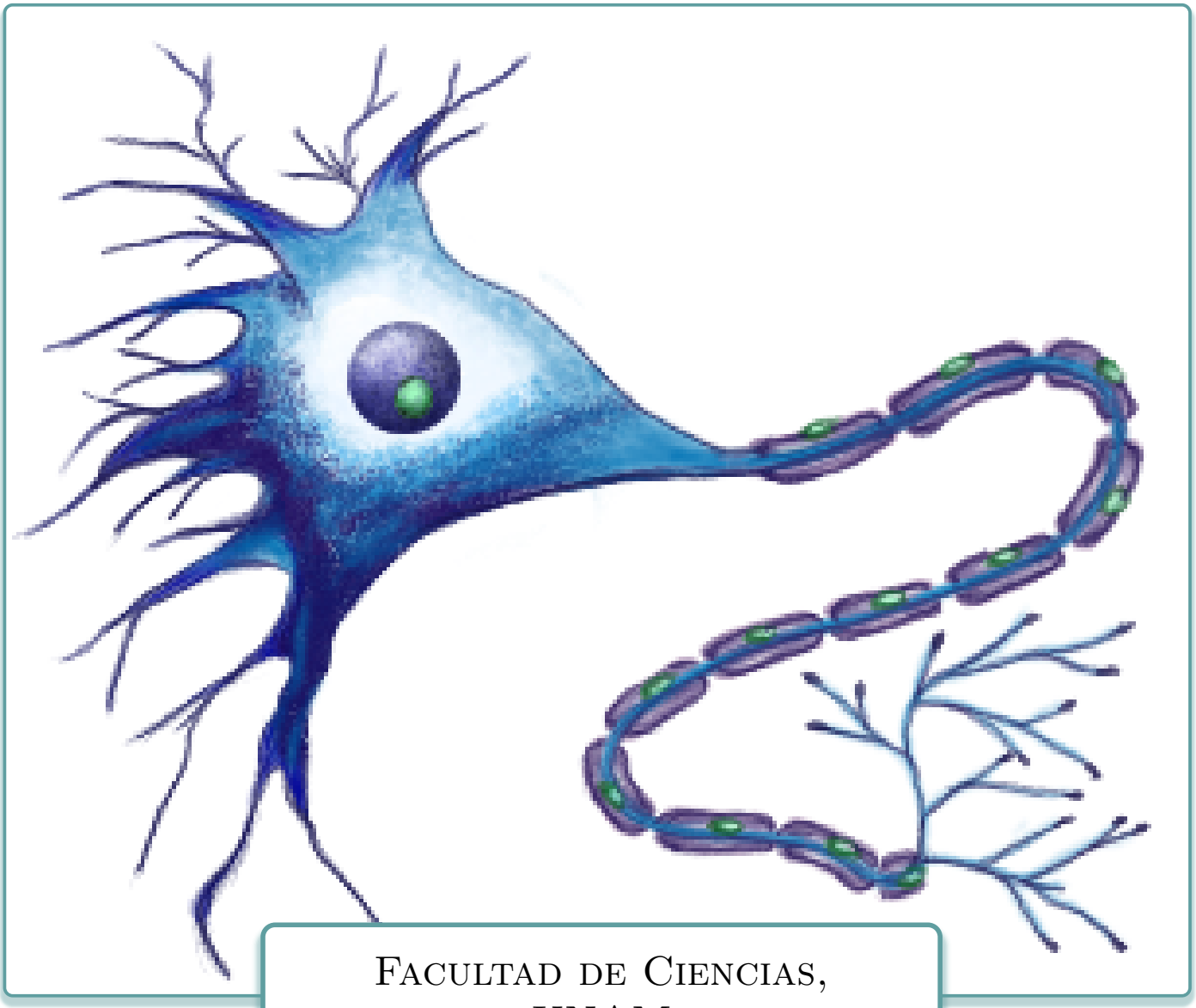
Redes Neuronales

Notas de clase

Karla Fernanda Jiménez Gutiérrez
Verónica Esther Arriola Ríos



Facultad de Ciencias,
UNAM

Índice general

Convenciones

A lo largo del texto se utilizará la siguiente notación para diversos elementos:

Conjuntos	C
Vectores	$\boldsymbol{x}$
Matrices	$\boldsymbol{M}$
Unidades	cm

Parte I
Introducción

1 | Historia de IA

Problemas representativos

Problema del marco

Agentes

Parte II
Resolución de problemas mediante búsqueda

2 | Espacio de estados

Espacio de estados continuo en física

Espacio de estados discreto

Mundo de juguete

3 | Búsqueda

Búsqueda ciega

Problemas de satisfacción de restricciones

Definición

Definición 3.1

Un problema de satisfacción de restricciones está definido por:

- Un conjunto de variables $V = \{X_1, X_2, ..., X_n\}$
- Un conjunto de restricciones $C = \{C_1, C_2, ..., C_m\}$
- Cada variable tiene un dominio asociado $\mathbb{D}_v$ con $\mathbb{D}_v \neq \emptyset$

Para este tipo de problemas

- Un estado es una asignación a unas o todas las variables $S = \{X_i = v_i, X_j = v_j, ...\}$
- Una *asignación consistente* es una asignación que no viola ninguna restricción.
- Una *asignación completa* es una asignación que menciona a todas las variables.
- Una solución es una asignación completa y consistente.

Entonces el sistema de transiciones en el espacio de estados $\Sigma = (S, A, \gamma)$, queda definido con los elementos siguientes:

- $S = \{s_1, s_2, ...\}$ el conjunto de todas las posibles asignaciones (parciales y completas) a las variables del problema.

	1	2	3	4	5	6	7	8	9
1		7				1			
2	5	3						6	
3						8			7
4	7				5			9	
5		4		1			5		
6	9			6		7	3	2	
7		6		3					
8	8		3		4		7		1
9	9					7			

Figura 3.1 Tablero de Sudoku al iniciar un juego.

- $A = \{a_1, a_2, ...\}$ el conjunto de las acciones que asignan a alguna variable X_i un valor v del dominio $\mathbb{D}_{X_i}$.
- $\gamma : S \times A \to S$ la función que aplica la acción a_i en el estado s_j, si al realizar la asignación a_i, s_j no queda en un estado inconsistente. Para todos los demás casos, devuelve $\emptyset$ y se dice que la acción no es aplicable.

De este modo, el problema de planeación $\mathcal{P} = (\Sigma, s_i, g)$ usa:

- $s_i = \{\}$ la asignación vacía.
- $g(s)$ la función que verifica si la asignación es completa.

Recorrido en el espacio de búsqueda

Para mostrar cómo se realiza una búsqueda en este espacio se utilizará el problema del *Sudoku*. Utilicemos como ejemplo el tablero mostrado en la Figura 3.1.

En este escenario, las variables son las casillas vacías del Sudoku. Sea X el conjunto de todas las casillas. Identifiquemos a cada casilla con el símbolo X_{ij}, con i y j en $\{1, 2, ..., 9\}$, según su posición en el tablero. Los valores posibles para cada casilla están en $\mathbb{D}_v = \{1, 2, ..., 9\}$.

Dado que el juego del Sudoku comienza con valores en algunas casillas, el estado inicial ya tiene valores asociados a algunas de las X_{ij}. Para el tablero del ejemplo se tendría la asignación inicial F:

$$\begin{aligned} F = \{&X_{12}=7, \quad X_{16}=1, \\ &X_{21}=5, \quad X_{22}=3, \quad X_{28}=6, \\ &\ldots \\ &X_{81}=8, \quad X_{83}=3, \quad X_{85}=4, \quad X_{87}=7, \quad X_{89}=1, \\ &X_{91}=9, \quad X_{96}=7\} \end{aligned}$$

Sin embargo, dado que los valores de estas casillas no pueden ser cambiados, este conjunto de casillas no pertenece al conjunto de variables. Más bien se tiene:

$$V = X - F$$

Estimemos la complejidad inicial de este problema, calculando el número de estados posibles con asignaciones completas:

1. Hay $|V|$ casillas vacías, cada una de las cuales puede albergar 9 valores distintos.
2. Esto da un total de $9^{|V|}$ asignaciones posibles.

En el ejemplo anterior, son $9^{54} \approx 3.5 \times 10^{51}$. Compárese con la capacidad de un disco duro de un terabyte, donde $1T \approx 10^{12}$bytes. No intentemos siquiera agregar las asignaciones incompletas. Claramente no es posible resolver este problema sin hacer uso de las restricciones, para evitar recorrer este espacio de estados.

