INTELIGENCIA ARTIFICIAL I

TOMA DE DECISIONES

Donde se verá cómo debe tomar decisiones un agente para obtener lo que desea, por lo menos, en promedio.

Dra. Ing. Selva S. RIvera

TOMA DE DECISIONES

La **teoría de la utilidad** es un marco teórico que se utiliza para explicar cómo las personas toman decisiones en situaciones de <u>incertidumbre</u>. Esta teoría se basa en la idea de que las personas asignan valores subjetivos a los diferentes resultados posibles de una decisión y luego eligen la opción que les proporciona el mayor valor esperado.

Por otro lado, las **redes de decisión** son una herramienta que se utiliza para modelar y analizar decisiones complejas. Estas redes se componen de nodos que representan eventos o decisiones, y arcos que representan las relaciones causales entre ellos.

Las redes de decisión se utilizan para identificar los factores clave que influyen en una decisión y <u>para evaluar cómo diferentes opciones pueden afectar el resultado final.</u>

AGENTES QUE TOMAN DECISIONES

- Cómo debe tomar decisiones un agente para **obtener lo que desea**, **por lo menos en promedio**.
- Esta clase de agentes **puede adoptar decisiones racionales basándose en lo que cree y desea**.
- Estos agentes pueden adoptar decisiones en situaciones en las que un agente lógico no tiene forma de decidir debido a la presencia de incertidumbre y/u objetivos contradictorios.

fuente: Inteligencia Artificial. Un enfoque Moderno - Russell y Norvig

Retrato de Daniel Bernoulli

Los matemáticos, en su teoría, valoran el dinero en proporción a la cantidad del mismo; la gente con sentido común, en la práctica, lo valora en proporción a la utilidad que puede obtener de él.

Paradoja de San Petersburgo

Consiste en un juego de apuestas con un valor esperado infinito. En esta situación, la teoría de decisiones recomienda que se admita cualquier apuesta por alta que sea, acción que ninguna persona racional seguiría.

Paradoja de San Petersburgo

El jugador tiene que pagar una apuesta para participar en el juego. A continuación, este realiza lanzamientos sucesivos de una moneda hasta que salga cruz por primera vez. Entonces se detiene el juego, se cuenta el nº de lanzamientos que se ha producido, y el jugador obtiene 2^n monedas.

Si sale cruz la primera vez el jugador gana 2^1 monedas.

Si sale cruz en el cuarto lanzamiento el jugador gana $2^4 = 16$ monedas.

¿Cuánto estarías dispuesto a apostar?

Supongamos un juego basado en el lanzamiento de un dado, donde se ganan:

$20 si sale el 6,

$8 si sale el 5,

y $ - 1 (se paga un peso además de la apuesta inicial) si salen del 1 al 4.

La ganancia esperada es, contando con una probabilidad de 1/6 para cada uno de los resultados posibles:

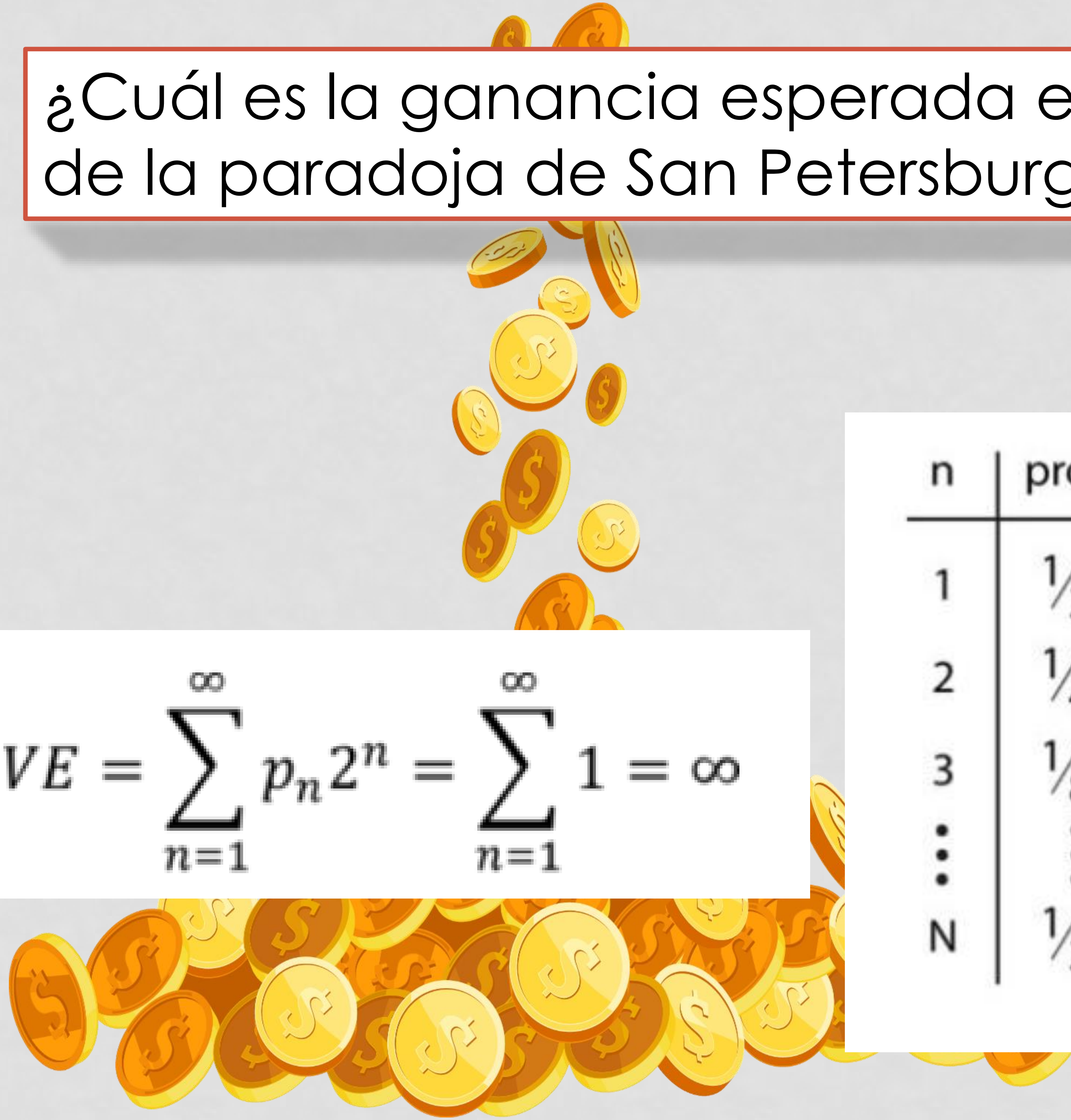
EM = 1/6 • 20 + 1/6 • 8 + 1/6 • -1 + 1/6 • -1 + 1/6 • -1 + 1/6 • -1 = 4

La ganancia esperada es $4

Jugador RACIONAL

Acepta si la **ganancia esperada** (la media del dinero que obtendría participando muchas veces) **es mayor que la suma exigida para entrar al juego.**

¿Cuál es la ganancia esperada en el juego de la paradoja de San Petersburgo?



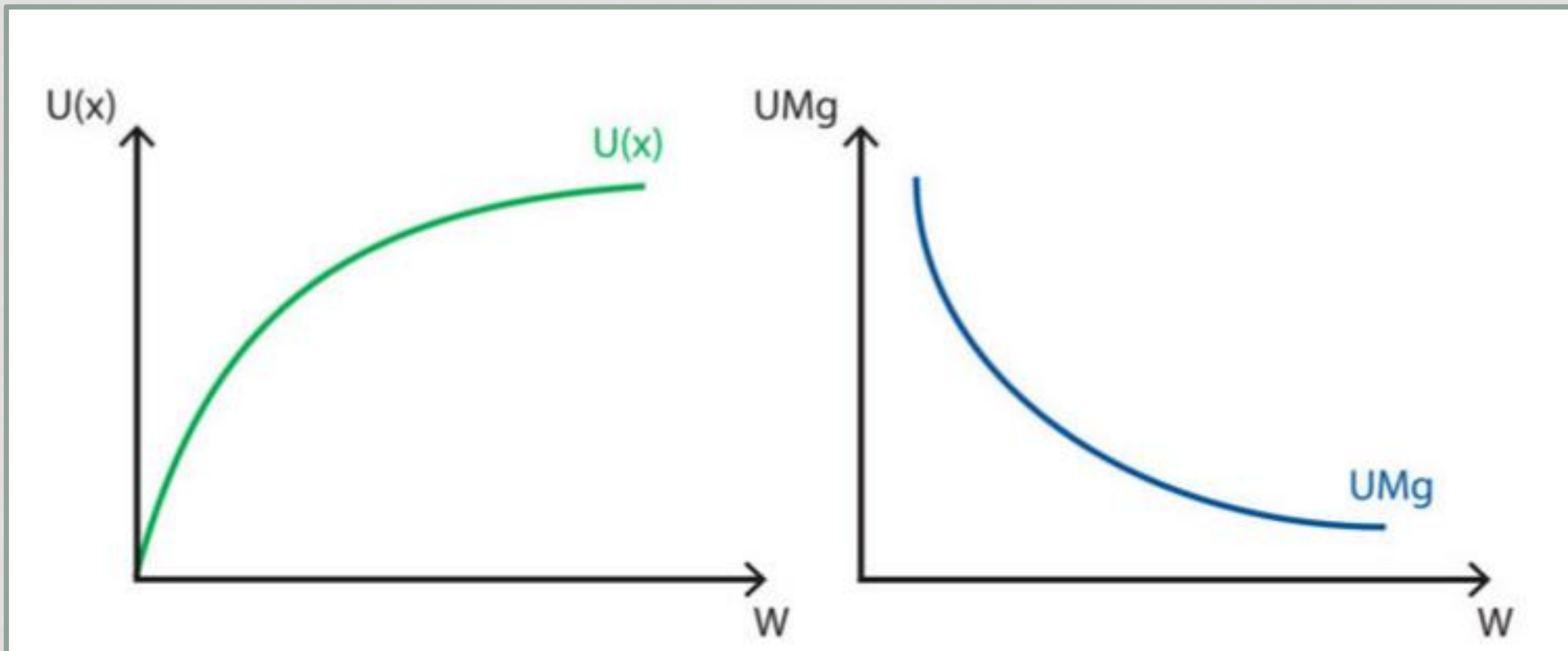
$$VE = \sum_{n=1}^{\infty} p_n 2^n = \sum_{n=1}^{\infty} 1 = \infty$$

n	prob.	pago	VE
1	$\frac{1}{2}$	2	1
2	$\frac{1}{4}$	4	1
3	$\frac{1}{8}$	8	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$
N	$\frac{1}{2}^n$	2^n	1

PARADOJA

- La paradoja surge entonces porque aunque siguiendo las directrices de la teoría de la decisión se debería apostar cualquier suma que nos exigiesen por elevada que parezca (ya que la apuesta será siempre favorable), las personas consideradas razonables no están dispuestas en general a apostar mucho.

La **función de utilidad** ($u(x)$) es el truco que los economistas usan para poder representar matemáticamente las preferencias de los agentes económicos, y en el caso de una persona racional, aunque es siempre creciente, crece de forma **cóncava** (es decir, crece cada vez más despacio). El sentido común apoya esta intuición. El valor "real" de 100 euros para alguien que tiene cero es muchísimo (ya que es una cuestión de supervivencia), pero para alguien que ya tiene un millón de euros, es ínfimo. Dicho de otra forma, la **utilidad marginal** del dinero es decreciente.



Bernoulli afirmaba que las decisiones que toman los individuos no surgen solo de los resultados esperados sino también de otros factores definidos a través de una función de utilidad.

Para Bernoulli, escoger una opción según su valor esperado no era racional porque en ese caso los individuos apostarían todo.

Su argumento era el siguiente: «la ganancia monetaria pueda incrementarse indefinidamente, pero la utilidad de esa riqueza no se incrementa de modo paralelo».

Bernoulli solucionó el problema utilizando la máxima utilidad esperada.

