Ejercicios 5: Razonamiento bayesiano

Departamento de Informática / Department of Computer Science
Universidad Carlos III de Madrid

Inteligencia Artificial
Grado en Ingeniería Informática
2019/20

Tabla de contenidos

Ejercicio 1: Hijos

- Una pareja tiene cuatro hijos. La probabilidad de tener un hijo de sexo masculino es 0.5
  1. ¿Cuál es la probabilidad de que los cuatro hijos tengan el mismo sexo?
  2. ¿Cuál es la probabilidad de que sólo dos hijos sean de sexo masculino?
  3. ¿Cuál es la probabilidad de que al menos dos hijos sean de sexo masculino?

- Todas las posibilidades son:

Hijos, Solución

- Todas las posibilidades son:

MMMM	MMMF	MMFM	MMFF
MFMM	MFMF	MFFM	MFFF
FMMM	FMMF	FMFM	FMFF
FFMM	FFMF	FFFM	FFFF

1. Probabilidad de que los cuatro hijos tengan el mismo sexo:

Hijos, Solución

- Todas las posibilidades son:

MMMM	MMMF	MMFM	MMFF
MFMM	MFMF	MFFM	MFFF
FMMM	FMMF	FMFM	FMFF
FFMM	FFMF	FFFM	FFFF

1. Probabilidad de que los cuatro hijos tengan el mismo sexo: 2/16
2. Probabilidad de que sólo dos hijos sean de sexo masculino:

Hijos, Solución

- Todas las posibilidades son:

MMMM	MMMF	MMFM	MMFF
MFMM	MFMF	MFFM	MFFF
FMMM	FMMF	FMFM	FMFF
FFMM	FFMF	FFFM	FFFF

1. Probabilidad de que los cuatro hijos tengan el mismo sexo: 2/16
2. Probabilidad de que sólo dos hijos sean de sexo masculino: 6/16
3. Probabilidad de que al menos dos hijos sean de sexo masculino:

Hijos, Solución

- Todas las posibilidades son:

MMMM	MMMF	MMFM	MMFF
MFMM	MFMF	MFFM	MFFF
FMMM	FMMF	FMFM	FMFF
FFMM	FFMF	FFFM	FFFF

1. Probabilidad de que los cuatro hijos tengan el mismo sexo: 2/16
2. Probabilidad de que sólo dos hijos sean de sexo masculino: 6/16
3. Probabilidad de que al menos dos hijos sean de sexo masculino: 11/16

Ejercicio 2: Dentista

Un destista ha registrado la siguiente información para un total de 1000 pacientes. Estos datos relacionan el dolor de muelas con la caries y el hecho de que la sonda del dentista se enganche en la muela

	dolor		¬ dolor	
	enganche	¬ enganche	enganche	¬ enganche
caries	108	12	72	8
¬ caries	16	64	144	576

- ¿Cuál es la distribución de probabilidad conjunta?
- ¿Cuál es la probabilidad de que el paciente tenga caries, $P(\text{caries})$?
- ¿Cuál es la probabilidad de que haya enganche y el paciente no tenga dolor $P(\neg\text{dolor}, \text{enganche})$?
- Sabemos que un paciente tiene dolor ¿Cuál es la probabilidad de que tenga caries?

Dentista, distribución de probabilidad conjunta

Un dentista ha registrado la siguiente información para un total de 1000 pacientes. Estos datos relacionan el dolor de muelas con la caries y el hecho de que la sonda del dentista se enganche en la muela

	dolor		¬ dolor	
	enganche	¬ enganche	enganche	¬ enganche
caries	0.108	0.012	0.072	0.008
¬ caries	0.016	0.064	0.144	0.576

Dentista, distribución de probabilidad conjunta

Un dentista ha registrado la siguiente información para un total de 1000 pacientes. Estos datos relacionan el dolor de muelas con la caries y el hecho de que la sonda del dentista se enganche en la muela

	dolor		$\neg$ dolor	
	enganche	$\neg$ enganche	enganche	$\neg$ enganche
caries	0.108	0.012	0.072	0.008
$\neg$ caries	0.016	0.064	0.144	0.576