Propagación de restricciones

Por la forma en que se planteó el problema en la sección anterior, se comienza la búsqueda de un solución, con la asignación vacía $s_i = \{\}$. ¿Cuáles son las acciones aplicables en este momento? Para comenzar: las acciones posibles son aquellas que asignan un valor a alguna de las casillas vacías. Obsérvese que se estableció como precondición para su aplicabilidad, que el valor asignado no viole ninguna de las restricciones del problema. En el caso del Sudoku, éstas son:

1. Que no haya otra casilla en el mismo renglón, con el mismo valor.
2. Que no haya otra casilla en la misma columna, con el mismo valor.

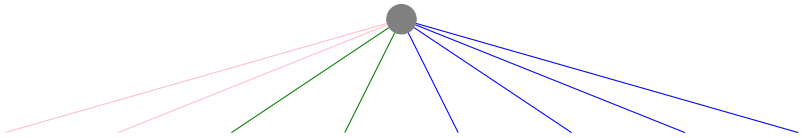


Figura 3.2 Grupos de acciones aplicables y sus sucesores para un problema de satisfacción de restricciones.

3. Que no haya otra casilla en el mismo cuadrante, con el mismo valor.

Consideremos ahora al conjunto $\mathbb{D}_\nu$ de valores posibles para cada variable. Estas tres condiciones implican que, cada vez que se asigne un valor ν, a una casilla X_{ij}, los valores que permiten una asignación consistente se reducen para las casillas siguientes:

1. Todas las casillas en el renglón i, ya no pueden tener a ν como valor.
2. Todas las casillas en la columna j, ya no pueden tener a ν como valor.
3. Todas las casillas en el cuadrante i, ya no pueden tener a ν como valor.

Antes de elegir la acción a aplicar, es necesario *propagar* el efecto de estas restricciones, para determinar cuáles son las acciones aplicables en un estado dado. Depués de hacer esto, quedarán grupos de acciones aplicables: un grupo por cada variable que aún no ha sido asignada, con tantas acciones asociadas, como valores sea posible asignarle. Visualicemos esto como en la Figura 3.2.

Todas las variables que no han sido asignadas deberán recibir alguna asignación en algún momento. Sin embargo, podemos seleccionar cual asignaremos primero y esto acelerará la búsqueda drásticamente. También podemos elegir qué valor probaremos primero, de entre los que aún se le puede asignar. Hay tres reglas que sugieren cómo realizar estas elecciones:

Búsqueda informada: uso de heurísticas

Alpinismo de colinas

Recocido simulado

A*

Búsqueda con adversarios

Min-max

Poda $\alpha - \beta$

Sistemas de planeación

STRIPS, PDDL y CLIPS

Hipótesis del mundo cerrado

Planeación con órdenes parciales (PoP)

Redes de planeación jerárquica

Estados de creencia

Razonamiento no monótono (conocimiento incompleto)

4 | Representación del conocimiento

Marcos (Frames)

El sistema de Marcos en la base de la programación Orientada a Objetos actual, por lo que los conceptos aquí citados resultarán familiares. Sin embargo, la teoría de marcos cubre casos que la mayoría de los lenguades de programación populares evitan, especialmente la herencia múltiple. Además, el término *marco* en general puede referirse tanto a una instancia como a una clase Winston 1992.

Dos notaciones comunes para representar a los marcos son:

```
1  (nombre ranura(valor-de-ranura) ranura-1(valor-de-ranura-1)...)
```

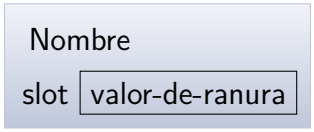


Figura 4.1 Diagrama de un marco.

Los marcos se originaron a partir de las redes semánticas. Una **red semántica** está compuesta por nodos y ligas entre ellos. Cada liga lleva por nombre el tipo de relación entre los nodos vinculados. Para reinterpretar una red semántica como un sistema de marcos se considera a cada nodo con las ligas vinculadas a él.

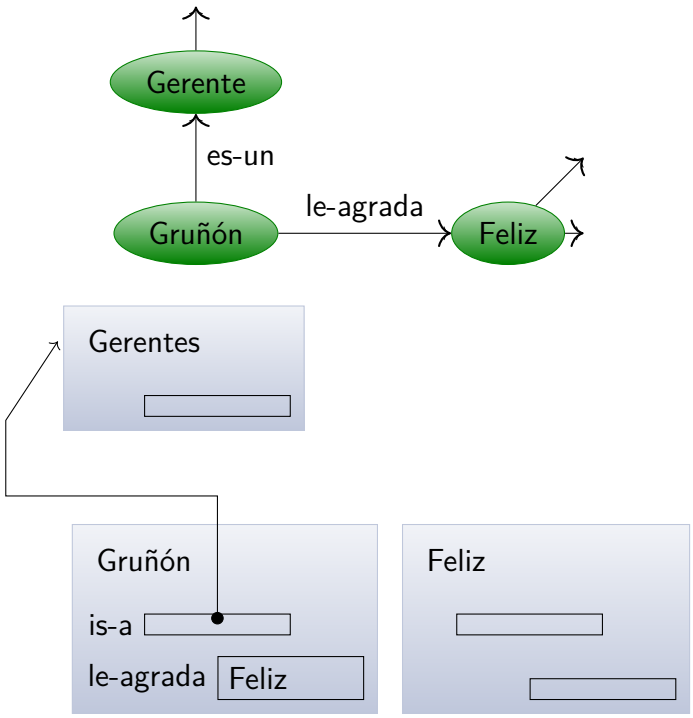


Figura 4.2 Arriba: red semántica. Abajo: diagrama de un marco; es posible usar ligas o nombres de otros marcos para llenar las ranuras.

Hay dos clases de marcos:

- Clases. a.k.o. = *a kind of* <class>
- Ejemplares (o instancias). is-a <class>

Procedimientos de acceso

Durante-la-construcción construye marcos instancia. Puede asignar valores por defecto a las ranuras mediante los **procedimientos-de-construcción** sugeridos por las clases y superclases de la instancia, de acuerdo con la **lista-de-precedencia**.

Durante-la-escritura Asigna valores a las rendijas. Ayudan a mantener restricciones entre ranuras relacionadas. Por ejemplo: si en la ranura *constitución-física* se escribe *delgado*, escribir *pequeño* en la ranura *apetito*.

Durante-la-lectura Devuelve los valores de las rendijas.

Cuando-se-solicite Sobrecargan los valores por defecto heredados. Pueden utilzar valores de otras ranuras que fueron llenadas durante la construcción de la instancia.

Con-respecto-a reciben como argumento el contexto en el que la propiedad debe ser evaluada. Por ejemplo, si se pide la estatura del enano Blimpy con repecto a un duente, el valor es *grande*. Si se pide su estatura con respecto a un enano promedio, tal vez el valor se *pequeño*.

Cuando-aplica semejante a los procedimientos *con-respecto-a*, pero para acciones. Calculan el mecanismo adecuado para cada acción dependiendo del contexto. Por ejemplo: `comer(sopa)` utiliza una cuchara, mientras que `comer(ensalda)` requiere un tenedor.

Lista de precedencia

Hay dos criterios útiles para la obtención de una lista de precedencia:

1. Cada clase debe aparecer en la lista de precedencia antes que cualquiera de sus superclases.

2. Cada superclase directa de una clase dada debe aparecer en la lista de precedencia antes que cualquier superclase directa que se encuentre a su derecha.

Si el tipo de herencia es simple, basta con recorrer la gráfica de marco primero en profundidad. Si la herencia es múltiple, pero no se forman rombos en la jerarquía, basta con agregar la regla **up-to-join-proviso**. Esta regla indica que, cada superclase que sea visitada más de una vez durante el recorrido primero en profundidad, de izquierda a derecha, debe ser ignorada hasta que la clase se encontrada por última vez. Sin embargo, para el caso general, es necesario aplicar el algoritmo **ordenamiento topológico** [Algoritmo 1].

