Inteligencia Artificial

Incertidumbre

Jorge Luis Guevara Diaz

www.jorge.sistemasyservidores.com

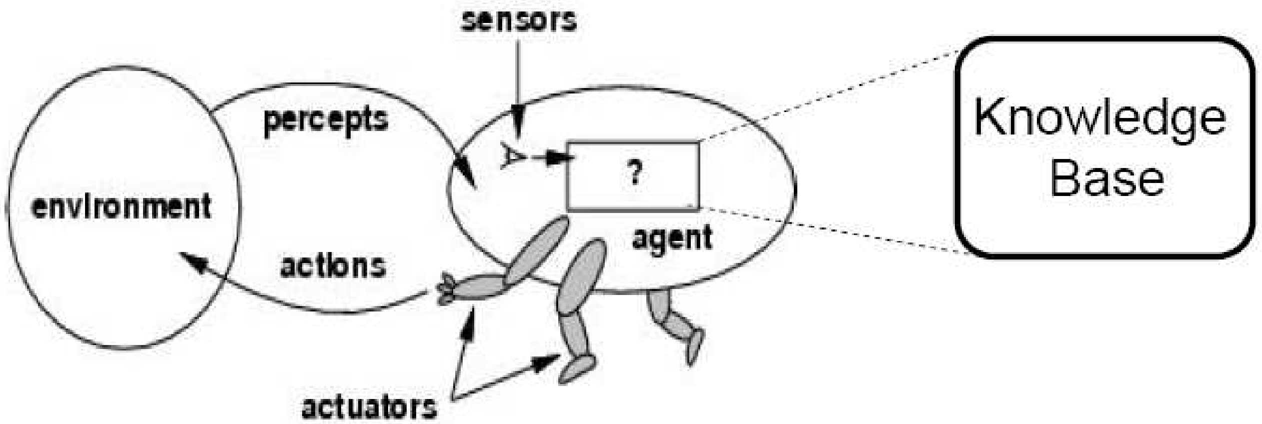
Que veremos hoy?

1. Actuando bajo Incertidumbre
2. Probabilidad, Notación básica
3. Axiomas de la Probabilidad
4. Inferencia Usando Distribucion conjunta completa
5. Independencia
6. Regla de Bayes
7. Ejemplo

Actuando bajo Incertidumbre

Actuando bajo Incertidumbre

- Agentes que son representados y actúan bajo incertidumbre
  - Comportamiento inteligente requiere conocimiento del mundo
  - Algunas veces uno está incierto del estado del mundo



Actuando bajo Incertidumbre

- La teoria de la probabilidad provee un marco para representar un agente que razone bajo incertidumbre
  - Representa las creencias acerca del mundo como sentencias (muy parecido a la lógica proposicional)
  - Asocia las probabilidades con las sentencias
  - Razona manipulando las sentencias, de acuerdo a reglas sólidas de inferencia probabilistica
  - El resultado de las inferencias son probabilidades asociadas con conclusiones, que son justificadas por las creencias y los datos (observaciones)

Permite a los agentes sustituir pensamiento por acciones en el mundo

Actuando bajo Incertidumbre

- Ejemplo:
- Creencias
  - Si los alumnos de IA estudian tiene un 60% de oportunidad de pasar el curso, y 40% de no pasarlo
  - Si los alumnos de IA no estudian tienen un 20% de oportunidad de aprobar el curso y un 80% de no aprobarlo

  Observacion: Los alumnos de IA no estudiaron

  Inferencia: ¿Cuál es la oportunidad de que puedan pasar el curso, y cual oportunidad de no pasarlo?
- Teoria de la probabilidad generaliza la lógica proposicional

  La teoría de la probabilidad asocia valores de probabilidad en el intervalo [0,1]

Actuando bajo Incertidumbre

La teoría de la probabilidad como representacion de conocimiento

- Compromiso Ontológico (Que queremos decir del mundo?)
  - Proposiciones que representan las creencias del agente con respecto al mundo
- Compromiso Epistemológico (Que podemos creer?)
  - Cual es la probabilidad que una proposicion sea verdadera (dada las creencias y las observaciones?)
- Sintaxis
  - Parecido a la lógica proposicional
- Semántica
  - Interpretación relativa a las frecuencias
  - Interpretación Bayesiana
- Teoria de Prueba
  - Basada en las leyes de la probabilidad

Actuando bajo Incertidumbre

Fuentes de incertidumbre

Un modelo de incertidumbre basado en aserciones probabilísticas puede ser hecho para

- Un mundo determinista
  - Pereza : Falla en enumeraciones
  - Ignorancia : falta de hechos relevantes
- Un mundo estocastico
  - Incertidumbre inherente