- ¿Cuál es la distribución de probabilidad conjunta?
  $P(\mathrm{Caries}, \mathrm{Dolor}, \mathrm{Enganche})$
- ¿Cuál es la probabilidad de que el paciente tenga caries?

$$P(\mathrm{caries}) = \sum_{caries} P(\mathrm{Caries}, \mathrm{Dolor}, \mathrm{Enganche}) = 0.108+0.012+0.072+0.008 = 0.2$$

- ¿Cuál es la probabilidad de que se produzca enganche y no tenga dolor?

$$P(\neg\mathrm{dolor}, \mathrm{enganche}) = 0.072 + 0.144 = 0.216$$

Dentista, distribución de probabilidad conjunta

Un dentista ha registrado la siguiente información para un total de 1000 pacientes. Estos datos relacionan el dolor de muelas con la caries y el hecho de que la sonda del dentista se enganche en la muela

	dolor		$\neg$ dolor	
	enganche	$\neg$ enganche	enganche	$\neg$ enganche
caries	0.108	0.012	0.072	0.008
$\neg$ caries	0.016	0.064	0.144	0.576

- Sabemos que un paciente tiene dolor ¿Cuál es la probabilidad de que tenga caries?

$$P(\text{caries}/\text{dolor}) = \frac{P(\text{caries}, \text{dolor})}{P(\text{dolor})} = \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6$$

Dentista, distribución de probabilidad conjunta

Un dentista ha registrado la siguiente información para un total de 1000 pacientes. Estos datos relacionan el dolor de muelas con la caries y el hecho de que la sonda del dentista se enganche en la muela

	dolor		¬ dolor	
	enganche	¬ enganche	enganche	¬ enganche
caries	0.108	0.012	0.072	0.008
¬ caries	0.016	0.064	0.144	0.576

- Sabemos que un paciente tiene dolor ¿Cuál es la probabilidad de que tenga caries?

$$P(\text{caries}/\text{dolor}) = \frac{P(\text{caries}, \text{dolor})}{P(\text{dolor})} = \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6$$

- Normalización: $P(\text{dolor})$ no se necesita

Dentista, distribución de probabilidad conjunta

Un dentista ha registrado la siguiente información para un total de 1000 pacientes. Estos datos relacionan el dolor de muelas con la caries y el hecho de que la sonda del dentista se enganche en la muela

	dolor		¬ dolor	
	enganche	¬ enganche	enganche	¬ enganche
caries	0.108	0.012	0.072	0.008
¬ caries	0.016	0.064	0.144	0.576

- Sabemos que un paciente tiene dolor ¿Cuál es la probabilidad de que tenga caries?

$$P(\mathrm{caries}/\mathrm{dolor}) = \frac{P(\mathrm{caries}, \mathrm{dolor})}{P(\mathrm{dolor})} = \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6$$

- Normalización: $P(\mathrm{dolor})$ no se necesita

$$\begin{aligned} P(\mathrm{Caries}/\mathrm{Dolor}) &= \alpha P(\mathrm{Caries}, \mathrm{Dolor}) \\ &= \alpha[P(\mathrm{Caries}, \mathrm{dolor}, \mathrm{enganche}) + P(\mathrm{Caries}, \mathrm{dolor}, \neg\mathrm{enganche})] \\ &= (0.6, 0.4) \end{aligned}$$

Ejercicio 3: Diagnóstico

Después de la revisión anual, el doctor tiene noticias malas y buenas. Las malas son que has dado positivo en una enfermendad seria, siendo el test preciso en un 99 % (es decir, la probabilidad de dar positivo cuando se tiene la enfermedad o de dar negativo cuando no se tiene es 0.99). Las buena noticia es que es una enfermedad rara, que solo se produce en 1 de cada $10,000$ personas de tu edad. ¿Por qué es una buena noticia que la enfermedad es rara? ¿Cuál es la probabilidad de que realmente tengas la enfermedad?