Algoritmo 1 Ordenamiento topológico.

listaMarcos ← Crear una lista de todos los marcos accesibles desde el nodo instancia a través de relaciones is-a y a.k.o, incluyendo al nodo instancia.
for all marco **in** listaMarcos **do**
    listaRestricciones.añadir(marco, primer clase a la izquierda)
    Añadir listaRestricciones cada par clase-clase de izquierda a derecha
    Si el marco no tiene clases o superclases añadirlo solo. Un elemento solo cuenta como estar a la izquierda del par.
end for
repeat
    Tomar al marco que aparezca a la izquierda de al menos un par, pero a la deracha de ninguno.
    if Hay más de un marco sólo a la izquierda **then**
        Recorrer la listaDePresedencia desde el final hacia el inicio.
        Seleccionar el marco que sea clase o superclase directa del elemento más cercano al final de la listaDePresedencia.
    end if
    Añadirlo a listaDePresedencia
    Eliminar a todos los pares donde aparezca.
until listaRestricciones $= \emptyset$

Aplicaciones

Extracción de información de textos en un contexto definido, así como producción de textos a partir de instancias con información. Almacenamiento de información para sistemas expertos como:

- CYC (http://www.cyc.com/ y http://www.cyc.com/platform/opencyc)
- OpenMind Common Sense http://commons.media.mit.edu/

Existe también una base datos del idioma Inglés donde las palabras están relacionadas unas con otras:

- WordNet http://wordnet.princeton.edu/

Otros

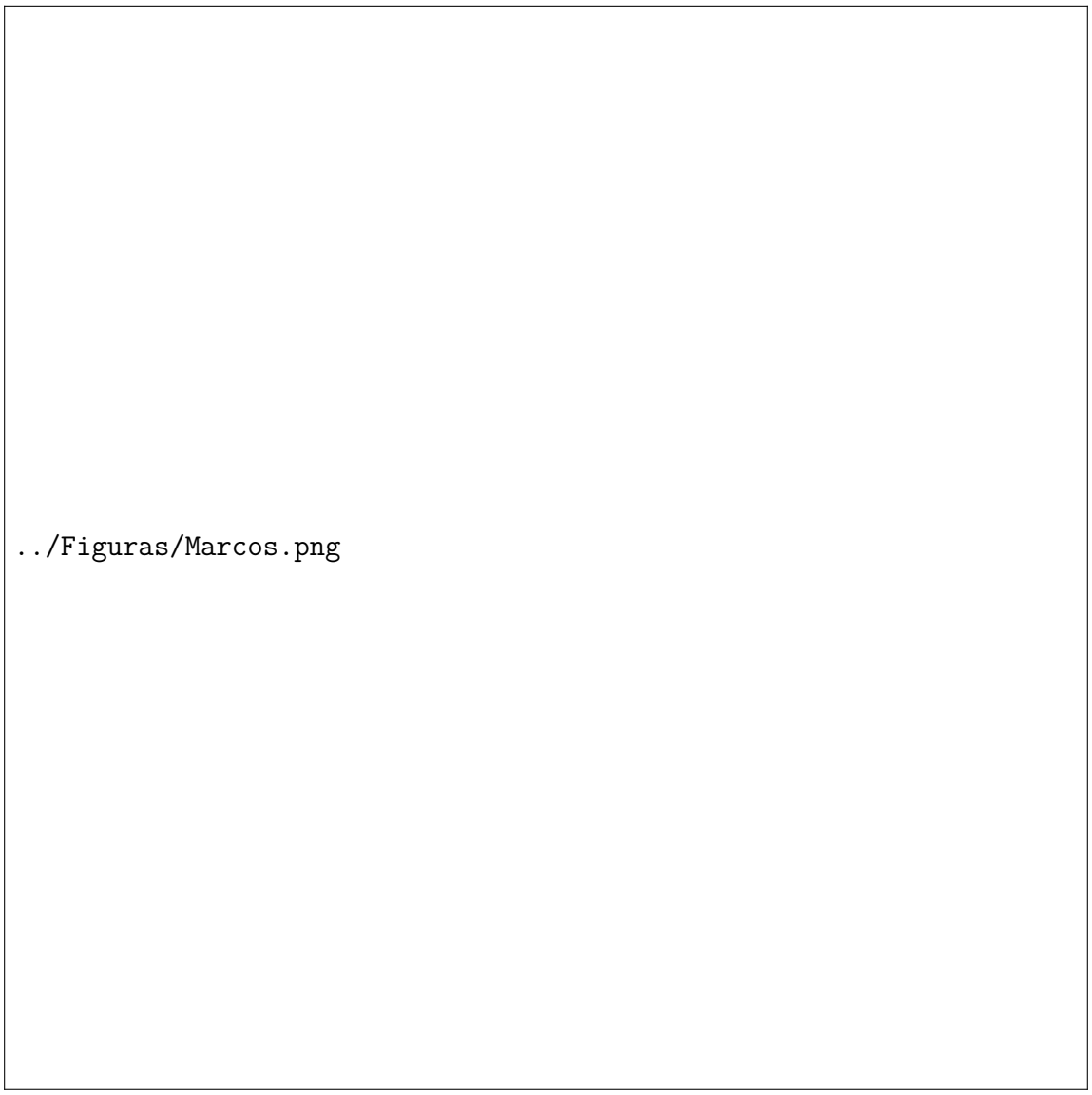


Figura 4.3 Objetos y relaciones diversas entre ellos. Particularmente: *un-tipo-de*, *es-un* que permiten heredar comportamientos.

Parte III
Aprendizaje de máquina

5 | Aprendizaje automático

Tipos de aprendizaje

Aprendizaje supervisado Aprender a predecir la salida, dado un vector de entrada.

Regresión La salida está en $\mathbb{R}^n$.

Clasificación La salida es discreta, usualmente etiquetas de clases.

Aprendizaje no supervisado Descubrir una buena representación para las entradas. Una buena representación tiene las características siguientes:

1. Es compacta, es una representación en pocas dimensiones de entradas en varias dimensiones.
2. Permite utilizar características que se pueden representar en forma económica (poco espacio).

Clasificación

Análisis de componentes principales

Aprendizaje reforzado Aprender a elegir una acción para maximizar una ganancia.

Sesgo inductivo

Conjuntos de entrenamiento, prueba y validación

Regresión polinomial

Mínimos cuadrados

Forma normal

Descenso por el gradiente

Regularización

Clasificación

Regresión logística

Regularización

Redes neuronales

Perceptrón

Compuertas lógicas

Evaluación: Propagación hacia adelante

Entrenamiento: Propagación hacia atrás

Árboles de decisión

Aprendizaje por refuerzo

K-medias

Parte IV
Razonamiento Bayesiano

6 | Antecedentes de probabilidad

Definiciones iniciales

Un *experimento causal, indeterminista o de azar* es aquel para el cual no necesariamente podemos predecir con certeza lo que va a ocurrir al realizarlo. Cuando se repite el mismo experimento bajo las mismas condiciones se puede obtener un resultado diferente.

Ej: "predecir el resultado de lanzar dos monedas."

Evento simple. Cualquier resultado elemental de un experimento aleatorio.

Ej: a: "obtener el 2 al lanzar un dado."

Evento. Cualquier conjunto de resultados posibles de un experimento aleatorio.

Ej: A: "obtener un par al lanzar un dado."
$A = \{2, 4, 6\} = \{$ "obtengo el dos", "o... cuatro", "o... seis"$\} = \{x | x = 2, 4, 6\}$

Espacio de eventos, muestras o Universo S. Conjunto de todos los eventos simples posibles en un experimento aleatorio.

Ej: "lanzar un dado."
$S = \{1, 2, 3, 4, 5, 6\}$

Tamaño de un evento. Número de eventos simples que satifacen la definición del evento (cardinalidad del conjunto).

Ej: tamaño de A: $|A| = 3$.

Definición 6.1

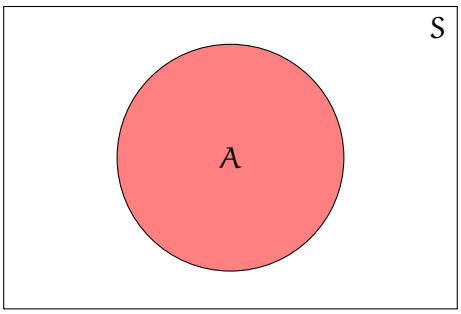


Figura 6.1 Evento A en el espacio de eventos S.

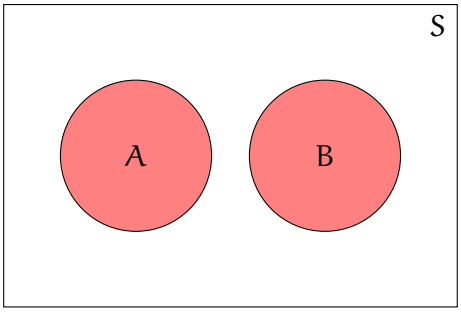


Figura 6.2 Eventos mutuamente exclusivos A y B en el espacio de eventos S.

Definimos a la probabilidad de un evento A como:

$$P(A) = \frac{|A|}{|S|}. \tag{6.1}$$

Destaca también lo que se conoce como la *definición frecuentista* de probabilidad, en la cual esta cantidad $P(A)$ se define con respecto a una secuencia potencialmente infinita de repeticiones del experimento aleatorio A.

Es posible representar gráficamente a los eventos utilizando *diagramas de Venn*. Por ejemplo, sea S el Universo y A un evento, su diagrama correspondiente se muestra en la Figura 6.1.