Probabilidad, Notación básica

Probabilidad, Notación básica

- La teoria de probabilidad se usa para medir el grado de confianza en la salida de algunas acciones (observaciones), las cuales no están definidas

- Espacio Muestral
  - S = conjunto de todas las posibles salidas ejemplo : 6 posibles tiros de un dado
  - w $\in$ S= punto de muestra/posible mundo/evento atómico

Probabilidad, Notación básica

- Modelo de Probabilidad o Espacio de Probabilidad

  Es un espacio muestral con una asignación de $P(w)$ para cada $w \in S$

  se puede interpretar como $P(w) = \frac{N_w}{N_s}$

  donde $\sum_w P(w) = 1$

- Ejemplo P(1)=P(2)=P(3)=P(4)=P(5)=P(6)=1/6

Probabilidad, Notación básica

- Evento A = Es cualquier subconjunto de S

$$P(A) = \sum_{w \in A} P(w)$$

ejemplo:

P(tiro < 4) = P(1)+P(2)+P(3)=1/6+1/6+1/6=1/2

puede tambíen ser vista como $P(A) = \frac{N_A}{N_s}$

Probabilidad, Notación básica

- Variable Aleatoria

  Es una función de los puntos de muestra hacia cualquier rango por ejemplo reales, booleanos, etc

  ejemplo Impar(1) = true
- P induce una Distribución de Probabilidad para cualquier variable aleatoria

$$P(X = x_i) = \sum_{w:X(w)=x_i} P(w)$$

ejemplo: P(Impar=verdadero)=P(1)+P(3)+P(5)=1/2

Probabilidad, Notación básica

Tipos Variable Aleatorias

1. Variables aleatorias booleanas

   ejemplo X tiene el dominio {true, false}

1. Variables aleatorias discretas

   ejemplo X toman valores del dominio de valores contables {soleado, nublado, lluvioso, nevado} ó {1,2,3,4,5,6}, etc

1. Variables aleatorias continuas toman valores del dominio de los números reales ejemplo X=4.02 ó X<= 4.02

Probabilidad, Notación básica

- Evento atómico : completa especificación del estado del mundo acerca del cual el agente esta incierto
  - Ejemplo si tenemos variables booleanas A y B tienen 4 posibles eventos atómicos:

    $A=true \wedge B=true \quad = \quad a \wedge b$

    $A=true \wedge B=false \quad = \quad a \wedge -b$

    $A=false \wedge B=true \quad = \quad -a \wedge b$

    $A=false \wedge B=false \quad = \quad -a \wedge -b$

Probabilidad, Notación básica

Los eventos atómicos tienen las siguientes propiedades

1. Mutuaente exclusivos : A lo mucho uno puede ser el caso por ejemplo a Λ b y a Λ -b no pueden ser ambos el caso

1. Exahustivos : Al menos uno es el caso es decir la disyunción de todos los eventos atómicos debe ser verdadera

1. Cualquier evento atómico implica la verdad o falsedad de cada proposición ejemplo: a Λ -b implica verdad para a y falsedad para b

1. Cualquier proposición es lógicamente equivalente a la disyunción de todos los elementos atómicos que implican la verdad de la proposiciónejemplo a es equivalente a la disyunción de eventos atómicos a Λ b y a Λ -b

Probabilidad, Notación básica

- Proposiciones: es un evento (conjunto de puntos de muestra)

  Dadas las variables aleatorias booleanas A y B

  - evento a : conjunto de puntos de muestras donde

    $$A(w) = true$$

  - evento $-a$: conjunto de puntos de muestras donde

    $$A(w) = false$$

  - evento $a \wedge b$:conjunto de puntos de muestras donde
  - $A(w) = true$ y $B(w) = true$

Probabilidad, Notación básica

- Con variables booleanas:
  - puntos de muestra = modelo proposicional lógico
    Ejemplo: $A(w) = true$, $B(w) = true$ ó $-a \wedge b$

- Proposición = disyunción de eventos atómicos en los que es verdad

$$(a \vee b) \equiv (\neg a \wedge b) \vee (a \wedge \neg b) \vee (a \wedge b)$$

$$\Rightarrow P(a \vee b) = P(\neg a \wedge b) + P(a \wedge \neg b) + P(a \wedge b)$$

Probabilidad, Notación básica

- Ejemplos de proposiciones

  A=verdadero

  Tiempo = lluvioso

  Temperatura = 21.6

  Temperatura <= 21.6

Probabilidad, Notación básica

- Probabilidad a priori

  llamada también propabilidad no condicional

  ejemplo

  P(A=true) = 0.2 P(Tiempo = soleado)=0.7

- Distribución de probabilidad

  valores para todos las posibles asignaciones

  P(Tiempo) =[0.7,0.2,0.08,0.02] (normalizadas suman 1)