- Paso 1: Notación de los estados (variables aleatorias):

Diagnóstico, variables

- Paso 1: Notación de los estados (variables aleatorias):
  - Test: variable Booleana
    - test: test positivo
      ¬test: test negativo
  - Enfermo: variable Booleana
    - enfermo: tienes la enfermedad
    - ¬enfermo: no tienes la enfermedad
- Paso 2: probabilidades conocidas

Diagnóstico, variables

- Paso 1: Notación de los estados (variables aleatorias):
  - Test: variable Booleana
    - test: test positivo
      ¬test: test negativo
  - Enfermo: variable Booleana
    - enfermo: tienes la enfermedad
    - ¬enfermo: no tienes la enfermedad
- Paso 2: probabilidades conocidas
  - Test=test
  - P(test|enfermo)=0.99
  - P(¬test|¬enfermo)=0.99
  - P(enfermo)=0.0001

- Paso 1: Notación de los estados (variables aleatorias):
  - Test: variable Booleana
    - test: test positivo
      ¬test: test negativo
  - Enfermo: variable Booleana
    - enfermo: tienes la enfermedad
    - ¬enfermo: no tienes la enfermedad
- Paso 2: probabilidades conocidas
  - Test=test
  - P(test|enfermo)=0.99
  - P(¬test|¬enfermo)=0.99
  - P(enfermo)=0.0001
- También conocemos:
  - P(¬enfermo)=0.9999, P(¬test|enfermo)=0.01, P(test|¬enfermo)=0.01

Diagnóstico, cálculo

- Lo que queremos saber es la probabilidad de tener realmente la enfermedad dado que el test es positivo

Diagnóstico, cálculo

- Lo que queremos saber es la probabilidad de tener realmente la enfermedad dado que el test es positivo
- Por lo tanto queremos conocer P(enfermo|test)

Diagnóstico, cálculo

- Lo que queremos saber es la probabilidad de tener realmente la enfermedad dado que el test es positivo
- Por lo tanto queremos conocer P(enfermo|test)

$$\text{P(enfermo|test)} = \frac{P(test|enfermo)P(enfermo)}{P(test)}$$

Diagnóstico, cálculo

- Lo que queremos saber es la probabilidad de tener realmente la enfermedad dado que el test es positivo
- Por lo tanto queremos conocer P(enfermo|test)

  $$\mathrm{P(enfermo|test)} = \frac{P(test|enfermo)P(enfermo)}{P(test)}$$

- Necesitamos P(test)?

Diagnóstico, cálculo

- Lo que queremos saber es la probabilidad de tener realmente la enfermedad dado que el test es positivo
- Por lo tanto queremos conocer P(enfermo|test)

  $$\mathrm{P(enfermo|test)} = \frac{P(test|enfermo)P(enfermo)}{P(test)}$$

- Necesitamos P(test)?

  $$\begin{aligned} \mathrm{P(test)} &= \mathrm{P(test|enfermo)P(enfermo)} + \mathrm{P(test|\neg enfermo)P(\neg enfermo)} \\ &= (0.99)(0.0001) + (0.01)(0.9999) \\ &= 0.010098 \end{aligned}$$

Diagnóstico, cálculo

- Lo que queremos saber es la probabilidad de tener realmente la enfermedad dado que el test es positivo
- Por lo tanto queremos conocer P(enfermo|test)

  $$\text{P(enfermo|test)} = \frac{P(test|enfermo)P(enfermo)}{P(test)}$$

- Necesitamos P(test)?

  $$\begin{aligned}\text{P(test)} &= \text{P(test|enfermo)P(enfermo)} + \text{P(test|}\neg\text{enfermo)P(}\neg\text{enfermo)} \\ &= (0.99)(0.0001) + (0.01)(0.9999) \\ &= 0.010098\end{aligned}$$

- Ahora calculamos P(enfermo|test)

Diagnóstico, cálculo

- Lo que queremos saber es la probabilidad de tener realmente la enfermedad dado que el test es positivo
- Por lo tanto queremos conocer P(enfermo|test)

$$\text{P(enfermo|test)} = \frac{P(test|enfermo)P(enfermo)}{P(test)}$$

- Necesitamos P(test)?

$$\begin{aligned}\text{P(test)} &= \text{P(test|enfermo)P(enfermo)} + \text{P(test}|\neg\text{enfermo)P(}\neg\text{enfermo)} \\ &= (0.99)(0.0001) + (0.01)(0.9999) \\ &= 0.010098\end{aligned}$$

- Ahora calculamos P(enfermo|test)

$$\begin{aligned}\text{P(enfermo|test)} &= \frac{P(test|enfermo)P(enfermo)}{P(test)} \\ &= \frac{(0.99)(0.0001)}{0.010098} \\ &= 0.009804\end{aligned}$$

P(test) se puede ver como un factor de normalización

Ejercicio 4: Estudio de mercado

Supongamos (por simplificar en dos grupos) que en una ciudad consideremos que el 51 % de los adultos son hombres y el resto mujeres. Un adulto se selecciona aleatoriamente para realizar un estudio relacionado con el uso de hábitos de consumo.