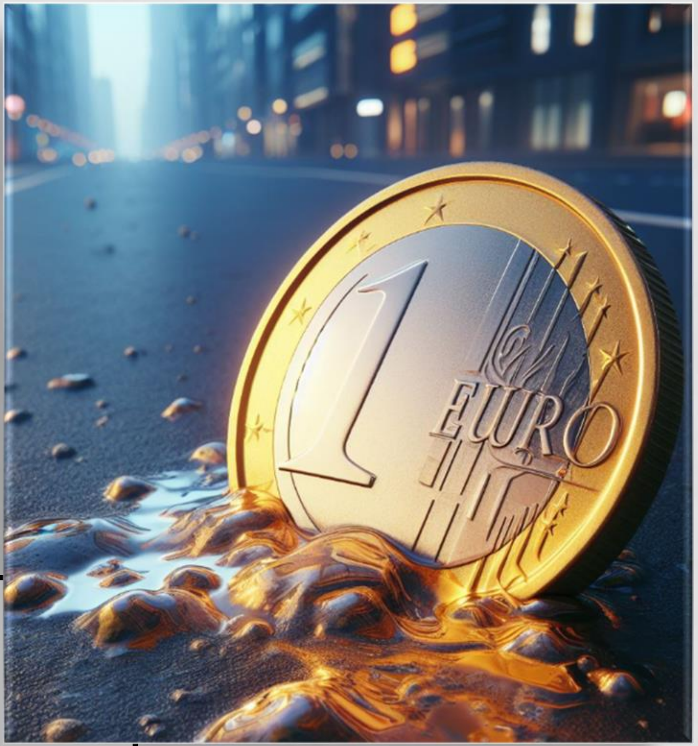
$$EU = p * U(x) + (1 - p) * U(y)$$

La **teoría de la utilidad esperada** es un modelo de elección racional donde los individuos toman decisiones con incertidumbre (resultados inciertos).

Cada resultado posible puede cuantificarse en términos de **útiles**, y representarse a través de la **función de utilidad**.

La elección preferida, según la teoría, será aquella cuya utilidad esperada sea la más alta; es decir, aquella utilidad que, estando ponderada por su probabilidad, sigue siendo mayor que el resto.

Ejemplo:



Dada la siguiente función de utilidad F(x)= √x

y la siguiente lotería:

- *(0.5:0.5) probabilidades de obtener (4;9)*

Valor esperado= *(1/2 x 4) + (1/2 x 9) = 6,5 unidades*

Utilidad Esperada= *(1/2 x √ 4) + (1/2 x √ 9) = 2,5 útiles*

Función de utilidad

- Son una representación numérica de nuestras preferencias
- No asignan un valor numérico a la preferencia propiamente dicha
- Simplemente indican orden y magnitud de preferencia
- Indican qué es lo que mas nos gusta y por cuanto

AGENTES QUE TOMAN DECISIONES

- Las **preferencias** de un agente respecto a ciertos estados del mundo se sustentan mediante una **FUNCIÓN DE UTILIDAD** que asigna una cantidad numérica para expresar lo deseable que es un estado.

- **U(S)** denotará la utilidad del estado **S**.

- Esta medida de la utilidad, combinada con la prob. de ocurrencia de las acciones da como resultado la **utilidad esperada EU(A|E)** (utilidad esperada de la acción A dada la evidencia E).

fuente: Inteligencia Artificial. Un enfoque Moderno - Russell y Norvig

Cálculo de Utilidad Esperada

$$EU(A|E) = \sum_i P(Resultado_i(A).|Realizar(A), E)U(Resultado_i(A))$$

$\boldsymbol{Resultado_i(A)}$: posibles resultados de una acción no determinista A, donde i recorre el rango de posibles resultados.

$P(Resultado_i(A)|Realizar(A), E)$: probabilidad de cada resultado donde el factor E engloba la evidencia disponible por el agente sobre el mundo y $Realizar(A)$ se refiere a la proposición consistente en ejecutar la acción A en el estado actual.

U : utilidad

El principio de la **Máxima utilidad esperada (MUE)** establece que un agente racional debe elegir aquella acción que maximice la utilidad esperada del agente.

fuente: Inteligencia Artificial. Un enfoque Moderno - Russell y Norvig

Se puede demostrar que el principio de MUE puede derivarse de las siguientes restricciones que expresan las preferencias de un agente sobre dos estados A y B:

- $A > B$ Se prefiere a A sobre B
- $A \sim B$ El agente es indiferente
- $A \gtrsim B$ El agente prefiere a A sobre B o ambos le son indiferentes

Acciones deterministas	Acciones no determinista
A y B son los estados resultantes de tales acciones.	A y B son loterías (experimento aleatorio). $L = [p_1, C_1; p_2, C_2; \ldots ; p_n, C_n]$

Una lotería con un único resultado podría escribirse como A o mediante el suceso seguro [1,A].

fuente: Inteligencia Artificial. Un enfoque Moderno - Russell y Norvig

AXIOMAS DE LA TEORÍA DE LA UTILIDAD

1) ORDENACIÓN $(A > B) \vee (B > A) \vee (A \sim B)$

Dados dos estados cualesquiera, un agente racional debe ser capaz de, o bien preferir uno de ellos, o bien establecer ambos igualmente preferibles.

Es decir, el agente no puede evitar tomar una decisión.

2) TRANSITIVIDAD $(A > B) \wedge (B > C) \rightarrow (A > C)$

Dados tres estados cualesquiera, si un agente prefiere A frente a B y a B sobre C, entonces el agente debe preferir A sobre C.

fuente: Inteligencia Artificial. Un enfoque Moderno - Russell y Norvig

AXIOMAS DE LA TEORÍA DE LA UTILIDAD

3) CONTINUIDAD A > B > C → $\exists\, p\ [p, A; 1-p, C] \sim B$

Si existe algún estado B entre A y C en la relación de preferencia, entonces existe un valor de probabilidad *p* para el que el agente racional le es indiferente entre considerar como seguro el estado B y la lotería que establece el estado A con probabilidad *p* y al estado C con probabilidad 1-*p*.

***Ejemplo de continuidad:** Supongamos una lotería A donde ganamos 10$ seguro, una lotería B donde no recibimos nada seguro y una lotería C donde vas a la cárcel seguro. A es preferido a B y B es preferido a C, pero esto significa que existe una probabilidad $p \in (0, 1)$ tal que estaría indiferente entre no recibir nada seguro (B) y una lotería compuesta L con p probabilidades de ganar 10$ y 1-p probabilidad de ir a la cárcel.*

fuente: Inteligencia Artificial. Un enfoque Moderno - Russell y Norvig

AXIOMAS DE LA TEORÍA DE LA UTILIDAD

4) SUSTITUCIÓN $A \sim B \rightarrow [p, A; 1\text{-}p, C] \sim [p, B; 1\text{-}p, C]$

Si un agente es indiferente entre dos loterías, A y B, entonces el agente es indiferente entre dos loterías más complejas en las que aparecen A y B. Esto es así, sin importar las probabilidades involucradas, ni el resto de resultados posibles en las loterías.

Ejemplo de sustitución: *Supongamos una lotería A donde ganás $ 10 seguro, una lotería B donde no recibes nada seguro y dos loterías en las que con cierta probabilidad ganás $1.000.000 (C) o $10 (A) o ganás $1.000.000 (C) o nada (B).*

Si un agente es indiferente entre A y B, entonces el agente es indiferente

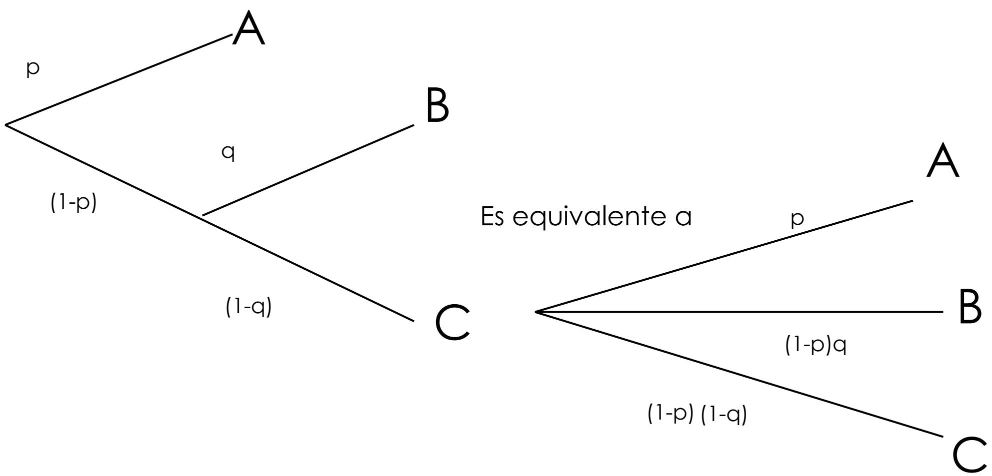
fuente: Inteligencia Artificial. Un enfoque Moderno - Russell y Norvig

AXIOMAS DE LA TEORÍA DE LA UTILIDAD

5) MONOTONICIDAD $A > B \rightarrow (p \geq q \leftrightarrow [p, A; 1-p, B] \gtrsim [q, A; 1-q, B])$

Se supone que existen dos loterías que tienen dos resultados posibles e iguales, A y B.
Si un agente prefiere una lotería A frente a la B, entonces el agente debe preferir la lotería que presenta una mayor probabilidad para A (y viceversa).

6) DESCOMPOSICIÓN $\left[p, A; 1-p, [q, B; 1-q, C]\right] \sim [p, A; (1-p)q, B; (1-p)(1-q), C]$



Las loterías compuestas pueden reducirse a una forma más simple empleando las propiedades de la teoría de la probabilidad.

Los axiomas de la teoría de la utilidad no dicen nada acerca de la utilidad, simplemente se refieren a las relaciones de preferencia.

La preferencia se asume que es una propiedad básica de los agentes racionales.

La existencia de una función de utilidad se deriva de los axiomas de la utilidad:

Principio de Utilidad

Si las preferencias de un agente obedecen a los axiomas de la utilidad entonces existe una función real U asociada a cada estado de forma que:

U(A) > U(B) ↔ A > B

U(A) = U(B) ↔ A ~ B

Principio de MUE

La utilidad de una lotería es la suma de la probabilidad de cada resultado multiplicada por la utilidad de cada resultado:

U(p1, S1; … ; pn, Sn]) = $\sum_i p_i \, U(S_i)$

fuente: Inteligencia Artificial. Un enfoque Moderno - Russell y Norvig

LA UTILIDAD DEL DINERO

- **Suponga que es el ganador de un concurso y se le plantea: o bien acepta directamente un premio de $1.000.000 o bien se juega el premio a cara o cruz, de forma que si sale cara, lo pierde todo y si sale cruz gana $3.000.000.**

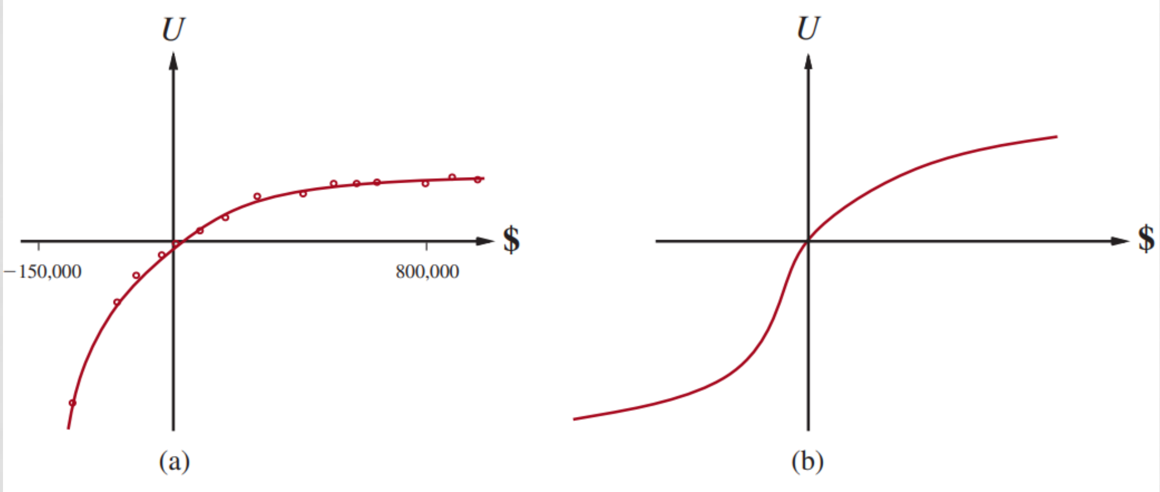
- VALOR MONETARIO ESPERADO (VME) de la apuesta es ½ ($0) + ½ ($3,000,000) = $1,500,000

- VME de aceptar el dinero es $1,000,000

  El valor que el agente estaría dispuesto a aceptar (en lugar de la lotería) se conoce como **equivalente de certeza** o **seguridad de la lotería**.

fuente: Inteligencia Artificial. Un enfoque Moderno - Russell y Norvig

LA UTILIDAD DEL DINERO



(a) Datos empíricos de Beard sobre un rango limitado.