Probabilidad, Notación básica

- **Distribución de probabilidad Conjunta**

Para un conjunto de variables aleatorias da la probabilidad de cada evento atómico en aquellas variables aleatorias

P(Tiempo,Caries) = Una matriz de 4 × 2 valores

Tiempo =	Sol	lluvia	nubes	nieve
Caries = true	0.144	0.02	0.016	0.02
Caries = false	0.576	0.08	0.064	0.08

Toda pregunta acerca de un dominio puede ser respondida por la distribución de probabilidad conjunta

Probabilidad, Notación básica

- **Distribución de probabilidad Conjunta completa**

  Conjunto completo de variables aleatorias usadas para describir el mundo, es una distribución de probabilidad conjunta pero que usa todo el conjunto de variables aleatorias por ejemplo: si el mundo consiste de tres variables A,B,C, entonces la distribución de probabilidad conjunta completa esta dado por

  $$P(A,B,C)$$

  Esta probabilidad conjunta puede ser representada por una tabla de 2x2x4

Probabilidad, Notación básica

- Probabilidad para variables aleatorias continuas

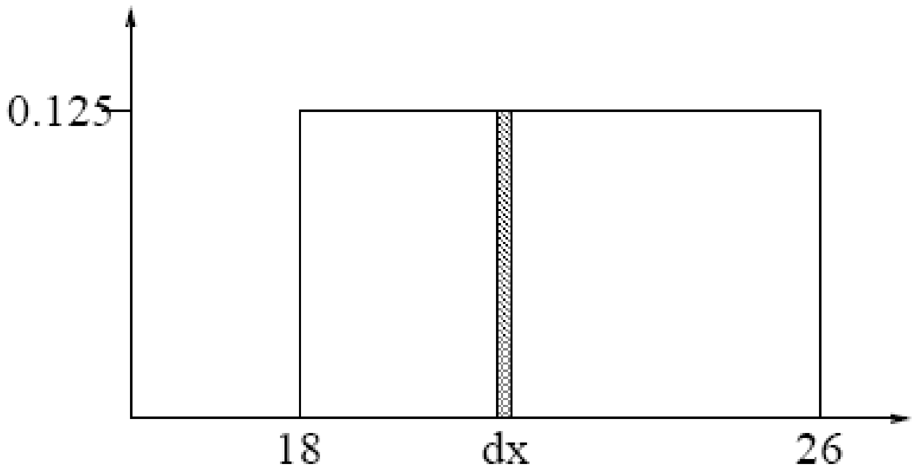
No es posible escribir la entera distribución en una tabla pues tiene infinitos valores,

La variable aleatoria toma su valor x de alguna función parametrizada de x

Las distribuciones de probabilidad para variables continuas son llamadas funciones de densidad de probabilidad

Probabilidad, Notación básica

Ejemplo de una distribución de densidad uniforme entre 18 y 26 $P(X = x) = U[18, 26](x)$