- Determina cuál es la probabilidad a priori de que la persona seleccionada sea hombre
- Después se da a conocer que la persona seleccionada es consumidor de un producto C. Se sabe que el 9.5 % de los hombres consume, mientras que lo hacen in 1.7 % de las mujeres. Utiliza esta información para determinar la probabilidad de que se haya seleccionado un hombre.

Estudio de mercado, variables

Utilizaremos la siguiente notación:

- *M* proposicional:
  - m = masculino
  - ¬m = femenino (o no masculino)
- *C* proposicional:
  - c = consume C
  - ¬c = no consume C

Antes de conocer la información dada en el apartado b, sólo sabemos que el 51 % de los adultos son de sexo masculino. Por lo tanto, la probabilidad de seleccionar aleatoriamente un adulto masculino es P(m) = 0.51.

- Identificamos qué información nos dan:
  - $P(m) = 0.51$ porque el 51 % de los adultos son masculinos
  - $P(\neg m) = 0.49$ porque el 49 % de los adultos son femeninos
  - $P(c|m) = 0.095$ porque el 9.5 % de los de sexo masculino consumen C (Esto es, la probabilidad de seleccionar alguién que consume el producto, dado que es de sexo masculino es 0.095)
  - $P(c|\neg m) = 0.017$. porque un 1.7 % de los de sexo femenino consume C (Esto es, la probabilidad de seleccionar alguien que consume C, dado que es de sexo femenino es 0.017)

- Queremos calcular $P(m|c)$

- Queremos calcular P(m|c)

  $P(m|c) = \frac{P(c|m)P(m)}{P(c)} =$

- Queremos calcular P(m|c)

$$P(m|c) = \frac{P(c|m)P(m)}{P(c)} = \frac{0.095x0.51}{P(c)}$$

$$= \frac{0.095x0.51}{P(c|m)P(m)+P(c|\neg m)P(\neg m)} = \frac{0.095x0.51}{0.095x0.51+0.017x0.49} = \frac{0.04845}{0.05678}$$

$$= 0.85329341$$

Antes de saber que la persona seleccionada consumía el producto, hay una probabilidad de 0.51 de que sea masculina. Sin embargo, después de conocer esa información, la probabilidad se revisa a 0.853. Esto tiene sentido, porque la probabilidad de ser hombre se incrementa en mucho con la información adicional de que la persona consume C (porque lo consumen más hombres que mujeres)

Ejercicio 5: Detección de fallos

- La empresa A fabrica el 80 % de un producto. La empresa B fabrica el 15 %, y la empresa C el 5 % restante. Los productos que fabrica la empresa A tienen un ratio de defectos del 4 %. Para la empresa B el ratio es del 6 % y para la C del 9 %, lo que explica por qué la empresa C tiene menos ventas.
- Si seleccionamos un producto de entre todos, determinar la probabilidad de que lo haya fabricado la empresa A
- Si un producto seleccionado aleatoriamente tiene defectos, determinar la probabilidad de que lo haya fabricado la empresa A

Detección de fallos, planteamiento

1. P(a)=?

1. P(a)=?

   Si un producto se selecciona aleatoriamente la probabilidad de que haya sido fabricado por la empresa A es 0.8

   P(a)=0.8

Detección de fallos, planteamiento

1. P(a)=?

   Si un producto se selecciona aleatoriamente la probabilidad de que haya sido fabricado por la empresa A es 0.8

   P(a)=0.8
2. Ahora queremos determinar:

Detección de fallos, planteamiento

1. P(a)=?

   Si un producto se selecciona aleatoriamente la probabilidad de que haya sido fabricado por la empresa A es 0.8