(b) Curva típca con rango completo

$$U(S_{k+n}) = -263.31 + 22.09 \log(n + 150,000)$$

$$n = -\$150,000 \text{ and } n = \$800,000$$

- Los agentes que manejan una curva sigmoidea sufrirán miedo al riesgo.
- En la región de desesperación o de valores muy negativos, el comportamiento se caracteriza por buscar el riesgo.

La diferencia entre el **valor monetario esperado** y su **equivalente de certeza** se conoce como **prima** del seguro.

Para pequeños cambios en la riqueza esperada frente a la actual, prácticamente cualquier curva se comporta linealmente. **(neutral al riesgo)**

fuente: Inteligencia Artificial. Un enfoque Moderno - Russell y Norvig

ESCALAS DE UTILIDAD Y EVALUACIÓN DE LA UTILIDAD

No hay especificación única de función de utilidad para un agente.

En contextos deterministas la función es ordinal **(función de valor)**.

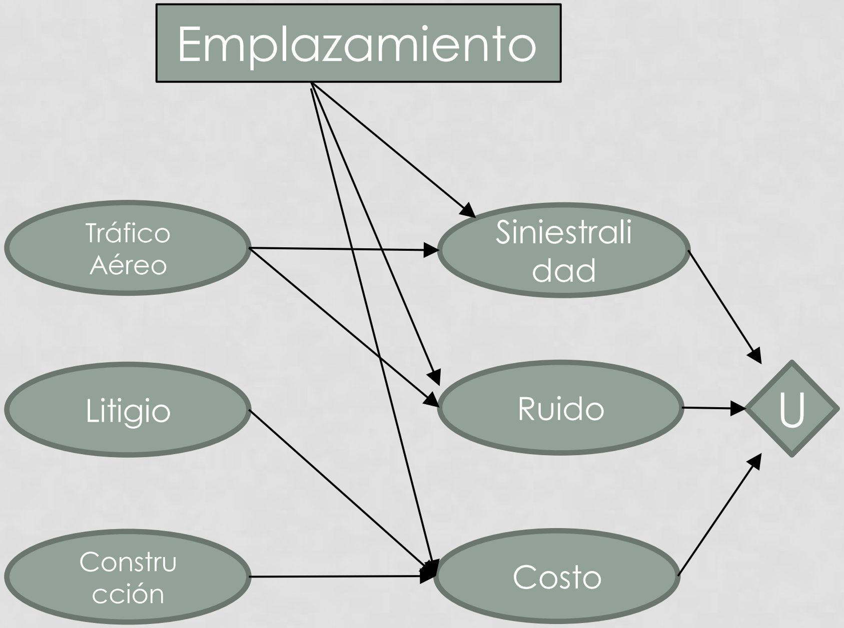
Un procedimiento para valorar utilidades es establecer una **escala**.

Los valores de utilidad normalizados varían entre 0 y 1.

fuente: Inteligencia Artificial. Un enfoque Moderno - Russell y Norvig

REDES DE DECISIÓN
(DIAGRAMA DE INFLUENCIAS)

- Este formalismo proporciona una forma de implementar agentes basados en utilidad.
- Una red de decisión representa información sobre el estado actual del agente, sus posibles acciones, el estado que resultará de la acción del agente y la utilidad del estado resultante.



Las Redes de decisión combinan Redes Bayesianas con nodos adicionales de acciones y utilidades

fuente: Inteligencia Artificial. Un enfoque Moderno - Russell y Norvig

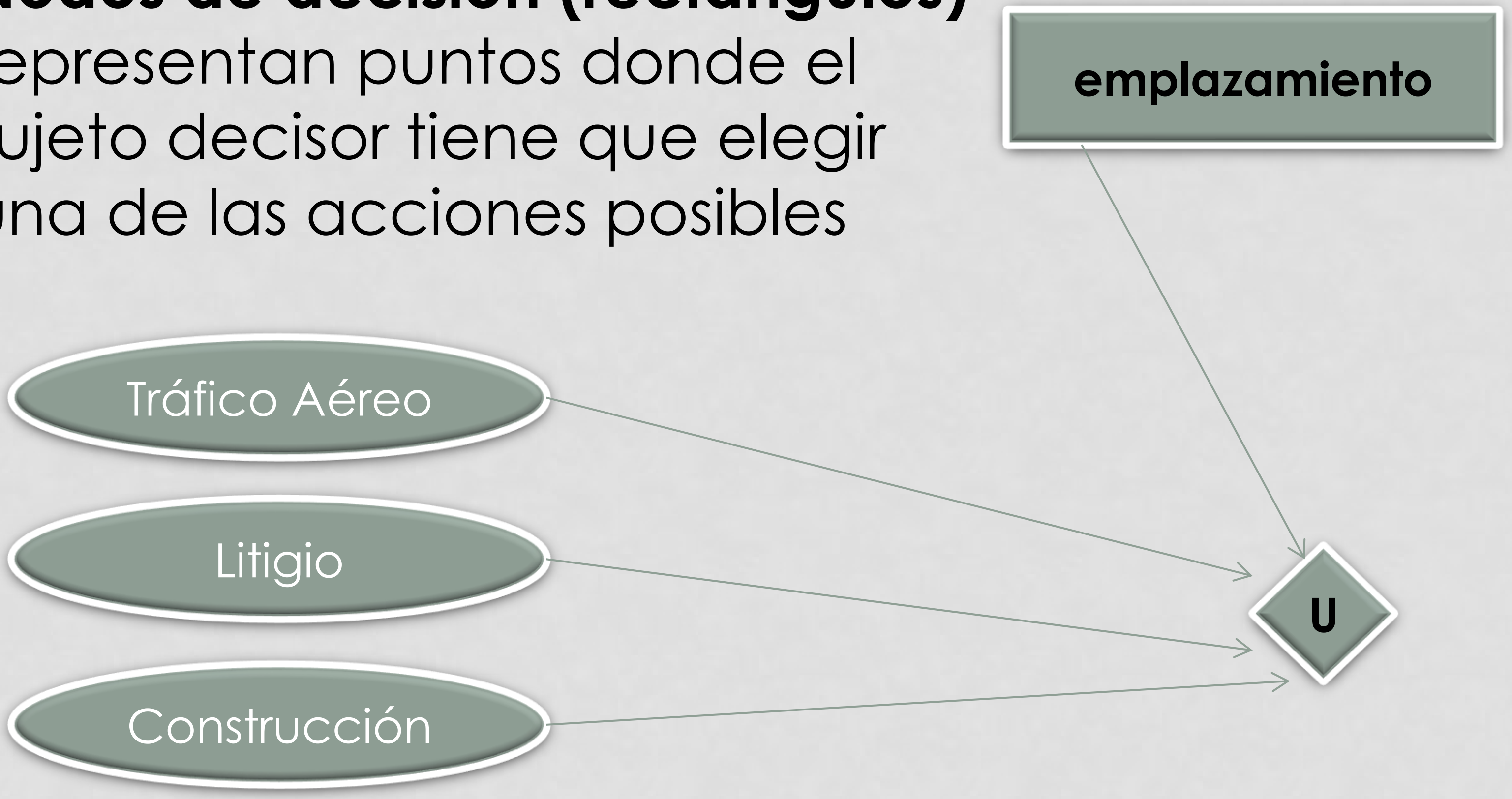
REDES DE DECISIÓN
(DIAGRAMA DE INFLUENCIAS)

1) **Nodos aleatorios (elipses)** representan **variables aleatorias** y están asociados con una distribución de probab. condicionada para cada combinación de valores de sus padres.
   Los padres pueden ser nodos aleatorios o nodos de decisión.

fuente: Inteligencia Artificial. Un enfoque Moderno - Russell y Norvig

REDES DE DECISIÓN
(DIAGRAMAS DE INFLUENCIA)

2) **Nodos de decisión (rectángulos)** representan puntos donde el sujeto decisor tiene que elegir una de las acciones posibles



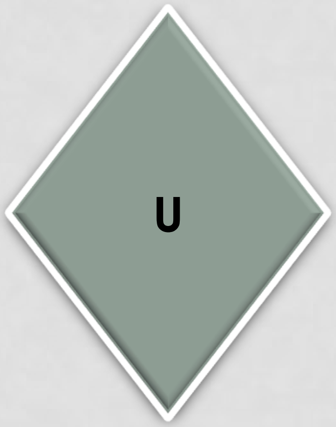
fuente: Inteligencia Artificial. Un enfoque Moderno - Russell y Norvig

REDES DE DECISIÓN
(DIAGRAMAS DE INFLUENCIA)

3. **Nodos de utilidad (rombos)** representan las funciones de utilidad del agente. Tiene como padres a todas aquellas variables descriptivas que influyen directamente sobre el valor de utilidad.

Asociado a estos nodos se tiene una descripción de la utilidad del agente como una función de los atributos padre.

La descripción puede ser una tabla de la función o una función parametrizada de tipo aditivo o multiplicativo.



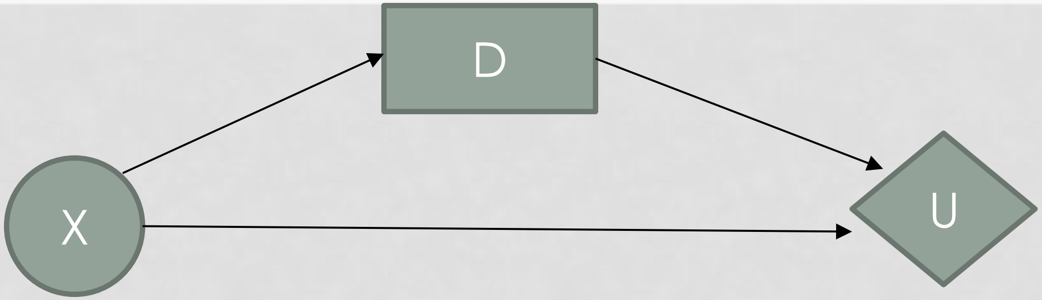
fuente: Inteligencia Artificial. Un enfoque Moderno - Russell y Norvig

SENSIBILIDAD

- Tanto las Redes Bayesianas como las Redes de Decisión son **muy sensibles** a los valores de probabilidad usados.

- Hay que realizar un gran esfuerzo por reunir **los mejores datos posibles** ya que se producen errores acumulativos.