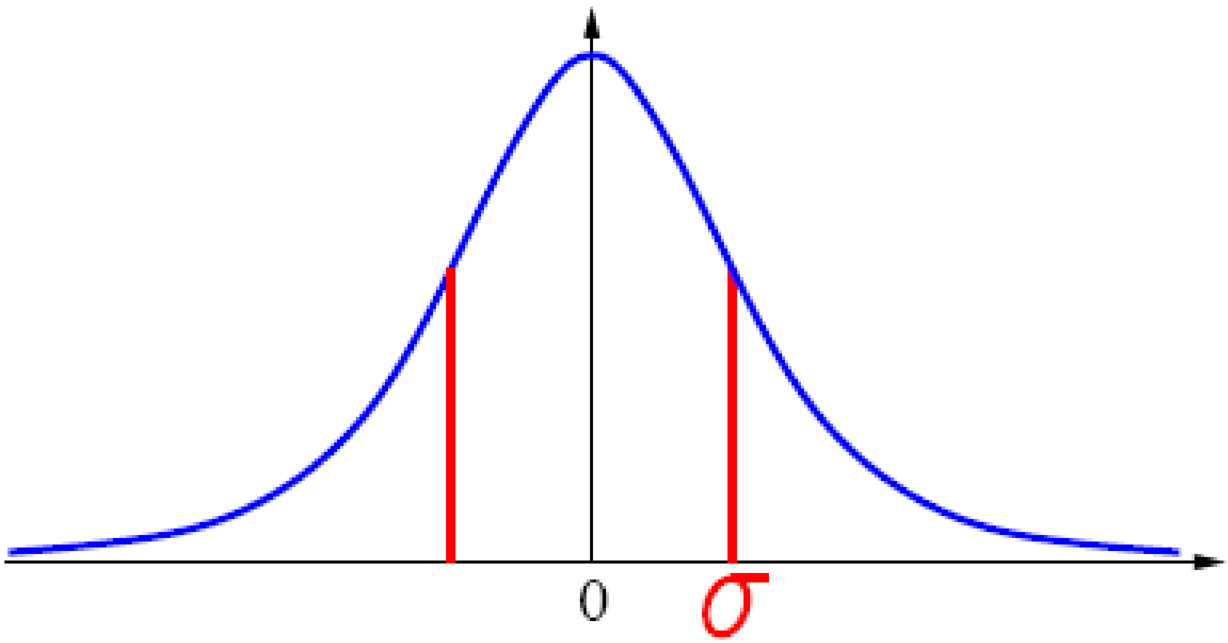
$P(X = 20.5) = 0.125$

$\lim_{dx \to 0} P(20.5 \leq X \leq 20.5 + dx)/dx = 0.125$

Probabilidad, Notación básica

- Densidad Gaussiana

$$P(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$$



Probabilidad, Notación básica

- Probabilidad Condicional (Probabilidad posterior)

$$P(a|b) = \frac{P(a \wedge b)}{P(b)} \text{ if } P(b) \neq 0$$

Regla del producto formulación alternativaç

$$P(a \wedge b) = P(a|b)P(b) = P(b|a)P(a)$$

Por ejemplo P(caries\dolor) =0.6

P(caries,dolor)=P(caries\dolor)P(dolor)=0.12

Probabilidad, Notación básica

Regla de la cadena es derivada por aplicaciones sucesivas de la Regla del producto

$$
\begin{aligned}
P(X_1, \ldots, X_n) &= P(X_1, \ldots, X_{n-1})\, P(X_n | X_1, \ldots, X_{n-1}) \\
&= P(X_1, \ldots, X_{n-2})\, P(X_{n_1} | X_1, \ldots, X_{n-2})\, P(X_n | X_1, \ldots, X_{n-1}) \\
&= \ldots \\
&= \Pi_{i=1}^{n} P(X_i | X_1, \ldots, X_{i-1})
\end{aligned}
$$

Probabilidad, Notación básica

- La siguiente notacion para distribuciones condicionales indicara todos los posibles valores para ***i,j***

$$P(X|Y) \text{ dado } P(X=x_i, Y= y_j)$$

por ejemplo

$$P(X,Y)=P(X|Y)P(y)$$

$$P(X=x_1 \wedge Y= y_1) = P(X=x_1| Y= y_1)P(Y=y_1)$$

$$P(X=x_1 \wedge Y= y_2) = P(X=x_1| Y= y_2)P(Y=y_2)$$

Axiomas de la Probabilidad

Axiomas de la Probabilidad

- Axiomas de Kolmogorov
  1. Todas las probabilidades estan entre cero y uno para cualquier proposución a

$$0 \leq P(a) \leq 1$$

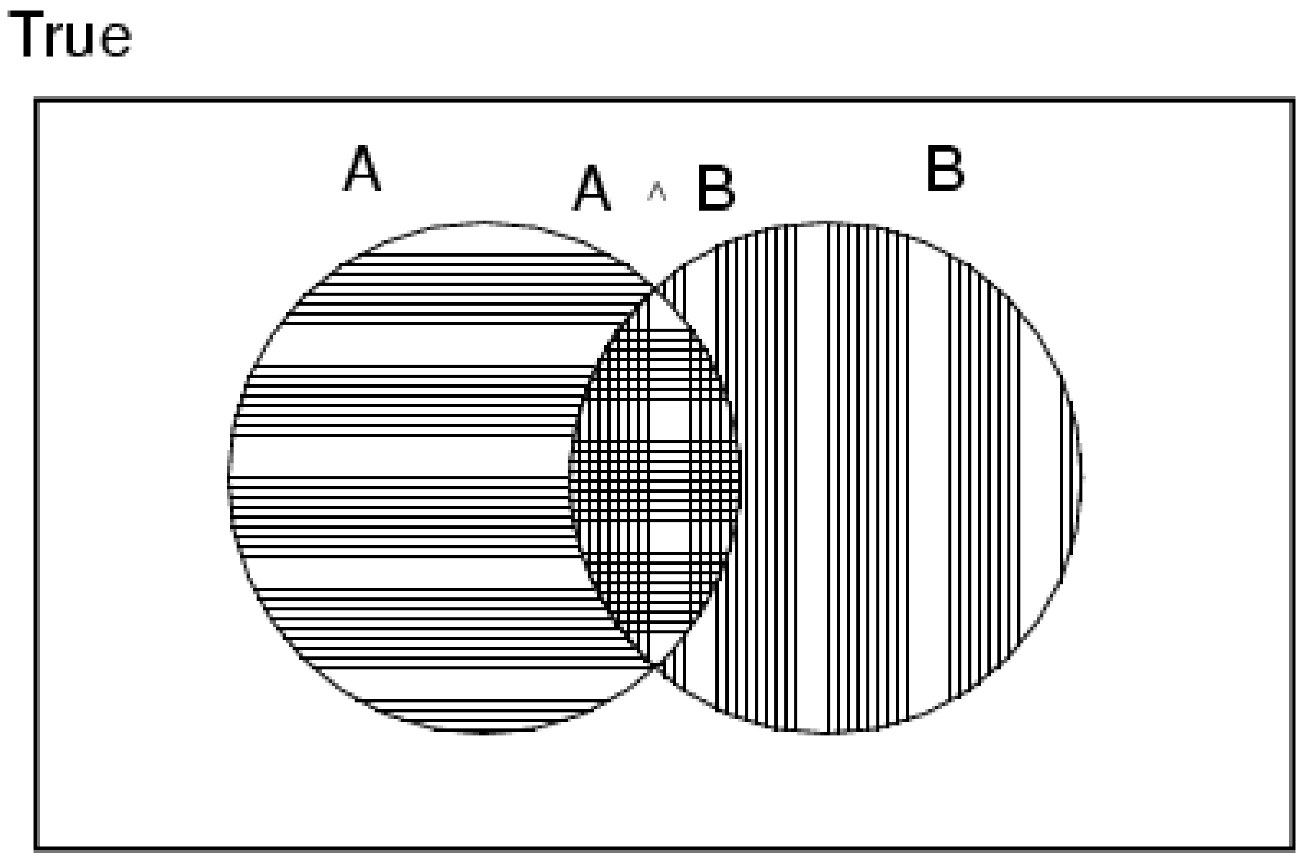
  2. Proposiciones necesariamente verdaderas tienen probabilidad 1 y proposiciones necesariamente falsas tienen probabilidad 0