   P(a)=0.8
2. Ahora queremos determinar:$P(a|d)$

1. P(a)=?

   Si un producto se selecciona aleatoriamente la probabilidad de que haya sido fabricado por la empresa A es 0.8

   P(a)=0.8

2. Ahora queremos determinar:$P(a|d)$

   - Probabilidades conocidas:
     - P(a)= 0.80
     - P(b)= 0.15
     - P(c)= 0.05
     - $P(d|a)$=0.04
     - $P(d|b)$= 0.06
     - $P(d|c)$= 0.09

1. P(a)=?

   Si un producto se selecciona aleatoriamente la probabilidad de que haya sido fabricado por la empresa A es 0.8

   P(a)=0.8

2. Ahora queremos determinar:P(a|d)

   - Probabilidades conocidas:
     - P(a)= 0.80
     - P(b)= 0.15
     - P(c)= 0.05
     - P(d|a)=0.04
     - P(d|b)= 0.06
     - P(d|c)= 0.09

   $P(a|d) = \frac{P(d|a)P(a)}{P(d)}$

1. P(a)=?

   Si un producto se selecciona aleatoriamente la probabilidad de que haya sido fabricado por la empresa A es 0.8

   P(a)=0.8

2. Ahora queremos determinar:P(a|d)
   - Probabilidades conocidas:
     - P(a)= 0.80
     - P(b)= 0.15
     - P(c)= 0.05
     - P(d|a)=0.04
     - P(d|b)= 0.06
     - P(d|c)= 0.09

$$
\begin{aligned}
P(a|d) &= \frac{P(d|a)P(a)}{P(d)} \\
&= \frac{P(d|a)P(a)}{P(d|a)P(a)+P(d|b)(b)+P(d|c)P(c)} \\
&= \frac{(0.04)(0.80)}{(0.04)(0.80)+(0.06)(0.15)+(0.09)(0.05)} \\
&= 0.703
\end{aligned}
$$

Detección de fallos, cálculo

- Calculando y luego normalizando

- Calculando y luego normalizando

$$
\begin{aligned}
P(a|d) &= \alpha P(d|a)P(a) \\
&= \alpha(0.04)(0.80) \\
&= \alpha 0.032
\end{aligned}
$$

Detección de fallos, cálculo

- Calculando y luego normalizando

$$
\begin{aligned}
P(a|d) &= \alpha P(d|a)P(a) \\
&= \alpha(0.04)(0.80) \\
&= \alpha 0.032
\end{aligned}
$$

$$
\begin{aligned}
P(b|d) &= \alpha P(d|b)P(b) \\
&= \alpha(0.06)(0.15) \\
&= \alpha 0.009
\end{aligned}
$$

Detección de fallos, cálculo

- Calculando y luego normalizando

$$
\begin{aligned}
P(a|d) &= \alpha P(d|a)P(a) \\
&= \alpha(0.04)(0.80) \\
&= \alpha 0.032
\end{aligned}
$$

$$
\begin{aligned}
P(b|d) &= \alpha P(d|b)P(b) \\
&= \alpha(0.06)(0.15) \\
&= \alpha 0.009
\end{aligned}
$$

$$
\begin{aligned}
P(c|d) &= \alpha P(d|c)P(c) \\
&= \alpha(0.09)(0.05) \\
&= \alpha 0.0045
\end{aligned}
$$

- Calculando y luego normalizando

$$
\begin{aligned}
P(a|d) &= \alpha P(d|a)P(a) \\
&= \alpha(0.04)(0.80) \\
&= \alpha 0.032
\end{aligned}
$$

$$
\begin{aligned}
P(b|d) &= \alpha P(d|b)P(b) \\
&= \alpha(0.06)(0.15) \\
&= \alpha 0.009
\end{aligned}
$$

$$
\begin{aligned}
P(c|d) &= \alpha P(d|c)P(c) \\
&= \alpha(0.09)(0.05) \\
&= \alpha 0.0045
\end{aligned}
$$

$$
\begin{aligned}
P(E|d) &= \alpha\{0.032, 0.009, 0.0045\} \\
&= \{0.032/0.0455, 0.009/0.0455, 0.0045/0.0455\} \\
&= \{0.703, 0.2, 0.1\}
\end{aligned}
$$