EJEMPLO MÉDICO 1
(AL MOMENTO DE TOMAR LA DECISIÓN SABEMOS SI HAY O NO HAY INFECCIÓN)

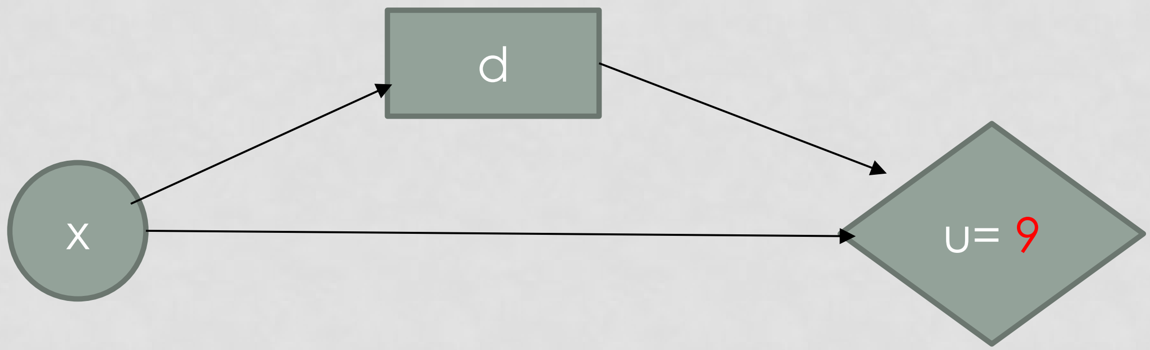


- Variable aleatoria: X (infección bacteriana); P(x)=0,86
- Decisión: D (administrar antibióticos)
- Utilidad: U (estado del paciente)

u(x,d)	x	¬ x
d	9	4
¬ d	1	10

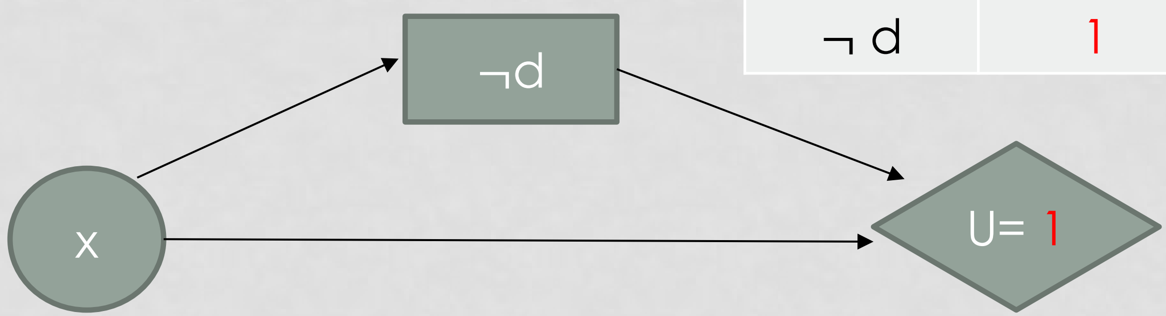
Evaluación en Redes de decisión

1- Establecer las variables de evidencia para el estado actual



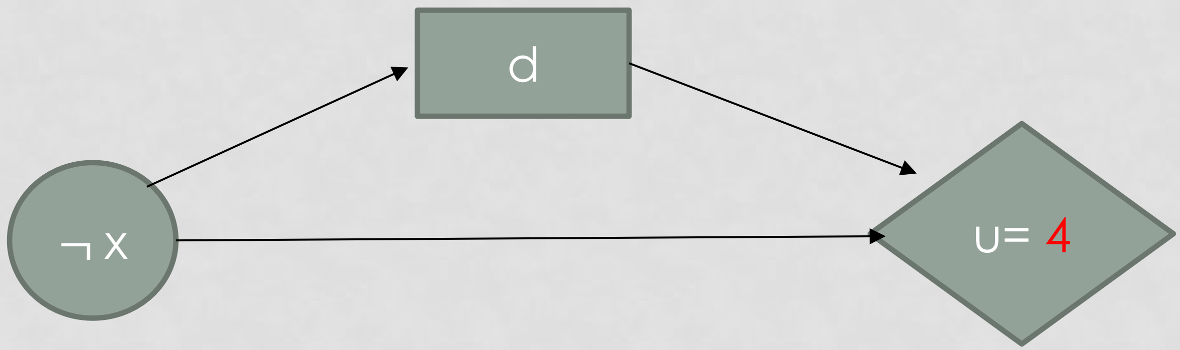
u(x,d)	x	¬ x
d	9	4
¬ d	1	10

Paciente enfermo



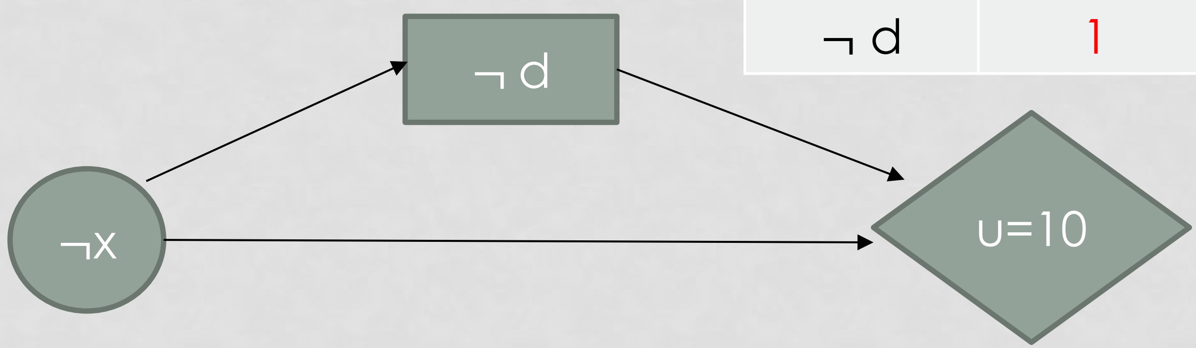
Evaluación en Redes de decisión

1- Establecer las variables de evidencia para el estado actual



u(x,d)	x	¬ x
d	9	4
¬ d	1	10

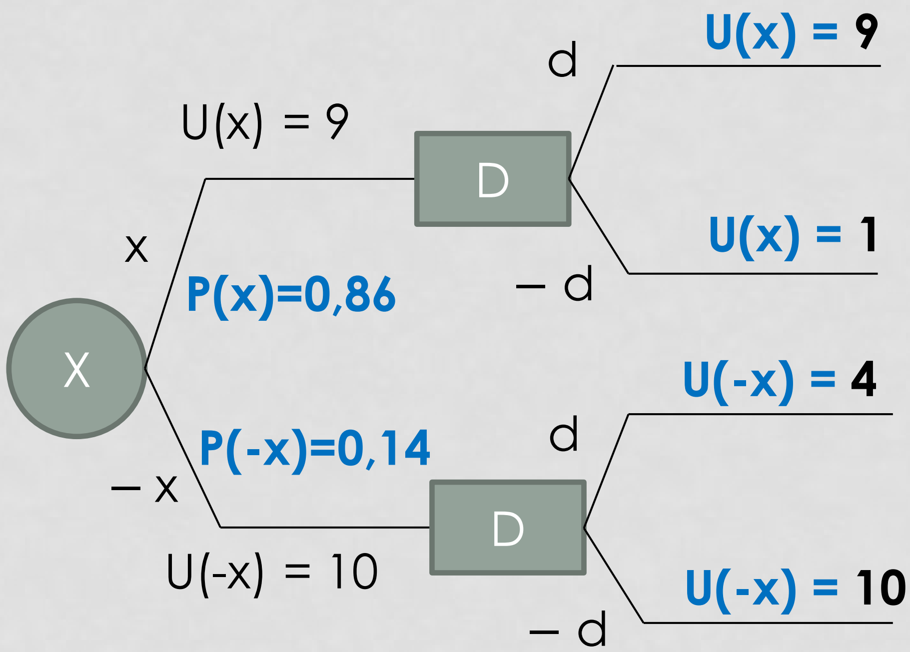
Paciente sano



Evaluación en Redes de decisión

1- Establecer las variables de evidencia para el estado actual

2- Para cada valor posible del nodo de decisión asignar el mayor valor de utilidad



DATOS

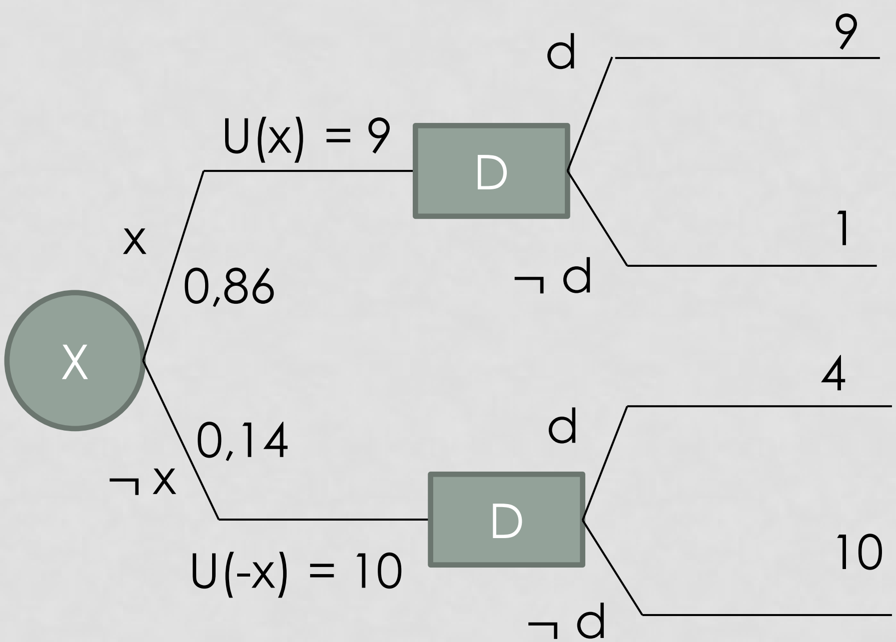
P(x)=0,86

U(X,D)	x	¬ x
d	9	4
¬ d	1	10

Evaluación en Redes de decisión

1- Establecer las variables de evidencia para el estado actual

2- Para cada valor posible del nodo de decisión calcular la utilidad resultante para la acción.



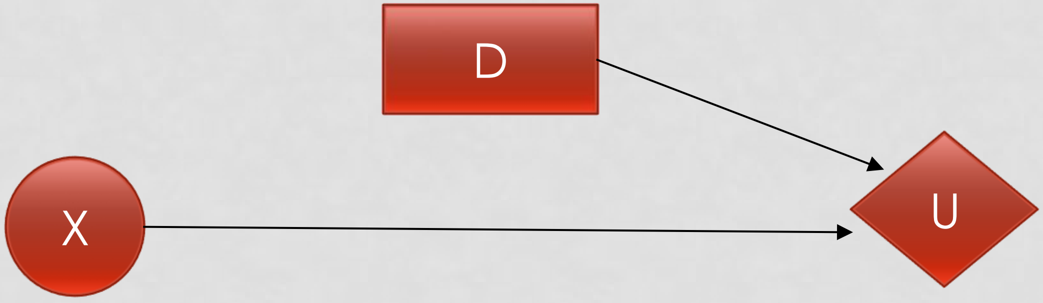
¿Decisión óptima?