$$P(true) = 1 \quad P(false) = 0$$

  3. La probabilidad de una disyunción está dada mediante:

$$P(a \vee b) = P(a) + P(b) - P(a \wedge b)$$

Axiomas de la Probabilidad



Inferencia Usando Distribucion conjunta completa

Inferencia Usando Distribucion conjunta completa

- Llamada también inferencia por enumeración

	dolor		-dolor	
	extraer	-extraer	extraer	-extraer
caries	0.108	0.012	0.072	0.008
-caries	0.016	0.064	0.144	0.576

- P(dolor) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2

Inferencia Usando Distribucion conjunta completa

	dolor		-dolor	
	extraer	-extraer	extraer	-extraer
caries	0.108	0.012	0.072	0.008
-caries	0.016	0.064	0.144	0.576

- P(caries V dolor) = 0.108 + 0.012 + 0.016 + 0.064+ 0.073+0.008= 0.28

Inferencia Usando Distribucion conjunta completa

	dolor		-dolor	
	extraer	-extraer	extraer	-extraer
caries	0.108	0.012	0.072	0.008
-caries	0.016	0.064	0.144	0.576

- $P(-caries|dolor) = \dfrac{P(dolor \wedge -caries)}{P(dolor)} =$

Inferencia Usando Distribucion conjunta completa

	dolor		-dolor	
	extraer	-extraer	extraer	-extraer
caries	0.108	0.012	0.072	0.008
-caries	0.016	0.064	0.144	0.576

- $P(\text{-caries}|\text{dolor}) = \dfrac{P(\text{dolor} \wedge \text{-caries})}{P(\text{dolor})} =$

$= (0.016+0.064)/(0.108 + 0.012 + 0.016 + 0.064) = 0.4$

Inferencia Usando Distribucion conjunta completa

	dolor		-dolor	
	extraer	-extraer	extraer	-extraer
caries	0.108	0.012	0.072	0.008
-caries	0.016	0.064	0.144	0.576

- $P(caries|dolor) = \frac{P(dolor \land caries)}{P(dolor)} =$

$= (0.108+0.012)/(0.108 + 0.012 + 0.016 + 0.064) = 0.6$

Inferencia Usando Distribucion conjunta completa

- Normalización

  Puede ser visto como una constante de normalización α

$$P(Caries \mid dolor) = \alpha P(Caries, dolor)$$

$$= \alpha[P(Caries. dolor, extraer) + P(Caries. dolor, -extraer)]$$

$$= \alpha[(0.108, 0.016) + (0.012, 0.064)]$$

$$= \alpha(0.12, 0.08) = (0.6, 0.4)$$

Inferencia Usando Distribucion conjunta completa

Idea General: Calcular la distribución de la variable de consulta manteniendo fijas las variables de evidencia y sumando sobre las variables ocultas