Se dice que dos eventos son *mutuamente exclusivos* si no pueden ser verdad al mismo tiempo y se pueden visualizar como dos conjuntos ajenos Figura 6.2.

Eventos posibles incluyen al universo S y al conjunto que no contiene resultados ϕ. Dados dos eventos E y F es posible crear otros eventos mediante las operaciones:

Unión. La unión de dos eventos $E \cup F$ es el conjunto de eventos simples en E o F inclusivamente. Figura 6.3 (Izquierda).

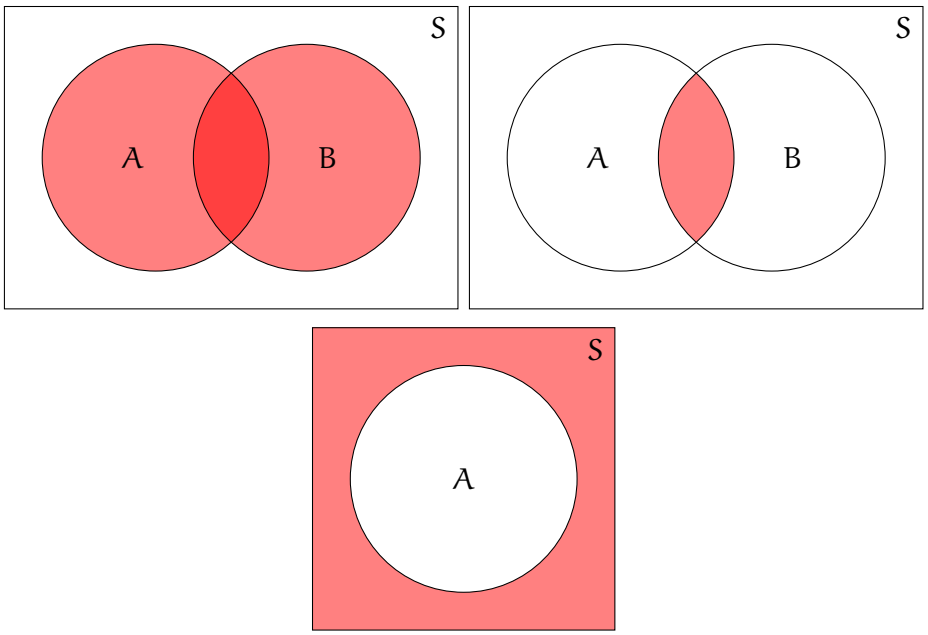


Figura 6.3 Arriba izquierda: Unión de eventos. Arriba derecha: Intersección. Abajo: Complemento.

Intersección. La intersección de dos eventos $E \cap F$ es el conjunto de eventos simples en ambos E y F al mismo tiempo. Figura 6.3 (Derecha).

Complemento. El complemento de un evento es el conjunto de los eventos simples que no se encuentran en E, denotado $\bar{E}$.

Variable aleatoria

Una variable aleatoria es una función que va del espacio de muestras a los reales $X : S \to \mathbb{R}$[1]. Esta función permite medir el aspecto de interés para un evento dado, en función de los resultados elementales obtenidos.

Ej: Para el evento E:"obtener un 7 al lanzar dos dados".
Sea la variable aleatoria X : "La suma del resultado en cada uno de dos dados".
$X(\mathrm{dado}_1, \mathrm{dado}_2) = \mathrm{dado}_1 + \mathrm{dado}_2$
Entonces el evento E se expresa como $E : X(\mathrm{dado}_1, \mathrm{dado}_2) = 7$.

Obsérvese que la función de la variable aleatoria puede ser, como un caso particular, la función identidad si el espacio de eventos está contenido en los reales. Otro caso

[1]Por convensión, escribiremos las variables aleatorias con mayúsculas y sus valores concretos con minúsculas.

particular sería una función que realice un mapeo uno a uno del espacio de eventos hacia los reales. Para ello es necesario que los valores en el dominio sean:

1. Mutuamente exclusivos.
2. Exhaustivos.

Ej: Formalmente la variable aleatoria $\mathrm{C}(Clima)$ tendría el rango:

$$\mathrm{C}(\mathrm{Clima}) =< 1, 2, 3, 4, 5 >$$

Dichos valores podrían venir de un mapeo directo con los valores del espacio de eventos:

$$Clima =< despejado, lluvioso, nublado, granizado, nevado >$$

$$\mathrm{C}(Clima) = \begin{cases} 1 & \text{si } despejado \\ 2 & \text{si } lluvioso \\ 3 & \text{si } nublado \\ 4 & \text{si } granizado \\ 5 & \text{si } nevado \end{cases}$$

Esto también se puede escribir $\mathrm{C} \in \{\mathrm{c}^1, \mathrm{c}^2, \mathrm{c}^3, \mathrm{c}^4, \mathrm{c}^5\}$, que abrevia los casos $\mathrm{C} = 1, \mathrm{C} = 2, \mathrm{C} = 3, \mathrm{C} = 4, \mathrm{C} = 5$.

De este modo es posible asociar una variable aleatoria a cada variable natural del espacio de eventos. Por comodidad y legibilidad, frecuentemente se hará referencia a la variable del espacio de eventos como una *variable aleatoria*, con un dominio diferente a los reales, aunque formalmente debe sobre-entenderse la existencia de un mapeo entre los elementos del espacio de eventos y los número reales.

Variables aleatorias discretas y continuas

Es posible clasificar a las variables aleatorias de acuerdo a las características de su rango en:

Discretas. Su rango consiste en un número finito de valores.

Un caso particular de variables discretas son las variables booleanas, donde los valores que pueden tomar son $\{0, 1\}$.

Es posible asignar una probabilidad a cada valor de una variable aleatoria discreta, definiendo lo que será entonces una *distribución de probabilidad*. Dado que los valores de la variable aleatoria son exclusivos, la suma de las probabilidades sobre todos los valores debe ser 1.

Ej:

$$P(\text{Clima}) = \begin{cases} 0.4 & \text{si } \textit{despejado} \\ 0.25 & \text{si } \textit{lluvioso} \\ 0.15 & \text{si } \textit{nublado} \\ 0.1 & \text{si } \textit{granizado} \\ 0.1 & \text{si } \textit{nevado} \end{cases}$$

Continuas. Al realizar un experimento aleatorio todos los valores sobre $\mathbb{R}$ o un intervalo $[a, b] \in \mathbb{R}$ son resultados posibles.

Ej: "$\text{Temperatura} \in [0, 60000]°\text{K}$".

Dado que el número de elementos en un intervalo continuo es infinito, la probabilidad de obtener un valor específico es cero. Por consiguiente se define una *función de densidad de probabilidad*[2]

$$f(x) \geqslant 0 \tag{6.2}$$

a partir de la cual se calcula la probabilidad de que X tome algún valor dentro de un subconjunto mesurable $B \in \mathbb{R}$ de sus valores posibles:

$$P(X \in B) = \int_B f(x)dx \tag{6.3}$$

En particular, si B es un intervalo de $\mathbb{R}$ entonces:

$$P(a \leqslant X \leqslant b) = \int_a^b f(x)dx \tag{6.4}$$

Dado que X debe tomar algún valor en $\mathbb{R}$, entonces $P(X)$ debe satisfacer:

$$P(X \in (-\infty, \infty)) = \int_{-\infty}^{\infty} f(x)dx = 1 \tag{6.5}$$

Extensión de la lógica proposicional

Es posible utilizar la teoría de probabilidades para extender a la lógica proposicional. El tipo más sencillo de proposición a utilizar es la afirmación de que una variable aleatoria tiene un valor particular tomado de su dominio.

Ej: "$\text{Temperatura} = 310$".

Dicho esto, es posible asociar una probabilidad al evento de que una proposición dada sea verdadera o falsa. Gracias a ello será posible realizar inferencias en condiciones de incertidumbre.

[2] Obsérvese que, en principio, $f(x)$ puede tomar valores mayores que uno Barber 2012.

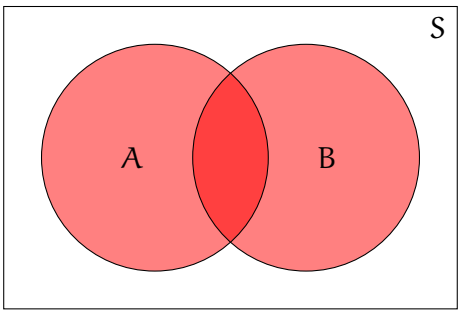


Figura 6.4 La probabilidad $P(a \vee b)$ corresponde a la probabilidad de la unión de ambos eventos.

