menor_50	sí	no	no	sí
Individuo 19	Menor_50	sí	no	sí	sí
Individuo 20	Menor_50	sí	no	no	no

Método probabilístico clásico

Se pide:

a) Calcular la distribución conjunta de las variables

b) A partir de la distribución conjunta calcular

- Las distribuciones marginales:
  P(Edad = Mayor_50) y P(Edad = Mayor_50, Hernia= sí)
- P(Hernia=si/Vómitos=si) (diagnóstico)
- P(Vómitos=si/Obesidad=si, Hernia=si) (predicción)

Método probabilístico clásico

Se pide:

a) Calcular la distribución conjunta de las variables

P(Edad= Mayor_50; Obesidad=no; Hernia=no; Indigestión=no, Vómitos=no) = 4/20
P(Edad= Mayor_50; Obesidad=no; Hernia=sí; Indigestión=no, Vómitos=sí) = 1/20
P(Edad= Mayor_50; Obesidad=sí; Hernia=sí; Indigestión=no, Vómitos=sí) = 3/20
P(Edad= Menor_50; Obesidad=no; Hernia=no; Indigestión=no, Vómitos=no) = 8/20
P(Edad= Menor_50; Obesidad=no; Hernia=no; Indigestión=si, Vómitos=si) = 1/20
P(Edad= Menor_50; Obesidad=sí; Hernia=no; Indigestión=no, Vómitos=sí) = 1/20
P(Edad= Menor_50; Obesidad=sí; Hernia=no; Indigestión=si, Vómitos=sí) = 1/20
P(Edad= Menor_50; Obesidad=sí; Hernia=no; Indigestión=no, Vómitos=no) = 1/20

b) A partir de la distribución conjunta calcular:

- Las distribuciones marginales:

  P(Edad = Mayor_50) y P(Edad = Mayor_50, Hernia= sí) **8/20** **4/20**

- P(Hernia=si/Vómitos=si) (diagnóstico) **4/7**
- P(Vómitos=si/Obesidad=si, Hernia=si) (predicción) **1**

¿Te parecen razonables estas probabilidades?

Repaso de probabilidad (III)

Dos variables X e Y son **condicionalmente independientes** dado una tercera variable Z si se tiene que

$$P(X/Y,Z) = P(X/Z)$$

***Caracterización de la independencia condicional*:**

X e Y son independientes dado Z sí y sólo sí $P(X,Y/Z) = P(X/Z) \cdot P(Y/Z)$

También se dice que **Z separa condicionalmente a X e Y**

Generalización de la ley de probabilidad total

$$P(Y_1/Y_2) = \sum_X P(Y_1/X,Y_2)P(X/Y_2)$$

Presentación intuitiva de red bayesiana

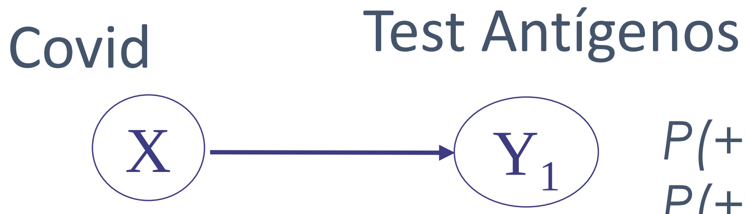
En una **red bayesiana**, cada nodo corresponde a una variable, que a su vez representa una entidad del mundo real

Notación:

- *X,* variable o nodo
- *x*, valor ($+x$=sí, $\neg x$=no)

Los arcos que unen los nodos representan relaciones de ***influencia causal***

Presentación intuitiva: ejemplo 1



$P(+y_1/+x) = 0.9$ $\quad$ $P(\neg y_1/+x) = 0.1$

$P(+y_1/\neg x) = 0.05$ $\quad$ $P(\neg y_1/\neg x) = 0.95$

$P(+x) = 0.099$ $\quad$ $P(\neg x) = 0.901$

Información cualitativa

- Nodos (variables)
- Enlaces (relaciones de influencia causal)

Información cuantitativa:

- La probabilidad a priori de los nodos que no tienen padres.
- La probabilidad condicionada de los nodos con padres.