+x → administrar antibiótico
-x → no administrar antibiótico

MU(x)=0,86 x 9 = 7,74

fuente: Inteligencia Artificial. Un enfoque Moderno - Russell y Norvig

EJEMPLO MÉDICO 2
(AL MOMENTO DE TOMAR LA DECISIÓN NO SABEMOS SI HAY O NO INFECCIÓN)

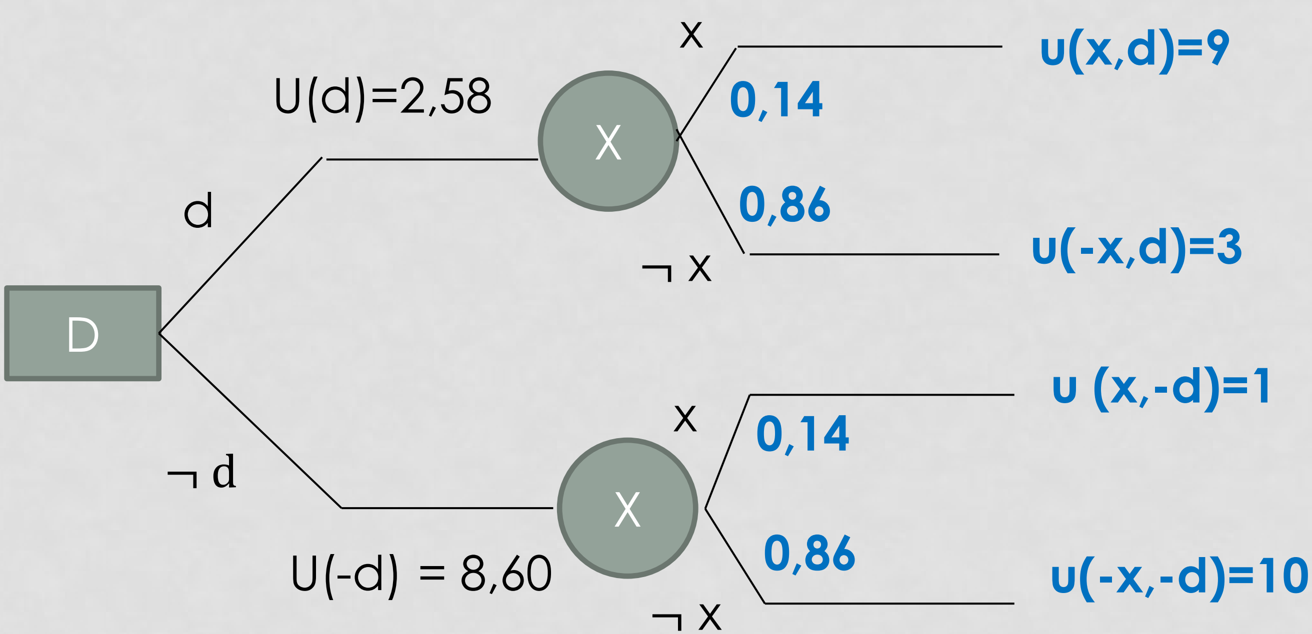


U(X,D)	x	¬ x
d	9	3
¬ d	1	10

EJEMPLO MÉDICO 2

(AL MOMENTO DE TOMAR LA DECISIÓN NO SABEMOS SI HAY O NO INFECCIÓN)

DATO
P(x)=0,14



Decisión óptima: No aplicar antibióticos

DATO

U(X,D)	x	¬ x
d	9	3
¬ d	1	10

EL VALOR DE LA INFORMACIÓN

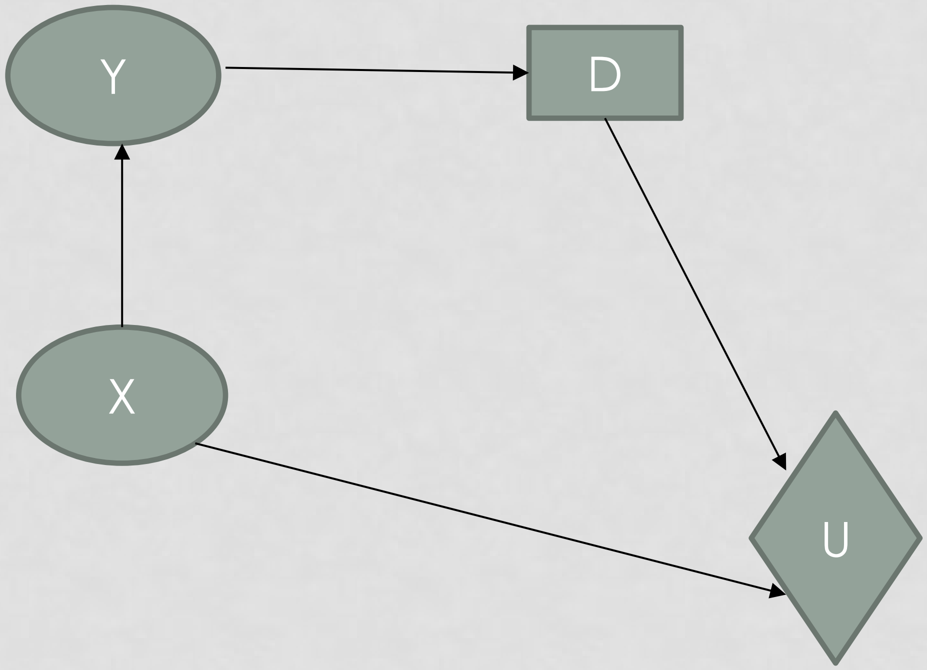
Uno de los aspectos más importantes en la toma de decisiones es el conocimiento de **"qué preguntas hacer".**

La **teoría del valor de la información** capacita a un agente para seleccionar qué información adquirir.
La adquisición de información se consigue mediante **acciones de percepción.**

En general el valor de la información se define como la diferencia entre el valor esperado de las mejores acciones antes y después de disponer de la información.

fuente: Inteligencia Artificial. Un enfoque Moderno - Russell y Norvig

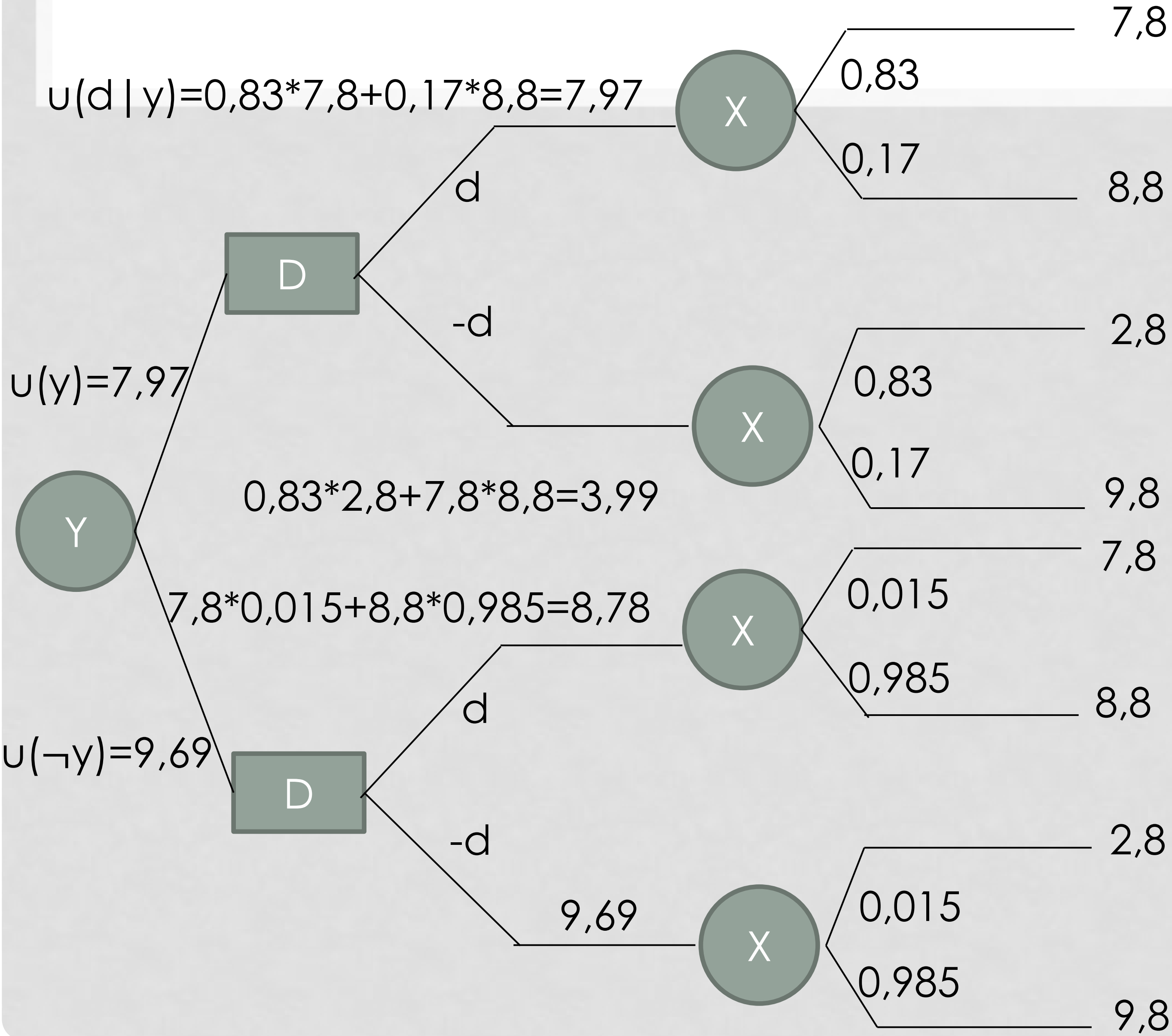
EJEMPLO MÉDICO 3
(AL MOMENTO DE TOMAR LA DECISIÓN DISPONEMOS DE UN ANÁLISIS "Y")



u(x,d)	x	¬ x
d	7,8	8,8
¬ d	2,8	9,8

EJEMPLO MÉDICO 3

(AL MOMENTO DE TOMAR LA DECISIÓN DISPONEMOS DE UN ANÁLISIS “Y”)



u(x,d)	**x**	**¬ x**
d	7,8	8,8
¬ d	2,8	9,8

x	**y**	**P(X | Y)**
V	V	0,83
F	V	0,17
V	F	0,015
F	F	0,985

Ejemplo ¿Cuánto debería pagar por la información?:

Una compañía petrolera pretende adquirir uno de los n bloques indistinguibles de derechos de perforaciones oceánicas. Sólo uno de los bloques contiene un valor petrolífero de C dólares y el precio de cada bloque es de C/n dólares.

Un geólogo ofrece a la Cia. los resultados de un estudio sobre el bloque 3.

¿Cuánto dinero debería estar dispuesta a pagar la Cía. por esa información?

$p_3 = 1/n$; Beneficio esperado $= C - C/n = \left(\frac{n-1}{n}\right)C$

$p_{\neg 3} = 1 - 1/n = \left(\frac{n-1}{n}\right)$; Beneficio esperado $= \left(\frac{1}{n-1}\, C \;- C/n\right) = \frac{1}{n\,(n-1)} C$

Siendo el beneficio esperado, conocido el resultado del informe

$$\frac{1}{n}\left(\frac{n-1}{n}\right) C + \left(\frac{n-1}{n}\right)\frac{1}{n\,(n-1)}\, C = C/n$$

fuente: Inteligencia Artificial. Un enfoque Moderno - Russell y Norvig

La información tiene valor en la medida en que tiene probabilidad de provocar la modificación de un plan y en la medida en que el nuevo plan resulte significativamente mejor que el anterior.

Se utiliza la expresión **VALOR DE LA INFORMACIÓN PERFECTA** (VIP) ya que habitualmente se supone la obtención de una evidencia exacta acerca de una variable aleatoria.

El Valor de la información es no negativo.
Esto implica que la información no es perjudicial.

fuente: Inteligencia Artificial. Un enfoque Moderno - Russell y Norvig

SISTEMAS EXPERTOS BASADOS EN LA TEORÍA DE LA INFORMACIÓN

La incorporación de Redes de decisión se traduce en que se pueden desarrollar SE que recomienden decisiones óptimas, considerando tanto las preferencias del usuario como la evidencia disponible.

Un SE basado en la teoría de la decisión se puede crear partiendo, al menos, de un equipo formado por "un experto del dominio" y "un ingeniero del conocimiento".

fuente: Inteligencia Artificial. Un enfoque Moderno - Russell y Norvig

SISTEMAS EXPERTOS BASADOS EN LA TEORÍA DE LA INFORMACIÓN

El análisis de la decisión se aplica en dominios donde entran en juego intereses importantes (mundo empresarial, militar, diagnóstico médico, diseños de ingeniería y gestión de recursos, entre otros).
El proceso incluye un estudio cuidadoso de todas las acciones posibles y sus consecuencias, así como las preferencias planteadas en cada resultado.

Sujeto decisor: el que establece preferencias entre los resultados

Analista de decisiones: enumera las acciones y resultados posibles y obtiene preferencias del sujeto decisor para determinar el mejor curso de la acción.