Axiomas de la probabilidad

Andrei Kolmogorov demostró cómo desarrollar el resto de la teoría probabilista a partir de los tres axiomas que llevan su nombre[3]:

1. Todas las probabilidades están entre 0 y 1. Para cualquier proposición a,

$$0 \leqslant P(a) \leqslant 1 \tag{6.6}$$

2. Las proposiciones necesariamente ciertas (es decir, válidas) tienen probabilidad 1, y las proposiciones necesariamente falsas (es decir, insatisfacibles) tienen probabilidad 0.

$$P(\mathrm{cierto}) = 1 \qquad P(\mathrm{falso}) = 0 \tag{6.7}$$

3. La probabilidad de una disyunción viene dada por

$$P(a \vee b) = P(a) + P(b) - P(a \wedge b) \tag{6.8}$$

Este último axioma es fácilmente interpretable utilizando teoría de conjuntos Figura 6.4, pues utilizando la Ecuación 6.1 se tiene:

$$P(a \vee b) = \frac{|a \cup b|}{|S|} = \frac{|a| + |b| - |a \cap b|}{|S|} = P(a) + P(b) - P(a \wedge b).$$

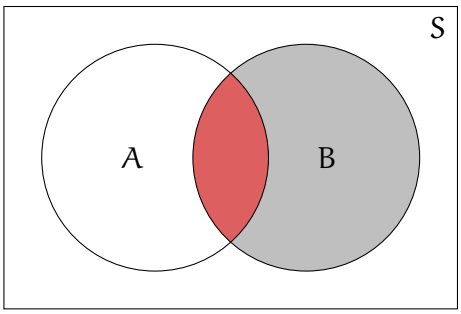


Figura 6.5 La probabilidad $P(a|b)$ corresponde a la probabilidad de la intersección de ambos eventos, como si B fuera el nuevo universo.

Probabilidad condicional

Es posible definir la probabilidad de un evento sujeto a la evidencia observada como (Figura 6.5):

$$P(a|b) = \frac{P(a \wedge b)}{P(b)} \tag{6.9}$$

Teorema de Bayes

Si tomamos las definiciones de probabilidad condicional:

$$P(a|b) = \frac{P(a \wedge b)}{P(b)} \qquad P(b|a) = \frac{P(a \wedge b)}{P(a)} \tag{6.10}$$

y despejamos la probabilidad conjunta en ambas expresiones:

$$P(a \wedge b) = P(b)P(a|b) = P(a)P(b|a) \tag{6.11}$$

obtenemos la fórmula del *teorema de Bayes*:

$$P(b|a) = \frac{P(a|b)P(b)}{P(a)} \tag{6.12}$$

A priori, a posteriori y verosimilitud

Asociada al teorema de Bayes existe una terminología particuar. La misma fórmula puede ser reescrita como sigue:

$$p(\theta|e) = \frac{p(e|\theta)p(\theta)}{p(e)} \tag{6.13}$$

[3] Russell y Norving 2004

donde a $p(\theta)$ se le conoce como la *creencia a priori* [4] de que ocurra θ. Esta variable puede representar cualquier cosa, por ejemplo una enfermedad como la varicela o el sarampeon. La característica especial de esta variable es que suele corresponder a algo que no podemos saber directamente si ocurrió o no. Lo único que podemos hacer es inferir la probabilidad de que haya ocurrido, dependiendo de la presencia una cierta evidencia e que sí podemos medir (por ejemplo la presencia de manchas rojas en la piel). Usualmente para obtener un diagnóstico, necesitaríamos la probabilidad $p(\theta|e)$ de que haya ocurrido θ dependiendo del valor de la evidencia e; como esta probabilidad se obtiene posteriormente a la presentación de la evidencia, se le suele conocer como *a posteriori* [5].

Desafortunadamente, es más fácil modelar $p(e|\theta)$, la probabilidad de que se presente la evidencia, dada su causa. A esta probabilidad se le conoce como *verosimilitud*[6].

A $p(e)$ no quisieron darle nombre porque representa a la probabilidad de que la evidencia se presente, en general y hay quienes la concideran *sólo una constante de normalización*, aunque no por ello deja de ser relevante. A menudo es frecuente encontrar la fórmula expresada como:

$$p(\theta|e) \propto p(e|\theta)p(\theta) \tag{6.14}$$

Distribuciones de probabilidad

Una *distribución de probabilidad* asocia a cada valor posible de una variable aleatoria la probabilidad de que se obtenga ese valor al realizar un experimento. Por ejemplo:

Lluvia	$\mathrm{P(Lluvia)}$
0	0.6375
1	0.3625
	$\sum = 1$

Una *distribución de probabilidad conjunta* asocia a cada combinación posible de los valores de varias variables que se revisan simulatáneamente, la probabilidad de que se obtenga esa combinación al realizar un experimento. Por ejemplo:

[4] *Prior* en inglés
[5] *Posterior* en inglés.
[6] *Likelihood* en inglés.

Estación	Lluvia	P(Lluvia ∧ Estación)
Primavera	0	0.1875
Primavera	1	0.0625
Verano	0	0.075
Verano	1	0.175
Otoño	0	0.175
Otoño	1	0.075
Invierno	0	0.2
Invierno	1	0.05
		$\sum = 1$

De esta tabla se pueden leer directamente respuestras a preguntas como "¿Cuál es la probabilidad de que sea primavera y llueva?", donde la respuesta sería 0.0625.

Si las variables siendo consideradas, son todas las variables del sistema, entonces se dice que se tiene la *distribución de probabilidad conjunta completa*.

La *distribución de probabilidad condicional* asocia una probabilidad a cada combinación de valores de variables, dado un valor determinado para un conjunto de variables evidencia. Por ejemplo: ¿cuál es la probabilidad de que llueva, sabiendo la estación del año? Un ejemplo concreto extraído de esta tabla sería "¿Cuál es la probabilidad de que llueva, si es primavera?" y la respuesta es 0.25.

Lluvia / Estación	Primavera	Verano	Otoño	Invierno
0	0.75	0.30	0.70	0.80
1	0.25	0.70	0.30	0.20
	$\sum = 1$	$\sum = 1$	$\sum = 1$	$\sum = 1$

La regla de la cadena

Así como se definió a la probabilidad condicional en términos de una probabilidad conjunta y la probabilidad de la variable evidencia en Ecuación 6.9, es posible generalizar la definición a varias variables. Para simplificar la notación a partir de este momento, en lugar de utilizar el tradicional símbolo $\wedge$ para denotar la relación y o $\mathrm{intersección}$, indicaremos a la probabilidad conjunta separando a las variables con comas. De este modo la ecuación citada se convierte en:

$$P(a|b) = \frac{P(a \wedge b)}{P(b)} = \frac{P(a, b)}{P(b)} \tag{6.15}$$

de donde habíamos despejado:

$$P(a, b) = P(a|b)P(b) \tag{6.16}$$

Cuando queremos expresar a la probabilidad de obtener la conjunción de varias variables, dada la conjunción de otra variables, la fórmula se verá como:

$$P(X_1, X_2, ..., X_n | Y_1, Y_2, ..., Y_n) = \frac{P(X_1, X_2, ..., X_n, Y_1, Y_2, ..., Y_n)}{P(Y_1, Y_2, ..., Y_n)} \tag{6.17}$$

de donde despejamos:

$$P(X_1, X_2, ..., X_n, Y_1, Y_2, ..., Y_n) = P(X_1, X_2, ..., X_n | Y_1, Y_2, ..., Y_n) P(Y_1, Y_2, ..., Y_n) \tag{6.18}$$

En particular, si comenzamos por condicionar a la primera variable con respecto a las demás en la conjunción, se verá así:

$$P(X_1, X_2, ..., X_n) = P(X_1 | X_2, ..., X_n) P(X_2, ..., X_n) \tag{6.19}$$

repitiendo el mismo proceso con los términos como $P(X_2, ..., X_n)$ obtenemos la fórmula conocida como *regla de la cadena*:

$$P(X_1, X_2, ..., X_n) = P(X_1 | X_2, ..., X_n) P(X_2 | X_3, ..., X_n) ... P(X_n) \tag{6.20}$$

Compuertas lógicas con probabilidades

Para iniciar con el uso de probabilidades para inferencia lógica, comenzaremos por ilustrar cómo la inferencia exacta puede ser vista como un caso particular de la inferencia con probabilidades, mediante la definición de la compuertas lógicas or y and en términos de probabilidades. Se dejará como un ejercicio para el lector definir a la compuerta or.

Para esto se definirán distribuciones de probabilidad condicional: la salida de la compuerta lógica, condicionada al valor de las entradas. Asignando probabilidad uno a las combinaciones correctas de variables de entrada/salida y cero a las opciones inválidas.