Presentación intuitiva ejemplo 1: significado de los parámetros

Covid

Test Antígenos

$P(+x) = 0.099$

$P(+y_1/+x) = 0.9$
$P(+y_1/\neg x) = 0.05$

Enfermedad

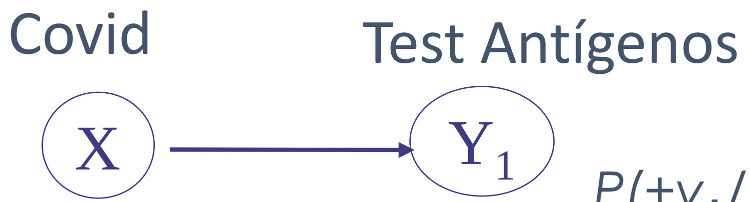
- $P(+x) = \mathbf{0.099}$ — **prevalencia**

Test

- $P(+y_1/+x) = \mathbf{0.9}$ — **sensibilidad**
- $P(\neg y_1/+x) = 0.1$ — falsos negativos
- $P(+y_1/\neg x) = \mathbf{0.05}$ — falsos positivos
- $P(\neg y_1/\neg x) = 0.95$ — **especificidad**

En medicina siempre se buscan las pruebas con **mayor grado de sensibilidad y especificidad**

Presentación intuitiva ejemplo 1: inferencias



$P(+x) = 0.099$

$P(+y_1/+x) = 0.9$
$P(+y_1/\neg x) = 0.05$

$P(+x) = \mathbf{0.099}$
$P(+y_1/+x) = \mathbf{0.9}$
$P(\neg y_1/+x) = 0.1$
$P(+y_1/\neg x) = \mathbf{0.05}$
$P(\neg y_1/\neg x) = 0.95$

Inferencias a partir de los datos

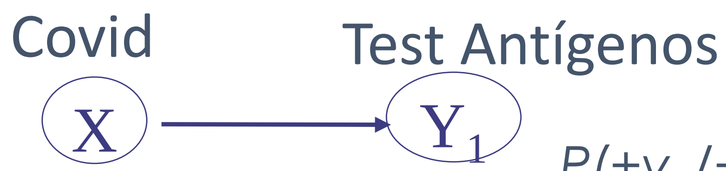
¿Cuál es la probabilidad de que una persona de positivo en un test de antígenos?

Probabilidad a priori de Y_1

$$P(+y_1) = P(+y_1/+x)\,P(+x) + P(+y_1/\neg x)\,P(\neg x) = 0.9 \cdot 0.099 + 0.05 \cdot 0.901 = \mathbf{0.13415}$$

$$P(\neg y_1) = P(\neg y_1/+x)\,P(+x) + P(\neg y_1/\neg x)\,P(\neg x) = 0.1 \cdot 0.099 + 0.95 \cdot 0.901 = 0.86585$$

Presentación intuitiva ejemplo 1: inferencias



$P(+x) = 0.099$

$P(+y_1/+x) = 0.9$
$P(+y_1/\neg x) = 0.05$

$P(+x)$ = **0.099**
$P(+y_1/+x)$ = **0.9**
$P(\neg y_1/+x) = 0.1$
$P(+y_1/\neg x)$ = **0.05**
$P(\neg y_1/\neg x) = 0.95$

Inferencias a partir de los datos

Si una persona ha dado positivo en el test de antígenos, ¿cuál es la probabilidad de que tenga Covid?

Probabilidades a posteriori dada una evidencia observada *e*: $P^*(x) = P(x/\mathbf{e})$.