Actualmente cada vez se automatizan más procesos de decisión y estas teorías se utilizan para garantizar que los procesos automáticos se comportan tal y como se desea.

fuente: Inteligencia Artificial. Un enfoque Moderno - Russell y Norvig

SISTEMAS EXPERTOS BASADOS EN LA TEORÍA DE LA INFORMACIÓN

1- **Crear un modelo causal** :

Determinar las variables intervinientes, dibujar los arcos indicando las relaciones entre ellas.

Algunas de estas relaciones serán conocidas por el experto del dominio y otras se extraerán de informes, documentación disponible, operarios, etc.

2- **Simplificar a un modelo de decisión cualitativo**

se eliminan variables que no están involucradas en la decisión a tomar.

3- **Asignar probabilidades**

Las prob. Iniciales pueden establecerse a partir de bases de datos, recopilación de información, etc.

fuente: Inteligencia Artificial. Un enfoque Moderno - Russell y Norvig

SISTEMAS EXPERTOS BASADOS EN LA TEORÍA DE LA INFORMACIÓN

4- **Asignar Utilidades**
Establecer escalas y funciones de utilidad que permitan su evaluación

5- **Verificar y refinar el modelo**
Para evaluar el sistema se necesitará un conjunto de pares (entrada, salida) que se conozca que son correctos (estándar de oro) para comparar.

6- **Realizar un análisis de sensibilidad**
Si cambios pequeños en las probabilidades y utilidades asignadas conducen a decisiones significativamente diferentes entonces podría merecer la pena gastar más recursos en obtener mejores datos.
El análisis revela que muchos de los valores necesitan especificarse sólo muy aproximadamente.
Si todas las variaciones conducen a la misma decisión, entonces el usuario tendrá una mayor confianza en que la decisión es la correcta.

fuente: Inteligencia Artificial. Un enfoque Moderno - Russell y Norvig

Árboles de decisión

Un Árbol de Decisión es un conjunto de condiciones organizadas en una estructura jerárquica, de tal manera que la decisión final a tomar se puede determinar siguiendo las condiciones que se cumplen desde la raíz del árbol hasta alguna de sus hojas.

¿Kara o Markus?

Algoritmo ID3

Una de las primeras técnicas asociadas a la inducción por medio de árboles de decisión fue la presentada por Ross Quinlan en 1983 y denomina **ID3 (Induction Decision Trees)**.

DATOS

Atributos: son los factores que influencian la clasificación o decisión.

La selección de atributos debe basarse en el conocimiento acumulado por la experiencia.

En este algoritmo cada atributo forma un nodo intermedio en un árbol cuyas hojas o nodos terminales son las clases o decisiones.

Clase: posibles valores de solución

Ejemplos: es el conjunto de combinaciones de atributos dados.

Dado el conjunto de ejemplos, el ID3 selecciona el atributo que subdivide los ejemplos de la mejor manera.

Algoritmo ID3

La elección del mejor atributo se establece mediante la **entropía**, eligiendo aquel valor que proporcione una mejor ganancia de información.

La función de entropía más usada es la binaria:

$$I(p,n) = -\left(\frac{p}{p+n}\right) log_2\left(\frac{p}{p+n}\right) - \left(\frac{n}{p+n}\right) log_2\left(\frac{n}{p+n}\right)$$

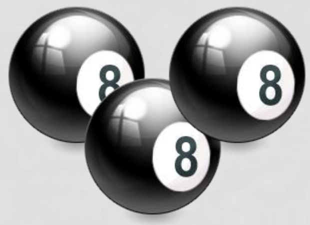
p: conj. de ejemplos positivos; **n**: conj. de ejemplo negativos

Entropía

Es la medida de la incertidumbre que hay en un sistema. Es decir, ante una determinada situación, la probabilidad de que ocurra cada uno de los posibles resultados.

Interpretación de la entropía

Un ejemplo de la entropía binaria podría ser sacar una bola de color rojo o negro de una bolsa.

Si en la bolsa hay  

el resultado es completamente desconocido, es decir la incertidumbre es máxima, es decir la entropía es 1.

Si en la bolsa hay 6 bolas negras el resultado es conocido de antemano, luego no existe incertidumbre y la entropía es 0.

Ganancia

Es la diferencia entre la entropía de un nodo y la inicial.

Es una heurística que sirve para la elección del mejor atributo.

Un buen criterio parece ser escoger el atributo que gana la mayor información.

ID3 examina todos los atributos y escoge el de máxima ganancia, forma la ramificación y usa el mismo proceso recursivamente para formar sub-árboles a partir de los nodos generados.

$$\boldsymbol{Ganancia(A) = I(p,n) - Einicial(p,n)}$$

Ejemplo ID3

9 clases P **5 clases N**

$$Ei(p,n) = -\left(\frac{p}{d}\right) log_2\left(\frac{p}{d}\right) - \left(\frac{n}{d}\right) log_2\left(\frac{n}{d}\right)$$

$$\mathbf{Ei}(9,5) = -\frac{9}{14} log_2\frac{9}{14} - \frac{5}{14} log_2\frac{5}{14}$$

= **0,940 entropía inicial**

Caso #	Aspecto	Temp.	Humedad	Viento	Clase Jugar tenis
1	Soleado	Alta	Alta	Calmo	N
2	Soleado	Alta	Alta	Ventoso	N
3	nublado	Alta	Alta	Calmo	P
4	Lluvioso	Normal	Alta	Calmo	P
5	Lluvioso	Fría	Normal	Calmo	P
6	Lluvioso	Fría	Normal	Ventoso	N
7	Nublado	Fría	Normal	Ventoso	P
8	Soleado	Normal	Alta	Calmo	N
9	Soleado	Fría	Normal	Calmo	P
10	Lluvioso	Normal	Normal	Calmo	P
11	Soleado	Normal	Normal	Ventoso	P
12	Nublado	Normal	Alta	Ventoso	P
13	Nublado	Alta	Normal	Calmo	P
14	Lluvioso	Normal	Alta	Ventoso	N

ELECCIÓN DEL NODO RAÍZ

G(H,Ei)=

$$-\frac{7}{14}.(-1) - \frac{7}{14}.(-1) - 0,94 = 0,06$$

G(V,EI) $-\frac{8}{14}.(-0,807) - \frac{6}{14}.(-1.222) - 0,94 = 0,045$

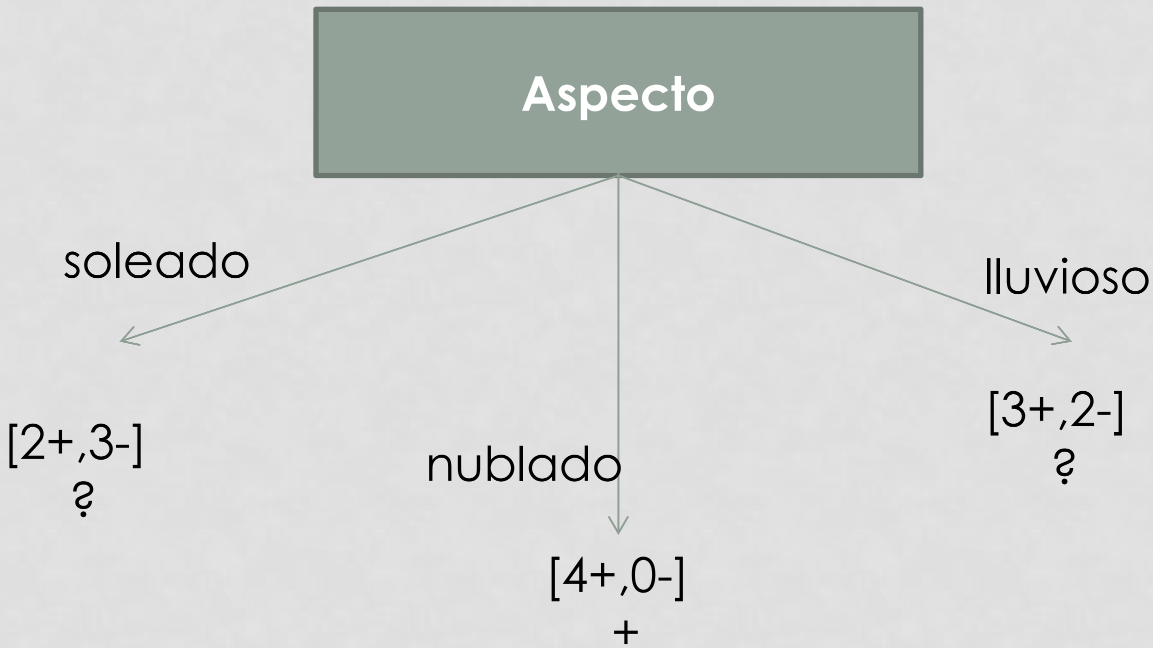
G(A,EI)=

$$-\frac{5}{14}.(-1,485) - \frac{4}{14}.(-1,807) - \frac{5}{14}.(-1,485) - 0,94 = \mathbf{0,637}$$

G(T,EI)=

$$-\frac{4}{14}.(-1,807) - \frac{6}{14}.(-1.22) - \frac{4}{14}.(-1,807) - 0,94 = 0,616$$

Atributo	Valores	P	N
Aspecto	Soleado	2	3
	Nublado	4	0
	lluvioso	3	2



Aspecto

soleado

Caso #	Temp.	Humedad	Viento	Clase
1	Alta	Alta	Calmo	N
2	Alta	Alta	Ventoso	N
8	Normal	Alta	Calmo	N
9	Fría	Normal	Calmo	P
11	Normal	Normal	Ventoso	P

nublado

Caso #	Temp.	Humedad	Viento	Clase
3	Alta	Alta	Calmo	P
7	Fría	Normal	Ventoso	P
12	Normal	Alta	Ventoso	P
13	Alta	Normal	Calmo	P

lluvioso

Caso #	Temp.	Humedad	Viento	Clase
4	Normal	Alta	Calmo	P
5	Fría	Normal	Calmo	P
6	Fría	Normal	Ventoso	N
10	Normal	Normal	Calmo	P
14	Normal	Alta	Ventoso	N

Aspecto

soleado

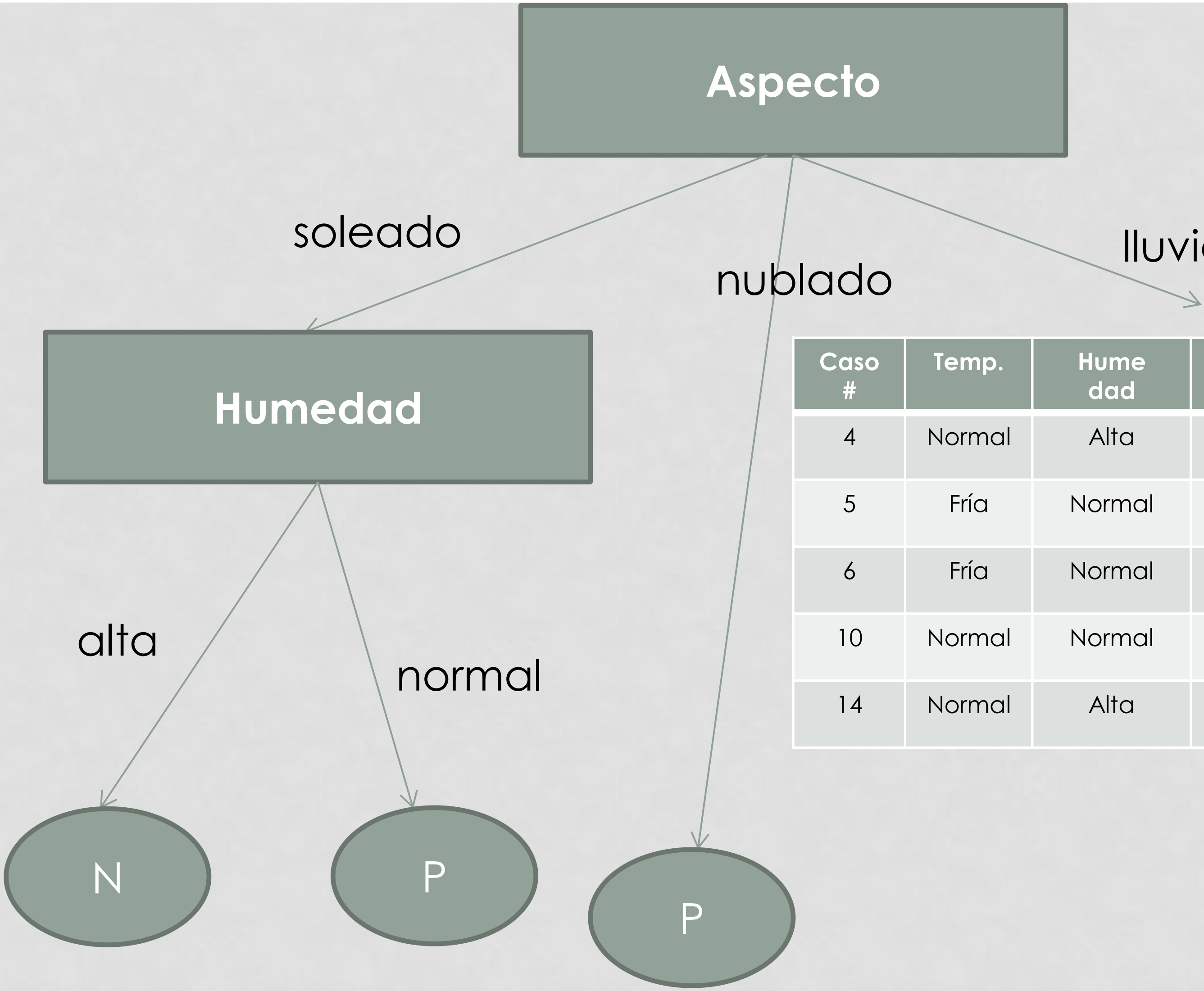
Caso #	Temp.	Humedad	Viento	Clase
1	Alta	Alta	Calmo	N
2	Alta	Alta	Ventoso	N
8	Normal	Alta	Calmo	N
9	Fría	Normal	Calmo	P
11	Normal	Normal	Ventoso	P