No: $\neg$

x	y	$P(y\|x)$
0	0	0
0	1	1
1	0	1
1	1	0

Y: $\wedge$

x_1	x_2	y	$P(y\|x_1, x_2)$
0	0	0	1
0	0	1	0
0	1	0	1
0	1	1	0
1	0	0	1
1	0	1	0
1	1	0	0
1	1	1	1

Ganamos la posibilidad de modelar compuertas ruidosas, que no siempre devuelven la respuesta correcta, asociando una pequeña probabilidad $\varepsilon \neq 0$ a las respuestas incorrectas y ajustando acordemente las correctas:

No: $\neg$		
x	y	$P(y\|x)$
0	0	ε
0	1	$1-\varepsilon$
1	0	$1-\varepsilon$
1	1	ε

Y: $\wedge$			
x_1	x_2	y	$P(y\|x_1, x_2)$
0	0	0	$1-\varepsilon$
0	0	1	ε
0	1	0	$1-\varepsilon$
0	1	1	ε
1	0	0	$1-\varepsilon$
1	0	1	ε
1	1	0	ε
1	1	1	$1-\varepsilon$

7 | Factores

Definición 7.1: Factor

Un factor $\phi(X_1, ..., X_k)$, sobre variables X_i, donde cada X_i puede tomar valores de un dominio D_{X_i} es una función:

$$\phi : Val(X_1, ..., X_k) \to \mathbb{R} \tag{7.1}$$

que asocia un número real a posibles asignaciones de valores a las variables $X_1, ..., X_k$. El *alcance* $\mathbb{A}$ de un factor son las variables cuyos posibles valores están siendo considerados.

$$\mathbb{A}(\phi(X_1, ..., X_k)) = \{X_1, ..., X_k\} \tag{7.2}$$

Ejemplo 7.1. *Un factor. Sean las variables y dominios:*

- Estación, $D_{\text{Estación}} = \{\text{Primavera}, \text{Verano}, \text{Otoño}, \text{Invierno}\}$
- Lluvia, $D_{\text{Lluvia}} = \{0, 1\}$

de modo que el alcance es $\mathbb{A} = \{\text{Estación}, \text{Lluvia}\}$

Estación	Lluvia	*Frecuencia (días/mes)*
Primavera	*0*	*18*
Primavera	*1*	*6*
Verano	*0*	*7*
Verano	*1*	*17*
Otoño	*0*	*17*
Otoño	*1*	*7*
Invierno	*0*	*20*
Invierno	*1*	*5*

Se utilizarán factores para representar las distribuciones de probabilidad, distribuciones de probabilidad conjuntas y condicionales.

Marginalización

Definición 7.2: Marginalización

Dada una variable X_i, en el alcance $\mathbb{A}$ de un factor $\phi = \phi(X_1, ..., X_k)$, que se desea marginalizar, se define a la operación como:

$$\text{marginalización}(\phi(X_1, ..., X_k), X_i) = \phi'(\mathbb{A}') \tag{7.3}$$

$$\mathbb{A}' = \mathbb{A}(\phi) - X_i \text{ con } i \in [1, k] \tag{7.4}$$

$$\phi'(x_1, ..., x_{i-1}, x_{i+1}, ..., x_k) = \sum_{Val\{X_i\}} \phi(x_1, ..., x_k) \tag{7.5}$$

donde x_i es un valor particular de X_i, $\phi'(x_1, ..., x_{i-1}, x_{i+1}, ..., x_k)$ es el renglón del factor ϕ' con $\{X_1 = x_i, ..., X_k = x_k\}$ y $Val\{X_i\}$ es el conjunto de valores posibles asignables a X_i.

Ejemplo 7.2. *Marginalizar* Estación

Estación	Lluvia	P(Lluvia, Estación)
Primavera	*0*	*0.1875*
Primavera	*1*	*0.0625*
Verano	*0*	*0.075*
Verano	*1*	*0.175*
Otoño	*0*	*0.175*
Otoño	*1*	*0.075*
Invierno	*0*	*0.2*
Invierno	*1*	*0.05*
		$\sum = 1$

$\Rightarrow$

Lluvia	P(Lluvia)
0	*0.63750*
1	*0.36250*
	$\sum = 1$

Reducción

Definición 7.3: Reducción

Dado un valor $x_i = a$ para una de las variables X_i en el alcance $\mathbb{A}$ del factor ϕ, se reduce el factor eliminando todas aquellas entradas en las cuales no se cumple que $X_i = a$. La operación se define como:

$$\text{reducción}(\phi(X_1, ..., X_k), X_i, a) = \phi'(\mathbb{A}') \tag{7.6}$$

$$\mathbb{A}' = \mathbb{A}(\phi) - X_i \text{ con } i \in [1, k] \tag{7.7}$$

$$\phi'(x_1, ..., x_{i-1}, x_{i+1}, ..., x_k) = \phi(x_1, ..., x_k) \text{ con } X_i = a \tag{7.8}$$

A diferencia de las distribuciones de probabilidad, la reducción en un factor no requiere renormalizar sus valores asociados, pues por definición, no es necesario que éstos sumen uno.

Ejemplo 7.3. *Reducir* $\text{Estación} = \text{Primavera}$

Estación	Lluvia	P(Lluvia, Estación)
Primavera	*0*	*0.1875*
Primavera	*1*	*0.0625*
Verano	*0*	*0.075*
Verano	*1*	*0.175*
Otoño	*0*	*0.175*
Otoño	*1*	*0.075*
Invierno	*0*	*0.2*
Invierno	*1*	*0.05*
		$\sum = 1$

$\Rightarrow$

Lluvia	P(Lluvia, Estación = primavera)
0	*0.1875*
1	*0.0625*
	$\sum = 0.25$

Normalización

Definición 7.4: Normalización

Dado un factor ϕ con n renglones sea

$$s = \sum_{i=1}^{n} \phi_i \tag{7.9}$$

con ϕ_i el valor asociado al renglón i, s es la suma de los valores de todos los renglones. Entonces:

$$\text{normalización}(\phi) = \phi' \tag{7.10}$$

$$\phi'_i = \frac{\phi_i}{s} \tag{7.11}$$

donde el valor en cada renglón de ϕ ha sido dividido entre s.

Ejemplo 7.4. *Normalizar*

Lluvia	P(Lluvia, primavera)
0	*0.1875 / 0.25*
1	*0.0625 / 0.25*
	$\sum = 0.25$

$\Rightarrow$

Lluvia	P(Lluvia|primavera)
0	*0.75*
1	*0.25*
	$\sum = 1$

Multiplicación

Definición 7.5: Multiplicación

Se define el producto de factores de tal modo que:

$$\phi_1 = \phi_1(X_1, ..., X_k) \tag{7.12}$$
$$\phi_2 = \phi_2(Y_1, ..., Y_l) \tag{7.13}$$
$$\phi = \phi_1 \phi_2 \tag{7.14}$$
$$\mathbb{A}(\phi) = \mathbb{A}(\phi_1) \cup \mathbb{A}(\phi_2) \tag{7.15}$$
$$\phi(z_1, ..., z_m) = \phi_1(x_1, ..., x_k) * \phi_2(y_1, ..., y_l) \tag{7.16}$$

donde $Z_k \in \mathbb{A}(\phi)$ por lo que:

$$\text{si } Z_k \in \mathbb{A}(\phi_1) \Rightarrow Z_k = X_i \wedge z_k = x_i \tag{7.17}$$
$$\text{si } Z_k \in \mathbb{A}(\phi_2) \Rightarrow Z_k = Y_j \wedge z_k = y_j \tag{7.18}$$

obsérvese que si $Z_k \in \mathbb{A}(\phi_1)$ y $Z_k \in \mathbb{A}(\phi_2)$ entonces $Z_k = X_i = Y_j$ y en cada renglón deberá cumplirse $z_k = x_i = y_j$.