Test de antígenos ha dado positivo: $+y_1$

$$P^*(+x) = P(+x/+y_1) = \frac{P(+y_1/+x)P(+x)}{P(+y_1)} = \mathbf{0.6642}$$

$$P^*(\neg x) = P(\neg x/+y_1) = \frac{P(+y_1/\neg x)P(\neg x)}{P(+y_1)} = 1 - P(+x/+y_1) = 0.3358$$

Presentación intuitiva ejemplo 1: inferencias

La expresión general del teorema de Bayes que hemos utilizado es:

$$P^*(x) = P(x/y) = \frac{P(y/x)P(x)}{P(y)}$$

P(+x) = **0.099**
$P(+y_1/+x)$ = **0.9**
$P(\neg y_1/+x) = 0.1$
$P(+y_1/\neg x)$ = **0.05**
$P(\neg y_1/\neg x) = 0.95$

que podemos reescribir como

$$P^*(x) = \alpha P(x)\lambda(x)$$

donde $\alpha = 1/P(y)$ y $\lambda(x) = P(y/x)$

Utilizando esta nueva expresión:

$$P^*(+x) = \alpha 0.099\ 0.9 = 0.0891\ \alpha$$

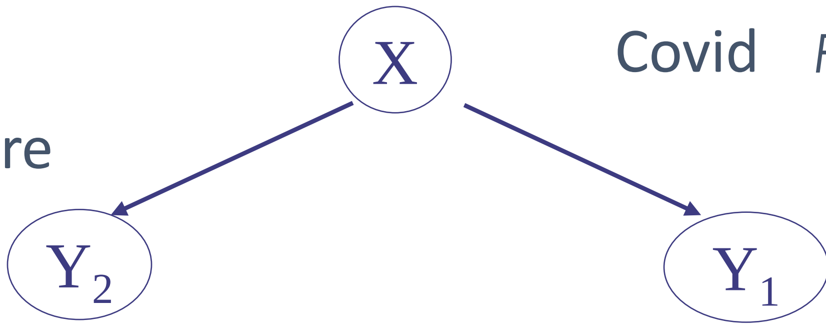
$$P^*(\neg x) = \alpha 0.901\ 0.05 = 0.04505\ \alpha$$

Y normalizando obtenemos el mismo resultado que antes

Presentación intuitiva: ejemplo 2

Covid $P(+x) = 0.099$

Fiebre



Test Antígenos

$P(+y_1/+x) = 0.9$

$P(+y_1/\neg x) = 0.05$

$P(+y_2/+x) = 0.6$

$P(+y_2/\neg x) = 0.1$

Inferencias

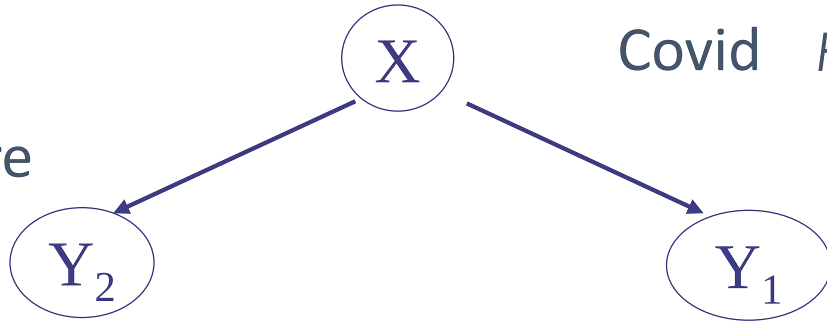
Si una persona tiene fiebre, ¿cuál es la probabilidad de que tenga Covid?

Opción A: mayor que antes
Opción B: menor que antes
Opción C: la misma de antes

Presentación intuitiva ejemplo 2: inferencias

Covid $P(+x) = 0.099$

Fiebre



Test Antígenos

$P(+y_1/+x) = 0.9$
$P(+y_1/\neg x) = 0.05$

$P(+y_2/+x) = 0.6$
$P(+y_2/\neg x) = 0.1$

Si una persona tiene fiebre, ¿cuál es la probabilidad de que tenga Covid?