X=variable de consulta (ejem: Caries),

E=variable de evidencia (ejem: Dolor)

e=valores observados para dicha variable

Y=las variables no observadas (ejemplo: extraer)

$$P(X\backslash e) = \alpha P(X,e) = \alpha \sum_{y} P(X,e,y)$$

Inferencia Usando Distribucion conjunta completa

- Análisis

  Para un dominio utilizando varibles booelanas , se requiere una tabla de tamaño $O(2^n)$ y un tiempo en procesar la tabla de $O(2^n)$

Problemas realistas tienen miles de variables

Independencia

Independencia

- A y B son independientes sii

  $P(A|B) = P(A)$ ó $P(B|A) = P(B)$ ó $P(A \wedge B) = P(A)P(B)$

Ejemplo:

Para nuestro ejemplo anterior agregamos una variable mas llamada Tiempo={sol, nubes, lluvia, nieve}

La distribución conjunta completa para

P(Dolor, Extraer, Caries, Tiempo)

será una tabla con 32 entradas

Independencia

- La tabla se ve reducida de 32 entradas a solo 12 si notamos la independencia de Tiempo frente a otras variables
- Es decir existe independencia

P(Dolor, Extraer, Caries, Tiempo) = P(Dolor, Extraer, Caries)P(Tiempo)

Ejemplo

P(dolor, extraer, caries, Tiempo=nubes)

= P(Tiempo=nublado\dolor, extraer, caries,) P(dolor, extraer, caries)

= P(Tiempo=nubes) P(dolor, extraer, caries)

Independencia

- Ejemplos de independencia

  El tiempo y los problemas dentales son independientes

  Jugar cara y sello en las monedas también es independiente

La independencia absoluta es poderosa pero rara

¿Qué hacer?

Independencia: Independencia condicional

- **P**(*Dolor, Caries, Extraer*) tiene $2^3 - 1 = 7$ entradas independientes
- Si se tiene caries, La probabilidad de que se realize extraccion no depende de el dolor:
  (1) **P**(*extrar* | *dolor, caries*) = **P**(*extraer* | *caries*)
- La misma independencia permanece si no se tiene caries:
  (2) **P**(*extraer*| *dolor,*¬*caries*) = **P**(*extraer* | ¬*caries*)
- *Extraer* es condicionalmente independiente de *Dolor* dado *Caries*:
  P(*Extraer*| *Dolor,Caries*) = **P**(*Extraer* | *Caries*)
- Estamentos equivalentes:
  P(*Dolor* | *Extraer, Caries*) = **P**(*Dolor* | *Caries*)
  P(*Dolor, Extraer*| *Caries*) = **P**(*Dolor* | *Caries*) **P**(*Extraer* | *Caries*)

Independencia: Independencia condicional

- Escribir la distribución conjunta total usando la regla de la cadena

  P(*Dolor, Extraer, Caries*)

  = **P**(*Dolor* | *Extraer, Caries*) **P**(*Extraer, Caries*)

  = **P**(*Dolor* | *Extraer, Caries*) **P**(*Extraer* | *Caries*) **P**(*Caries*)

  = **P**(*Dolor* | *Caries*) **P**(*Extraer* | *Caries*) **P**(*Caries*)

  I.e., 2 + 2 + 1 = 5 números independientes

- En muchos casos el uso de independencia condicionalreduce el tamáño de la representacion de la distribución conjunta de exponencial en *n* a lineal en *n*.□

- Independencia condicional es nuestra mas básica y robusta forma de conocimiento en entornos inciertos.□

Independencia: Independencia condicional

- *X* es condicionalmente independiente de Y dado *Z* si la distribución de probabilidad que gobierna *X* es independiente del valor de *Y* dado el valor de *Z*:

- *P* (*X* |*Y*, *Z*) = *P* (*X* |*Z*) esto es si

$$(\forall x_i, y_j, z_k) P(X = x_i \mid Y = y_j, Z = z_k) = P(X = x_i \mid Z = z_k)$$

Independencia: Independencia condicional

- La definición general de independencia condicional de dos variables X y Y, dada una tercera variable Z es

$$P(X,Y \setminus Z) = P(X \setminus Z)P(Y \mid Z)$$

Regla de Bayes

Regla de Bayes

- Regla del Producto $P(a \wedge b) = P(a|b)P(b) = P(b|a)P(a)$

  $\Rightarrow$ Regla de Bayes: $P(a|b) = \dfrac{P(b|a)P(a)}{P(b)}$

- O a manera de distribución

$$\mathrm{P}(Y|X) = \frac{\mathrm{P}(X|Y)\mathrm{P}(Y)}{\mathrm{P}(X)} = \alpha \mathrm{P}(X|Y)\mathrm{P}(Y)$$

- Util para evaluar la probabilididad de un diagnostico de una probabilidad causal:
  - P(Causa|Efecto) = P(Efecto|Causa) P(Causa) / P(Efecto)
  - Ejemplo., *M* = meningitis, *S* Rigidez en el cuello:

$$P(m|s) = \frac{P(s|m)P(m)}{P(s)} = \frac{0.8 \times 0.0001}{0.1} = 0.0008$$

Nota: La probabilidad a postiori de la meningitis es muy pequeña!