Ejemplo 7.5. *Multiplicar* P(Lluvia|Estación)P(Estación)

Estación	Lluvia	P(Lluvia|Estación)
Primavera	*0*	*0.75*
Primavera	*1*	*0.25*
Verano	*0*	*0.30*
Verano	*1*	*0.70*
Otoño	*0*	*0.70*
Otoño	*1*	*0.30*
Invierno	*0*	*0.80*
Invierno	*1*	*0.20*

$\times$

Estación	P(Estación)
Primavera	*0.25*
Verano	*0.25*
Otoño	*0.25*
Invierno	*0.25*
	$\sum = 1$

$=$

Estación	Lluvia	P(Lluvia, Estación)
Primavera	*0*	*0.1875*
Primavera	*1*	*0.0625*
Verano	*0*	*0.075*
Verano	*1*	*0.175*
Otoño	*0*	*0.175*
Otoño	*1*	*0.075*
Invierno	*0*	*0.2*
Invierno	*1*	*0.05*
		$\sum = 1$

Ejemplo 7.6. *Multiplicar* P(Lluvia)P(Estación)

Lluvia	P(Lluvia)
0	*0.63750*
1	*0.36250*
	$\sum = 1$

$\times$

Estación	P(Estación)
Primavera	*0.25*
Verano	*0.25*
Otoño	*0.25*
Invierno	*0.25*
	$\sum = 1$

$=$

Sea R = P(Lluvia)P(Estación)

Estación	Lluvia	R
Primavera	*0*	*0.159375*
Primavera	*1*	*0.090625*
Verano	*0*	*0.159375*
Verano	*1*	*0.090625*
Otoño	*0*	*0.159375*
Otoño	*1*	*0.090625*
Invierno	*0*	*0.159375*
Invierno	*1*	*0.090625*
		$\sum = 1$

$=$

Estación	Lluvia	R
Primavera	*0*	*0.159375*
Verano	*0*	*0.159375*
Otoño	*0*	*0.159375*
Invierno	*0*	*0.159375*
Primavera	*1*	*0.090625*
Verano	*1*	*0.090625*
Otoño	*1*	*0.090625*
Invierno	*1*	*0.090625*
		$\sum = 1$

Distribuciones de probabilidad con factores

Claramente, los factores pueden contener, en particular, tablas de distribuciones de probabilidad sobre variables discretas. Las variables aleatorias del sistema aparecerán en el alcance del factor, las combinaciones posibles de asignaciones a estas variables quedarán registradas en los renglones del factor y la probabilidad de que ocurran será el número real asociado a esa entrada.

Distribuciones de probabilidad condicional con factores

Para escribir la distribución de probabilidad condicional utilizando factores, la notación cambia un poco. Por ejemplo:

Estación	Lluvia	P(Lluvia|Estación)	
Primavera	0	0.75	$\sum = 1$
Primavera	1	0.25	
Verano	0	0.30	$\sum = 1$
Verano	1	0.60	
Otoño	0	0.70	$\sum = 1$
Otoño	1	0.30	
Invierno	0	0.8	$\sum = 1$
Invierno	1	0.1	

8 | Dependencia e independencia probabilística

Independencia

Dos eventos α y β son independientes $P \models \alpha \perp \beta$[1] si:

$$P(\alpha \cap \beta) = P(\alpha)P(\beta) \tag{8.1}$$
$$P(\alpha|\beta) = P(\alpha) \tag{8.2}$$
$$P(\beta|\alpha) = P(\beta) \tag{8.3}$$

Se dice que dos variables aleatorias A y B son *independientes* $P \models A \perp B$ si se cumple que:

$$P(A, B) = P(A)P(B) \tag{8.4}$$
$$P(A|B) = P(A) \tag{8.5}$$
$$P(B|A) = P(B) \tag{8.6}$$

No hay forma de representar el concepto de independencia con diagramas de Venn. Obsérvese que independencia y exclusión ($A \cap B = \emptyset$) son dos conceptos distintos.

Independencia condicional

Se dice que dos eventos/variables aleatorias E_1 y E_2 son *condicionalmente independientes* dado F, $P \models (E_1 \perp E_2|F)$ si, dado que F ocurre, la probabilidad condicional de que E_1 ocurra no cambia al obtenerse información sobre si E_2 ocurre o no. Esto se escribe:

[1] Se lee: P satisface que α es independiente de β.

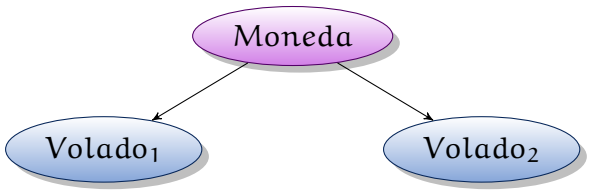


Figura 8.1 Un ejemplo de independencia condicional. El resultado de cada volado con una moneda depende de si se eligió una moneda normal o una moneda cargada. Si se sabe qué moneda se eligió, conocer el resultado de un volado ya no dice nada nuevo sobre el resultado de otro volado con la misma moneda.

$$P(E_1|E_2F) = P(E_1|F) \tag{8.7}$$
$$P(E_2|E_1F) = P(E_2|F) \tag{8.8}$$
$$\tag{8.9}$$

o equivalentemente:

$$P(E_1E_2|F) = P(E_1|F)P(E_2|F) \tag{8.10}$$

Ejemplo: Se tienen dos monedas, una cargada de modo que 90 % del tiempo el resultado de un volado con ella será sol y otra moneda normal $Moneda =< Cargada, Normal >$. Se elige una moneda al azar y se lanzan dos volados $Volado_i =< Sol, Águila >$ [Figura 8.1]. Inicialmente conocer el resultado del primer volado brinda información sobre la posibilidad de haber elegido la moneda cargada y por ende acerda del posible resultado del siguiente volado. Sin embargo, si se sabe qué moneda se eligió, esa información es suficiente para conocer la probabilidad de cada resultado en el segundo volado, la información del primer volado ya no es relevante. En este caso se dice que el resultado del segundo volado es independiente del resultado del primero, dado que se sabe qué moneda fue elegida.

Redes Bayesianas

Definición 8.1: Red Bayesiana

Una *Red Bayesiana* es:

- Una gráfica acíclica dirigida (GDA) G cuyos nodos representan a las variables aleatorias $X_1, ..., X_n$.

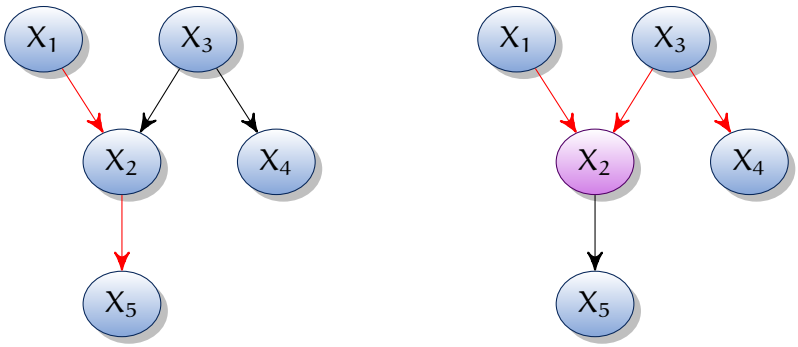


Figura 8.2 Izquierda: Ruta activa $X_1 - X_2 - X_5$. Derecha: Ruta $X_1 - X_2 - X_3 - X_4$ con estructura-v entre $X_1 - X_2 - X_3$, activada porque X_2 fue observada.

- Para cada nodo X_i define una distribución de probabilidad condicional

$$P(X_i|Padres_G(X_i)) \tag{8.11}$$

La red de Bayes representa un distribución de probabilidad conjunta para la cual se cumple, debido a la relación de independencia condicional entre las variables, que:

$$P(X_1, ..., X_n) = \prod_i P(X_i|Padres_G(X_i)) \tag{8.12}$$

A esta fórmula se le conoce como regla de la cadena para Redes Bayesianas. Cuando un distribución de probabilidad conjunta P satisface esta relación, se dice que P *se factoriza sobre* G.

Independencia en redes bayesianas

Definición 8.2: Ruta

Una *ruta* $X_1 - ... - X_k$ es una secuencia de nodos que se encuentran conectados entre sí mediante una sola arista (no dirigida) en la gráfica.

Ruta activa

Sea $\mathbb{Z}$ el conjunto de variables evidencia cuyo valor ha sido observado.

Definición 8.3

- Una ruta se encuentra *activa* si no tiene *estructuras-v* (dos padres X_{i-1}, X_{i+1}

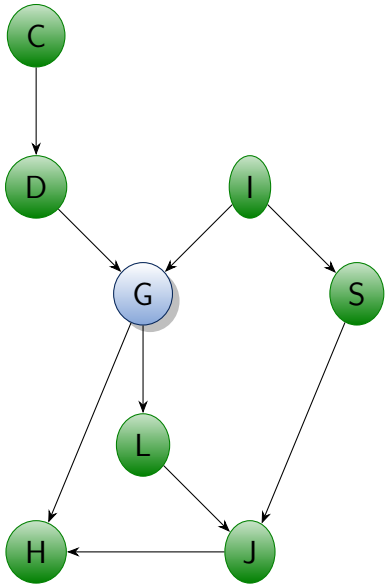


Figura 8.3 Ejemplo de gráfica de Bayes con G observado.

de un nodo común X_i, $X_{i-1} \to X_i \leftarrow X_{i+1}$) [Figura 8.2 (Izquierda)].