Probabilidades a posteriori dado que hay fiebre: $+y_2$

$$P^{**}(x) = P(+x/+y_2) = \alpha P(x)\lambda(x)\ ,\lambda_{y_2}(x) = P(y_2/x)$$

$$P^{**}(+x) = P(+x/+y_2) = \alpha\, 0.6\ 0.099 = 0.0594\ \alpha = \mathbf{0.3973}$$

$$P^{**}(\neg x) = P(\neg x/+y_2) = \alpha\, 0.1\ 0.901 = 0.0901\ \alpha = 0.6027$$

$$\alpha = \frac{1}{0.0594+0.0901} = \frac{1}{0.1495}$$

Es menor que si da positivo en Test

$P^{*}(+x) = P(+x/+y_1) = 0.6642$

Presentación intuitiva ejemplo 2: inferencias

Si una persona tiene fiebre y ha dado positivo en el test de antígenos, ¿cuál es la probabilidad de que tenga Covid?

Probabilidades a posteriori dado que tiene fiebre y ha dado positivo en el test:{*+y$_1$, +y$_2$*}

?

$$P^*(x) = P(x/y_1, y_2) = \frac{P(y_1, y_2/x)P(x)}{P(y_1, y_2)}$$

Hipótesis independencia condicional:

$P(y_1,y_2/x) = P(y_1/x)\, P(y_2/x)$

$$\lambda(x) = \lambda_{y_1}(x) \cdot \lambda_{y_2}(x)$$

Y entonces: $P^*(x) = \alpha P(x)\lambda(x)$

Presentación intuitiva ejemplo 2

$\lambda(x) = P(y/x)$

$P(+x) = 0.099$
$P(+y_1/+x) = 0.9$
$P(+y_1/\neg x) = 0.05$
$P(+y_2/+x) = 0.6$
$P(+y_2/\neg x) = 0.1$

Si seguimos con los cálculos:

$$P^{***}(+x) = \alpha P(+x)\lambda_{+y_1}(+x)\ \lambda_{+y_2}(+x) = \alpha\ 0.099 \cdot 0.9 \cdot 0.6 = \mathbf{0.9223}$$
$$P^{***}(\neg x) = \alpha P(\neg x)\lambda_{+y_1}(\neg x)\ \lambda_{+y_2}(\neg x) = \alpha\ 0.901 \cdot 0.05 \cdot 0.1 = 0.0777$$

Ejercicio propuesto:

Utilizando esta formulación, hallar la probabilidad de que una persona tenga Covid si tiene fiebre y ha dado negativo en el test de antígenos

Solución: 0.0649

Presentación intuitiva ejemplo 2: inferencias

Si una persona tiene fiebre, ¿cuál es la probabilidad de que dé positivo en el test de antígenos?

Probabilidad de dar positivo en el test dado que tiene fiebre $+y_1/+y_2$

$P^*(y_1) = P(y_1/y_2) = \sum P(y_1/x, y_2)\, P(x/y_2) = \sum P(y_1/x)\, P(x,y_2)P(y_2)^{-1}$

Independencia condicional

$P(+x) = 0.099$
$P(+y_1/+x) = 0.9$
$P(+y_1/\neg x) = 0.05$
$P(+y_2/+x) = 0.6$
$P(+y_2/\neg x) = 0.1$

$\lambda(x) = P(y/x)$

Sea $\pi_{Y_1}(x) = P(x, y_2) = P(x)P(y_2/x)$, $\alpha = [P(y_2)]^{-1}$

Entonces $P^*(y_1) = \alpha \sum_X P(y_1/X)\pi_{Y_1}(X)$

Y finalmente:

$$P^*(+y_1) = \alpha\, [\pi_{Y_1}(+x)\, P(+y_1/+x) + \pi_{Y_1}(\neg x)\, P(+y_1/\neg x)] = \mathbf{0.3877}$$

$$P^*(\neg y_1) = \alpha\, [\pi_{Y_1}(+x)\, P(\neg y_1/+x) + \pi_{Y_1}(\neg x)\, P(\neg y_1/\neg x)] = 0.6123$$

Ejercicio propuesto:

Utilizando esta formulación, hallar la probabilidad de que una persona que ha dado positivo en el test de antígenos tenga fiebre

Solución: **0.4321**

Presentación intuitiva: CONCLUSIONES

Las redes bayesianas permiten hacer inferencias:

- **Abductivas**: ¿cúal es el diagnóstico que mejor explica los hallazgos?
- **Predictivas**: ¿cuál es la probabilidad de obtener cierto resultado en el futuro?