Regla de Bayes : independencia condicional

$\mathbf{P}(\mathit{Caries} \mid \mathit{dolor} \wedge \mathit{extraer})$

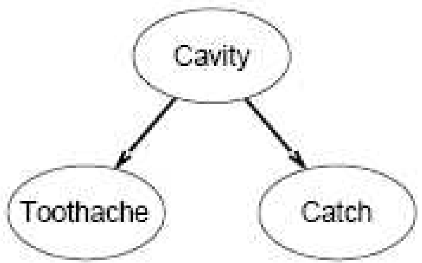
$= \alpha\mathbf{P}(\mathit{dolor} \wedge \mathit{extraer} \mid \mathit{Caries})\ \mathbf{P}(\mathit{Caries})$

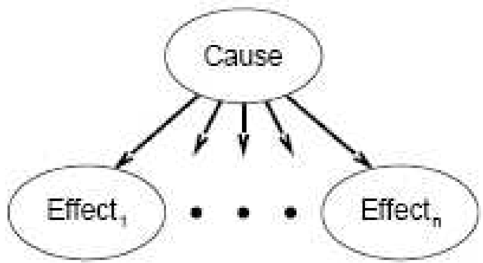
$= \alpha\mathbf{P}(\mathit{dolor} \mid \mathit{Caries})\ \mathbf{P}(\mathit{extraer} \mid \mathit{Caries})\ \mathbf{P}(\mathit{Caries})$

□

- Esto es un ejemplo del modelo ingenuo de Bayes

  $\mathbf{P}(\mathrm{Causa}, \mathrm{Efecto}_1, \ldots, \mathrm{Efecti}_n) = \mathbf{P}(\mathrm{Causa})\ \pi_i \mathbf{P}(\mathrm{Efecto}_i \mid \mathrm{Causa})$□□





- El número de parámetros es lineal en *n*

Ejemplo

Ejemplo

- P_{ij} =true sii [i; j] contiene un pozo
- B_{ij} =true sii [i; j] tiene brisa
- Incluimos solamente $B_{1;1}$;$B_{1;2}$;$B_{2;1}$ en el modelo de probabilidad

Ejemplo: especificando el modelo de probabilidad

- La distribucion conjunta completa es

$$\mathbf{P}(P_{1,1}, \ldots, P_{4,4}, B_{1,1}, B_{1,2}, B_{2,1})$$

- Aplicando la regla del producto

$$\mathbf{P}(B_{1,1}, B_{1,2}, B_{2,1} \mid P_{1,1}, \ldots, P_{4,4})\mathbf{P}(P_{1,1}, \ldots, P_{4,4})$$

- Hacerlo de tal manera de obtener P(efecto\causa)
- Reglas 1 si exiten pozos adyacentes a las brisas, o cc pozos estan ubicados aleatoriamente con una probabilidad de 0.2 por cuadrado

$$\mathbf{P}(P_{1,1}, \ldots, P_{4,4}) = \Pi_{i,j=1,1}^{4,4} \mathbf{P}(P_{i,j}) = 0.2^n \times 0.8^{16-n}$$

Para n pozos

Ejemplo: observaciones y consulta

- Se conocen los siguientes hechos

$$b = \neg b_{1,1} \wedge b_{1,2} \wedge b_{2,1}$$

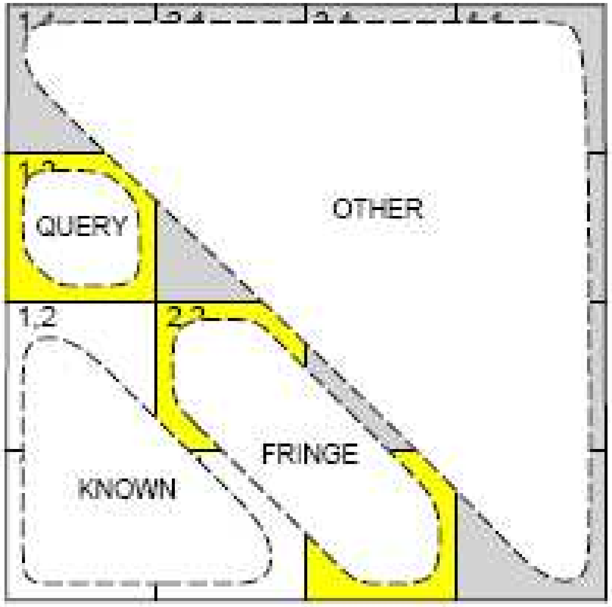
$$conocido = \neg p_{1,1} \wedge \neg p_{1,2} \wedge \neg p_{2,1}$$