- Una ruta $X_1 - \ldots - X_k$ está *activa dado* $\mathbb{Z}$ si:
  - ⋆ Para cualquier *estructura-v* $X_{i-1} \to X_i \leftarrow X_{i+1}$ se tiene que X_i o alguno de sus descencientes están en $\mathbb{Z}$.
  - ⋆ Ningún otro X_i está en $\mathbb{Z}$.

Ejercicio 8.1. *Dada la red Bayesiana en la Figura 8.3, determinar si los senderos siguientes son activos, dado que la variable* G *fue observada:*[2]

1. $C \to D \to G \leftarrow I \to S$ ✓
2. $I \to G \to L \to J \to H$
3. $I \to S \to J \to H$ ✓
4. $C \to D \to G \leftarrow I \to S \to J \leftarrow L$

D-separación

[2] Ejericio tomado del curso en línea de Daphne Koller.

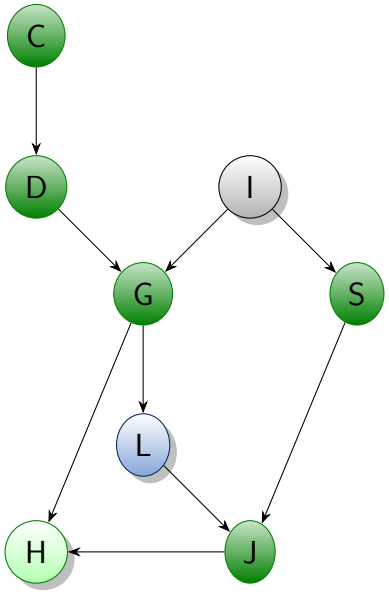


Figura 8.4 Ejemplo de gráfica de Bayes con I, L y H observados. I bloquea el flujo, H lo habilita y L varía su función dependiendo del sendero analizado.

Definición 8.4

Dos variables aleatorias X y Y se encuentran *d-separadas* en G dado $\mathbb{Z}$ si no existe una ruta activa entre X y Y dado $\mathbb{Z}$.

$$\mathrm{d}-\mathrm{sep}_{\mathrm{G}}(\mathrm{X},\mathrm{Y}|\mathrm{Z}) \tag{8.13}$$

Teorema 8.1. *Si la distribución de probabilidad* P *se factoriza sobre* G *y* $\mathrm{d}-\mathrm{sep}_{\mathrm{G}}(\mathrm{X},\mathrm{Y}|\mathrm{Z})$ *entonces* P *satisface* $(\mathrm{X} \perp \mathrm{Y}|\mathrm{Z})$.

Cualquier nodo se encuentra $\mathrm{d}-\mathrm{separado}$ de sus no descendientes dados sus padres $\Rightarrow$ Si P se factoriza sobre G, entonces en P cualquier variable es independiente de sus no-descendientes dados sus padres.

Ejemplo 8.1. *Indique si existe una d-separación en los casos siguientes:*

1. $\mathrm{d}-\mathrm{sep}(\mathrm{D},\mathrm{I}|\mathrm{L})$
2. $\mathrm{d}-\mathrm{sep}(\mathrm{D},\mathrm{J}|\mathrm{L})$
3. $\mathrm{d}-\mathrm{sep}(\mathrm{D},\mathrm{J}|\mathrm{L},\mathrm{I})$ ✓
4. $\mathrm{d}-\mathrm{sep}(\mathrm{D},\mathrm{J}|\mathrm{L},\mathrm{H},\mathrm{I})$

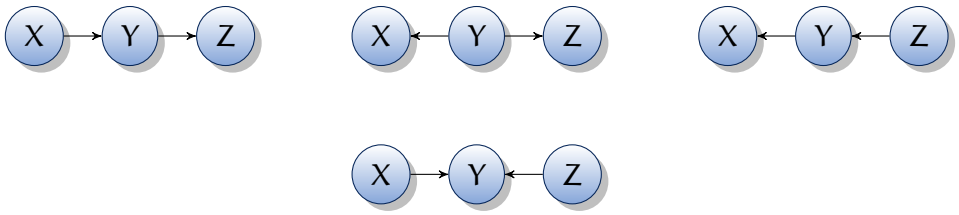


Figura 8.5 Arriba: Tres gráficas $I-equivalentes$. Abajo: Gráfica no $I-Equivalente$ a las anteriores.

Mapas de independencias (Mapas-I)

Si la distribución de probabilidades conjuntas P satisface todas las relaciones de independencia I implicadas por las $d-separaciones$ de una gráfica G se dice que G es un mapa de independencias de P.

$$I(G) = \{(X \perp Y|Z) : d-sep_G(X, Y|Z)\} \tag{8.14}$$

Los mapas de independencias para P no son únicos y no necesariamente contienen todas las relaciones de independencia que son válidas en P. Si G es un mapa de independencias para P, entonces P se factoriza sobre G, para demostrar esto se utiliza la regla de la cadena y el hecho de que cualquier nodo es independiente de sus no descendientes dados su padres.

I-Equivalencia

Dos gráficas G_1 y G_2 sobre $X_1, ..., X_n$ son $I-Equivalentes$ si los conjuntos de independencias derivadas de ambas gráficas son iguales Figura 8.5.

$$I(G_1) = I(G_2) \tag{8.15}$$

Inferencia

Eliminación de variables

Modelo Ingenuo de Bayes

Es un modelo particular utilizado para problemas de clasificación, que asume que todas las características C_i de una clase $Clase$ son independientes entre sí, dada información sobre la clase, esto es $\forall C_i, C_j(C_i \perp C_j|Clase)$ [Figura 8.6], de tal modo que la factorización siguiente es válida:

$$P(Clase, C_1, ..., C_n) = P(Clase)\prod_{i=1}^{n} P(C_i|Clase) \tag{8.16}$$

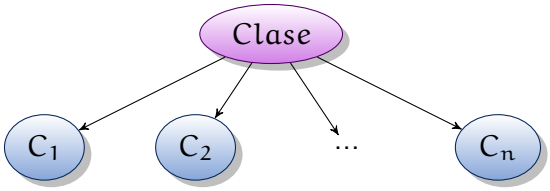


Figura 8.6 En modelo Ingenuo de Bayes, las características de una clase son independientes, dada la clase.

Inferencia aproximada

Muestreo directo

Verosimilitud ponderada

Muestreo Montecarlo en cadenas de Makov (MCCM)

Redes Bayesianas Dinámicas

Sección incompleta.

Sea X una variable aleatoria cuyo valor cambia con el tiempo. La probabilidad de una secuencia de eventos que ocurren a tiempos discretos $t = \{0, ..., T\}$ se expresa:

$$P(X^{(0:T)}) = P(X^{(0)}) \prod_{t=0}^{T-1} P(X^{(t+1)}|X^{(0:t)})$$

utilizando la regla de la cadena.

La *hipótesis de Márkov* asume que:

$$(X^{(t+1)} \perp X^{(0:t-1)}|X^{(t)}) \tag{8.17}$$

entonces se cumple que:

$$P(X^{(0:T)}) = P(X^{(0)}) \prod_{t=0}^{T-1} P(X^{(t+1)}|X^{(t)}) \tag{8.18}$$

Otra simplificación que se puede hacer al modelo es asumir *independencia temporal*, a lo cual también se le conoce como asumir que el sistema es *estacionario*. Esto se expresa:

$$P(X^{(t+1)}|X^{(t)}) = P(X'|X) \qquad \forall t \tag{8.19}$$

Es decir, la dinámica del sistema no cambia con el tiempo.

Un sistema puede estar descrito por un conjunto de variables para cada tiempo y las premisas de independecia se aplican sobre ellas. Para aliviar las restricciones del modelo, estas variables pueden codificar información derivada de otras relaciones temporales. Por ejemplo: en un problema de localización de un vehículo, se puede agregar a la velocidad como variable que describe el estado.

Ross 2013; Bolstad 2007

Inferencia en redes de Markov

Aprendizaje en redes Bayesianas

Máxima verosimilitud

Estimación Bayesiana

Parte V
Robótica Inteligente

9 | Cierre: Inteligencia artificial y el cerebro

Bibliografía

Barber, David (2012). *Bayesian Reasoning and Machine Learning*. Cambridge University Press.

Bolstad, William M. (2007). *Introduction to Bayesian Statistics*. Wiley.

Koller, Daphne y Nir Friedman (2009). *Probabilistic Graphical Models, Principles and Techniques*. MIT Press Cambridge.

Ross, Sheldon (2013). *A First Course in Probability*. Ed. por Deirdre Lynch. 9th. University of Southern California: Pearson.

Russell, Stuart y Peter Norving (2004). *Inteligencia Artificial, Un Enforque Moderno*. 2a. Pearson Prentice Hall.

Winston, Patrick Henry (jun. de 1992). *Artificial Inteligence*. Ed. por Addison Wesley. 3rd. U.S.A.