nublado

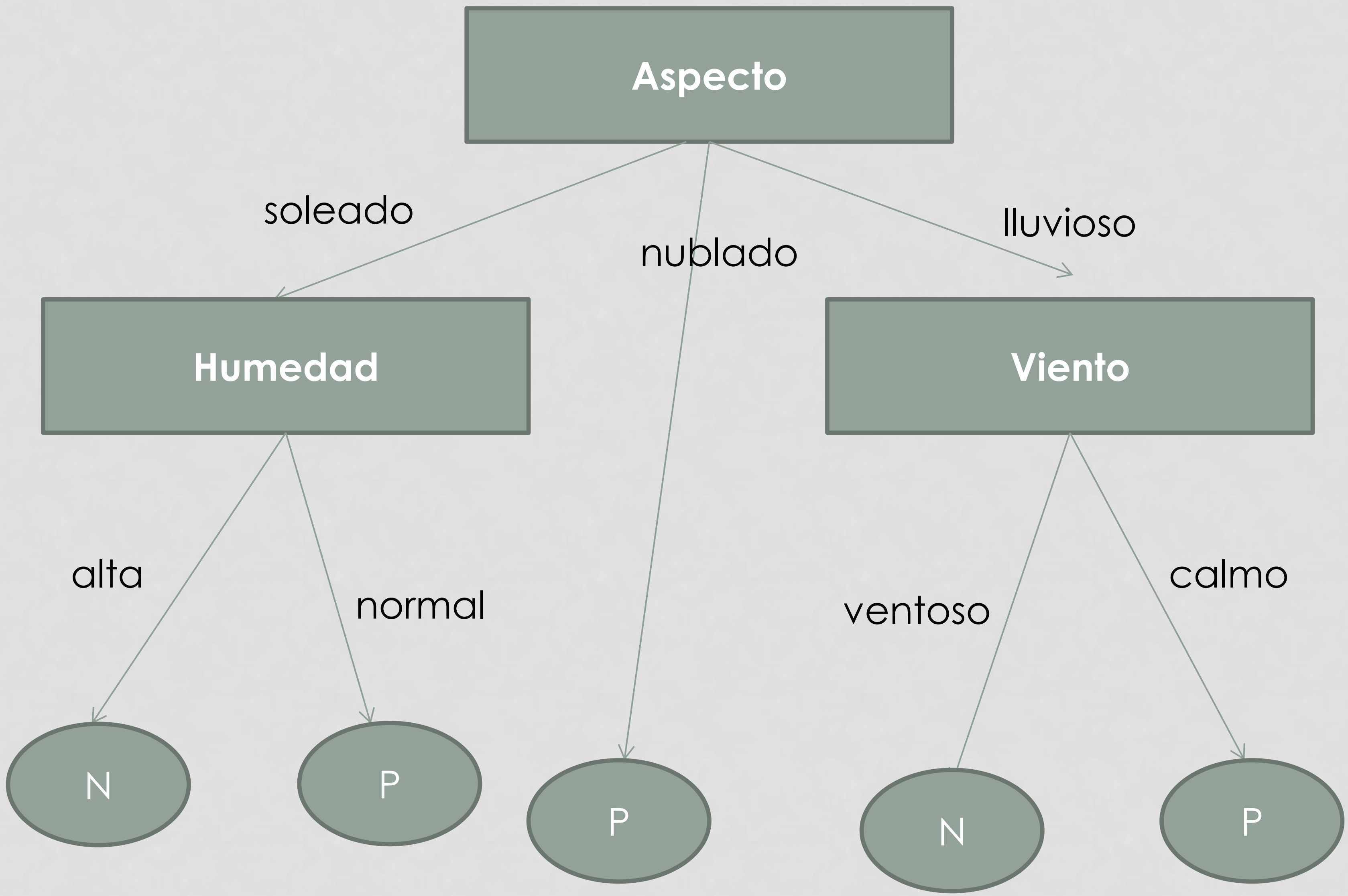
P

lluvioso

Caso #	Temp.	Humedad	Viento	Clase
4	Normal	Alta	Calmo	P
5	Fría	Normal	Calmo	P
6	Fría	Normal	Ventoso	N
10	Normal	Normal	Calmo	P
14	Normal	Alta	Ventoso	N



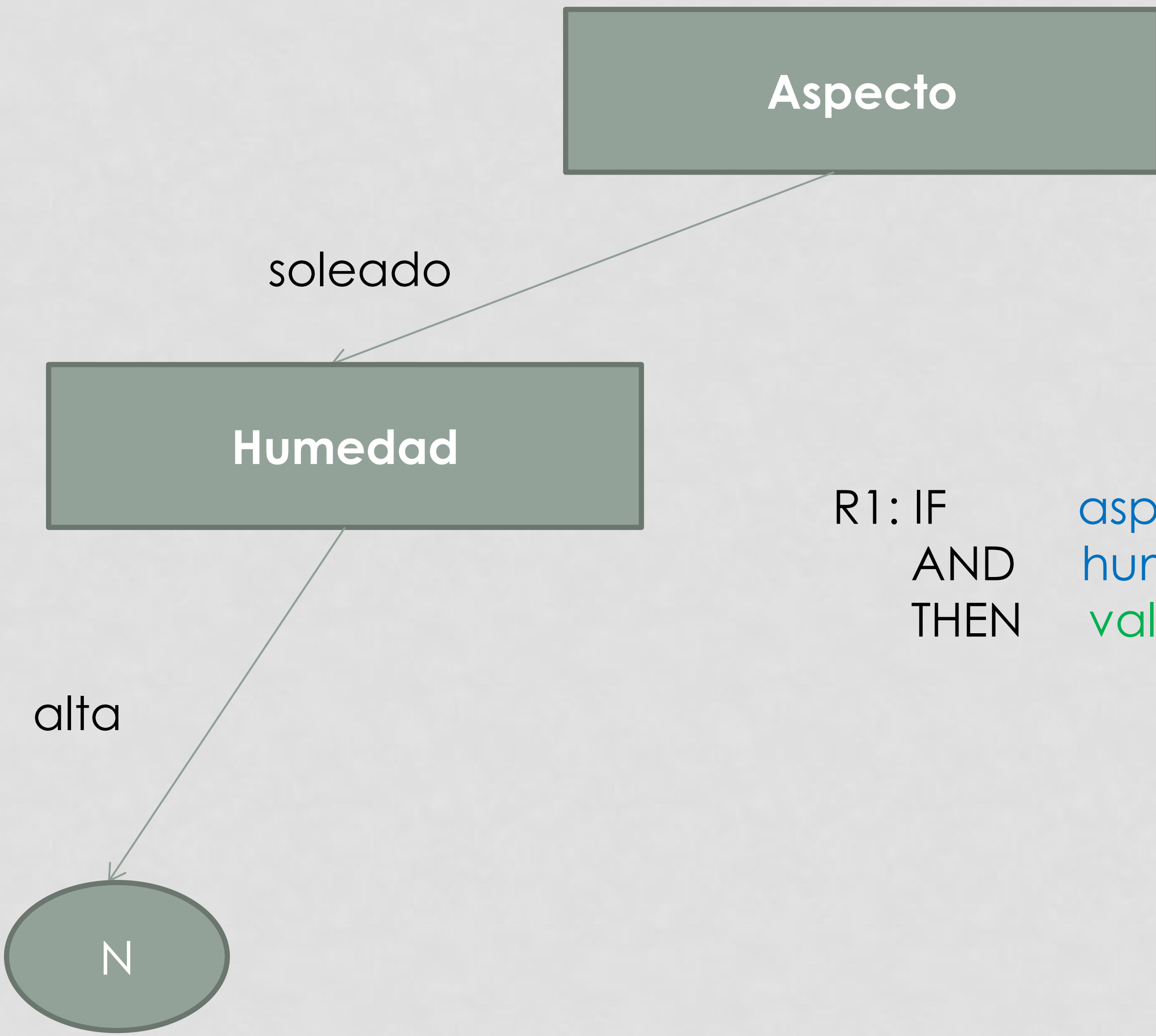
Caso #	Temp.	Hume dad	Viento	Clase
4	Normal	Alta	Calmo	P
5	Fría	Normal	Calmo	P
6	Fría	Normal	Ventoso	N
10	Normal	Normal	Calmo	P
14	Normal	Alta	Ventoso	N


Aspecto
soleado
nublado
lluvioso
Humedad
Viento
alta
normal
ventoso
calmo
N
P
P
N
P