- La consulta es $\boxed{P(p_{1,3} \setminus b, conocido)}$
- Definimos desconocido como todos los p_{ij} diferentes de p_{13}
- Por inferencia mediante ennumeracion (distribucion conjunta completa tenemos)

$$P(p_{1,3} \setminus b, conocido) = \alpha \sum_{desconocido} P(p_{1,3}, desconocido, b, conocido)$$

- Crece exponencialmente con el número de cuadrados

Ejemplo: Usando independencia condicional



Desconocido = Fringe U Otro

$$P(b \setminus p_{1,3}, conocido, desconocido) = P(b \setminus p_{1,3}, conocido, fringe)$$

Ejemplo: Usando independencia condicional

$$P(p_{1,3} \backslash b, conocido) = \alpha \sum_{desconocido} P(p_{1,3}, b, conocido, desconocido)$$

$$= \alpha \sum_{desconocido} P(b \backslash p_{1,3}, conocido, desconocido) P(p_{1,3}, conocido, desconocido)$$

$$= \alpha \sum_{fringe} \sum_{otro} P(b \backslash p_{1,3}, conocido, fringe, otro) P(p_{1,3}, conocido, fringe, otro)$$

$$= \alpha \sum_{fringe} \sum_{otro} P(b \backslash p_{1,3}, conocido, fringe) P(p_{1,3}, conocido, fringe, otro)$$

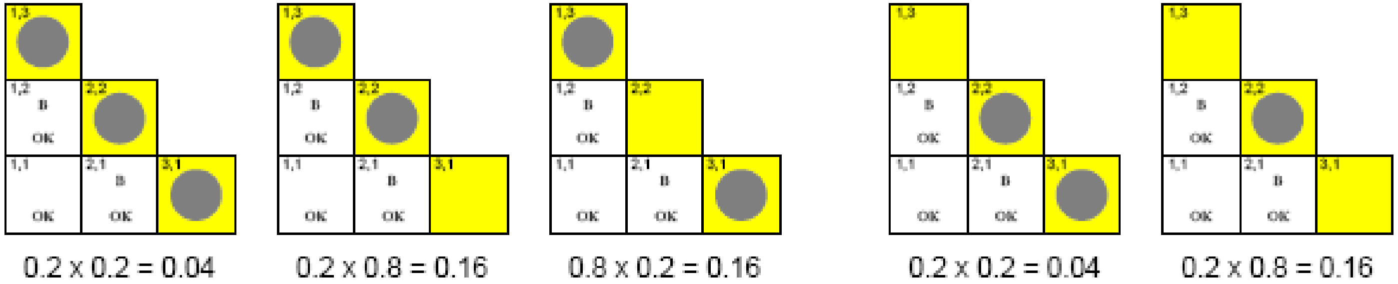
$$= \alpha \sum_{fringe} P(b \backslash p_{1,3}, conocido, fringe) \sum_{otro} P(p_{1,3}, conocido, fringe, otro)$$

$$= \alpha \sum_{fringe} P(b \backslash p_{1,3}, conocido, fringe) \sum_{otro} P(p_{1,3}) P(conocido) P(fringe) P(otro)$$

$$= \alpha P(p_{1,3}) P(conocido) \sum_{fringe} P(b \backslash p_{1,3}, conocido, fringe) P(fringe) \sum_{otro} P(otro)$$

$$= \alpha' P(p_{1,3}) \sum_{fringe} P(b \backslash p_{1,3}, conocido, fringe) P(fringe)$$

Ejemplo: Usando independencia condicional



$$P(p_{1,3} \setminus conocido, b) = \alpha'[0.2(0.04+0.16+0.16), 0.8(0.04+0.16)]$$

$$= [0.31, 0.69]$$

$$P(p_{2,3} \setminus conocido, b) = [0.86, 0.14]$$

Resumen

- La probabilidad es un formalismo riguroso para conocimiento incierto□
- La distribucion conjunta de la probabilidad especifica la probabilidad de cada evento atomico
- Las consultas pueden ser respondindas sumando todos los eventos atómicos□
- Para dominios no triviales, de debe buscar una manera de reducir el tamaño conjunto□
- Independencia y independencia condicional proveen las herramientas□

Referencias Bibliográficas

- Capitulo 13 Artificial Intelligence: A Modern Approach, Russell and Norvig
- http://www.aaai.org/AITopics/pmwiki/pmwiki.php/AITopics/Uncertainty