INCONVENIENTES

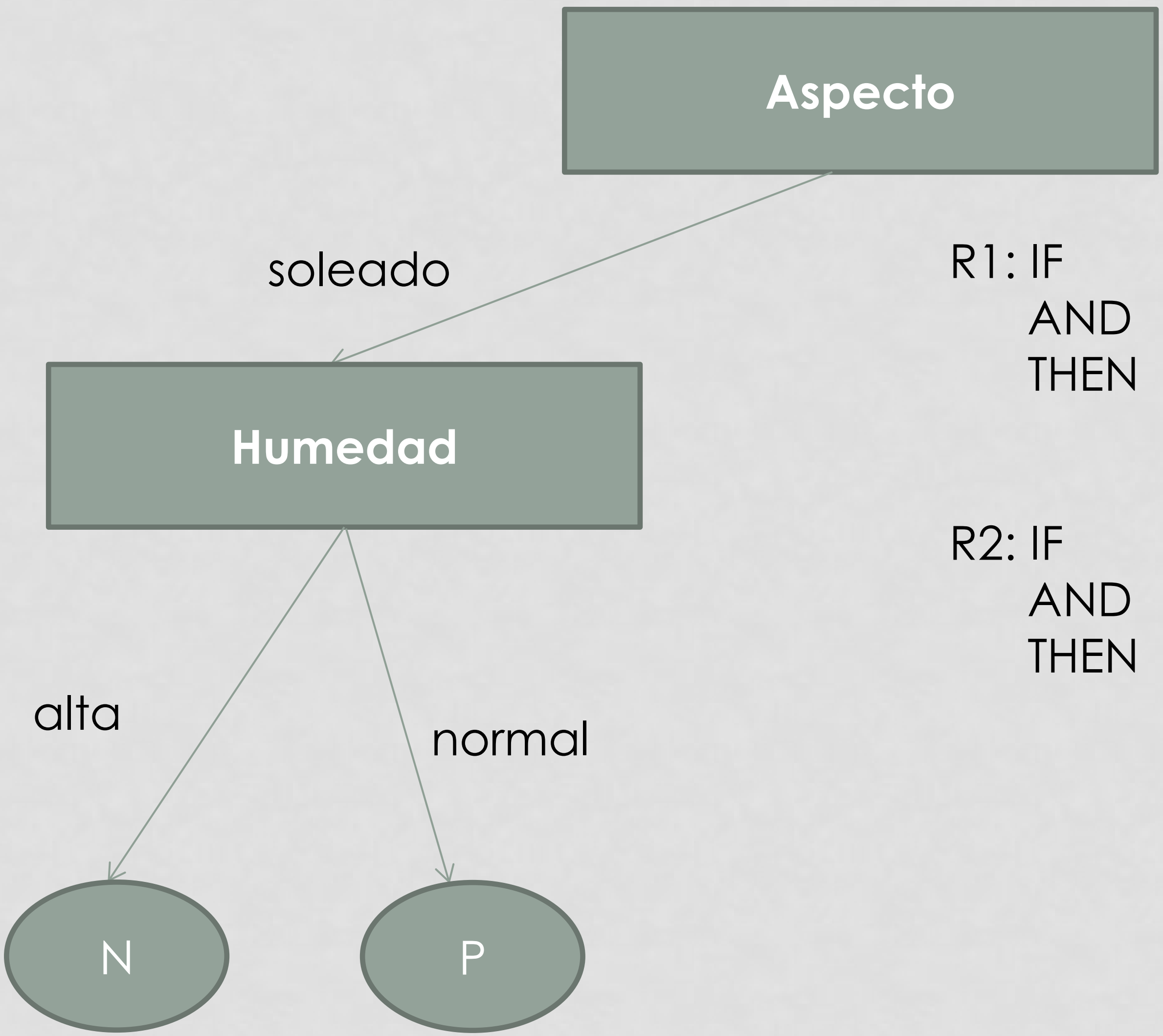
- Favorece indirectamente a aquellos atributos con muchos valores, los cuales no tienen que ser los más útiles.
- Genera árboles de decisión a partir de ejemplos de partida.
- Generación de grandes árboles de decisión que no representan garantía de reglas eficientes.
- Aplicables sólo a problemas de clasificación y diagnóstico.

SISTEMA EXPERTO
REGLAS DE PRODUCCIÓN



R1: IF aspecto = soleado
AND humedad= alta
THEN valor=N

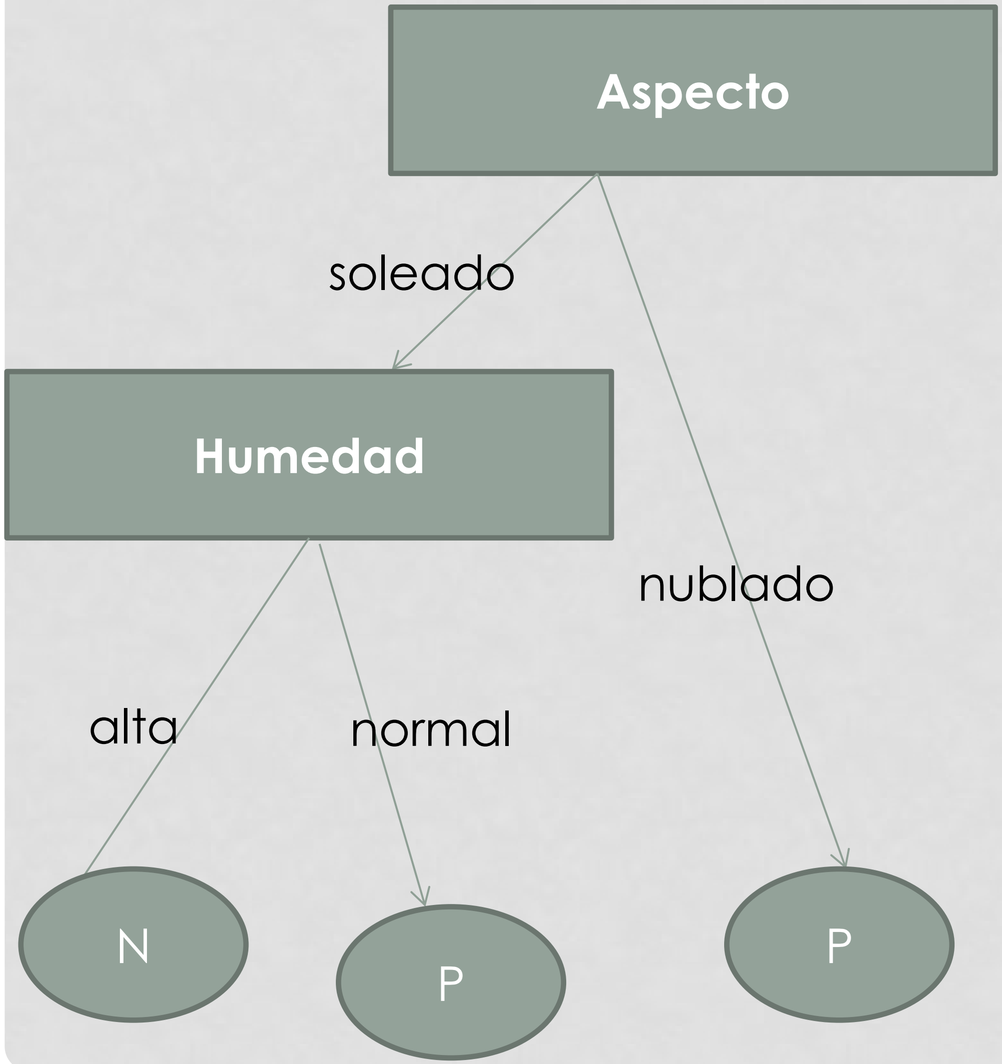
SISTEMA EXPERTO
REGLAS DE PRODUCCIÓN



R1: IF aspecto = soleado
AND humedad= alta
THEN valor=N

R2: IF aspecto = soleado
AND humedad= normal
THEN valor=P

SISTEMA EXPERTO
REGLAS DE PRODUCCIÓN

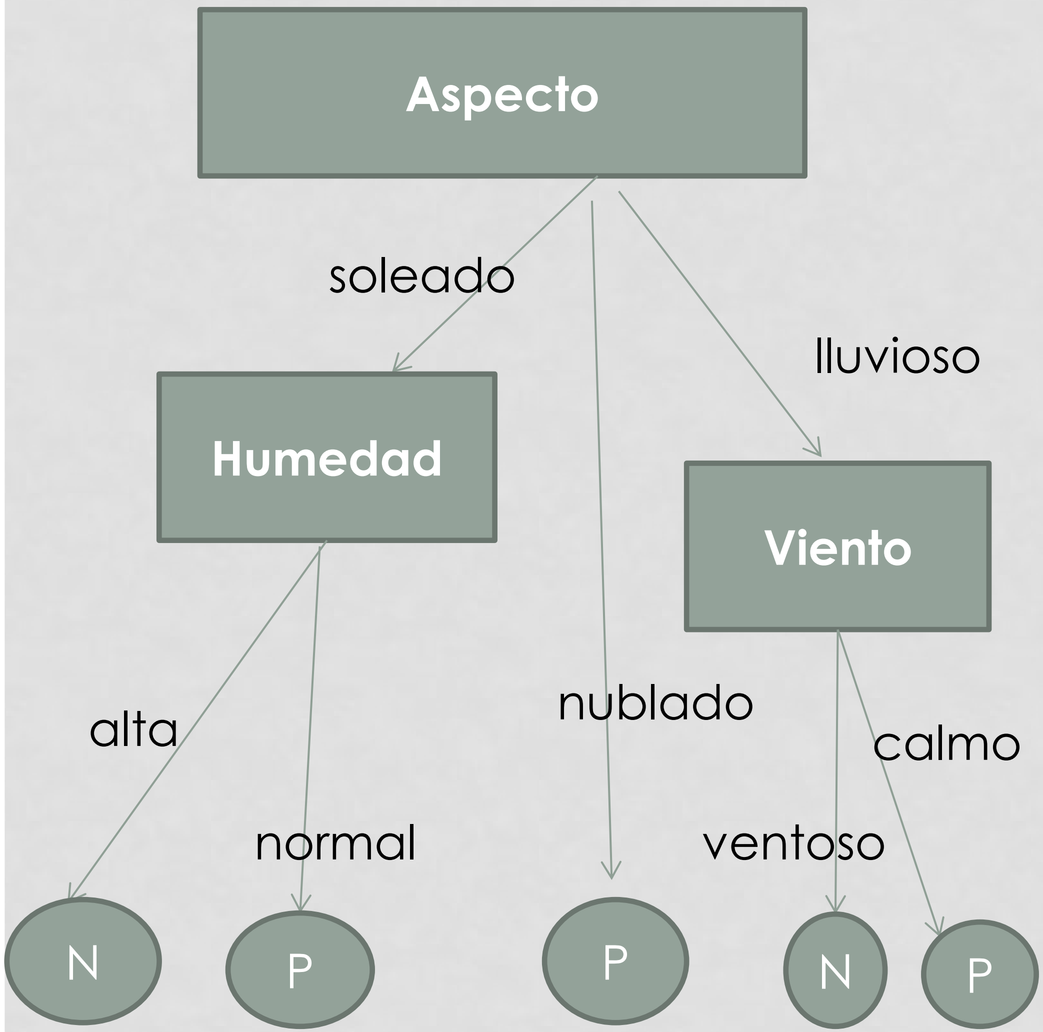


R1: IF aspecto = soleado
AND humedad= alta
THEN valor=N

R2: IF aspecto = soleado
AND humedad= normal
THEN valor=P

R3: IF aspecto = nublado
THEN valor=P

SISTEMA EXPERTO
REGLAS DE PRODUCCIÓN



R1: IF aspecto = soleado
AND humedad= alta
THEN valor=N

R2: IF aspecto = soleado
AND humedad= normal
THEN valor=P

R3: IF aspecto = nublado
THEN valor=P

R4: IF aspecto = lluvioso
AND viento= ventoso
THEN valor=N

R5: IF aspecto = lluvioso
AND viento= calmo
THEN valor=P

LENGUAJES PARA DESARROLLAR SISTEMAS EXPERTOS

- **Específicos**

LISP
Se trata de uno de los lenguajes de alto nivel más antiguos. Es un lenguaje cuya principal estructura de datos son las listas, aún cuando se han ido incorporando otras estructuras más sofisticadas como pueden ser los objetos. Tiene como ventaja el manejo de sus estructuras a muy alto nivel lo que facilita la implementación rápida de los modelos y su facilidad de modificación. Como desventaja está su relativa lentitud frente a lenguajes de propósito general como C.

PROLOG
Es un lenguaje declarativo que, a partir de los datos introducidos deduce nuevos hechos y resuelve el problema automáticamente. PROLOG tiene incluido, por tanto, un motor de inferencia que se encarga de realizar búsquedas en su base de hechos. Programar con PROLOG, por tanto, consiste en insertar hechos sobre objetos y preguntar al sistema sobre sus relaciones.

LENGUAJES PARA DESARROLLAR SISTEMAS EXPERTOS

- **Smalltalk** fue el primer lenguaje de programación que fue diseñado para basarse exclusivamente en objetos.

Shells para desarrollar S.E

Un Shell (intérprete) es un Sistema Experto que contiene la máquina de inferencia, la interface con el usuario y la base de conocimiento vacía y, puede ser empleado en la creación de diversos SE. **ART, LOOPS, KEE, EMYCIN, KAS, ARIES y el CLIPS**, que además de ser un lenguaje de programación para S.E es un Shell.

CLIPS

A Tool for Building Expert Systems

C Language Integrated Production System (CLIPS)

Desarrollado por NASA’s Johnson Space Center from 1985 to 1996