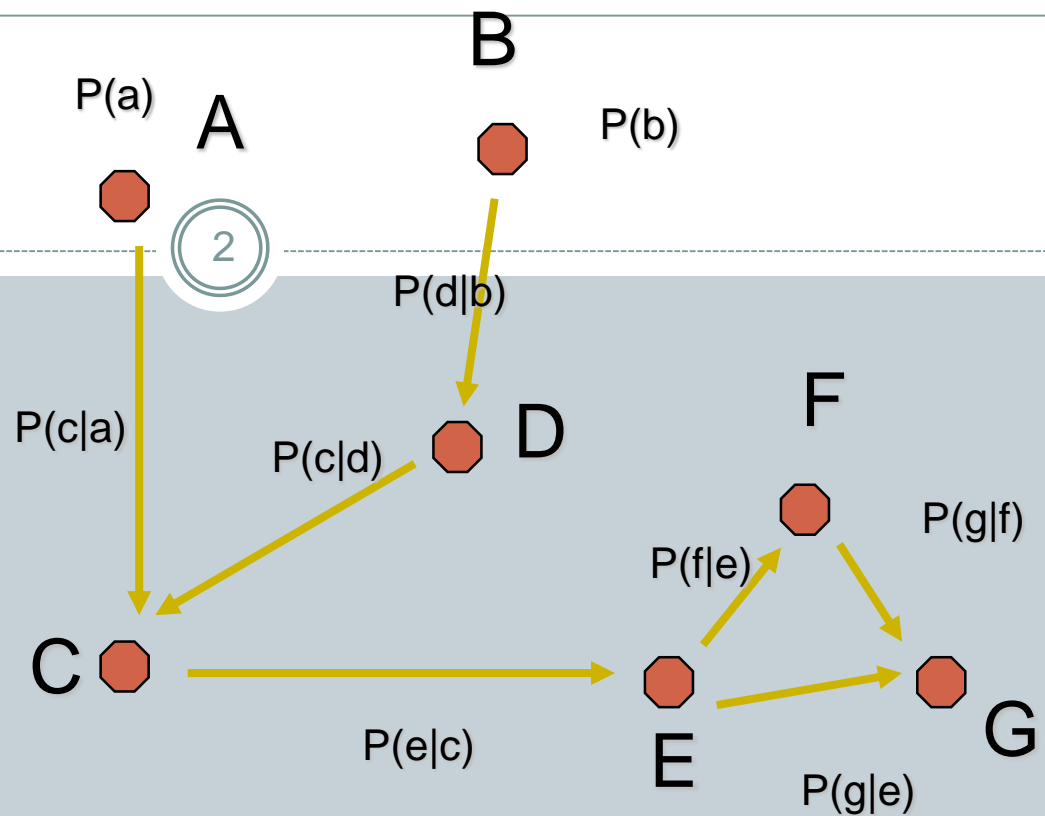


# Bayesian Belief networks

1

- Los métodos descritos hasta el momento suponen que las distribuciones por clase pueden ser parametrizadas por un vector  $\theta$
- Si existe algún conocimiento acerca de  $\theta$  también pudo ser usado
- A veces, existe información sobre las dependencias entre las componentes del vector de características
- Estas relaciones pueden representarse gráficamente como redes bayesianas.

# Red Bayesiana Discreta

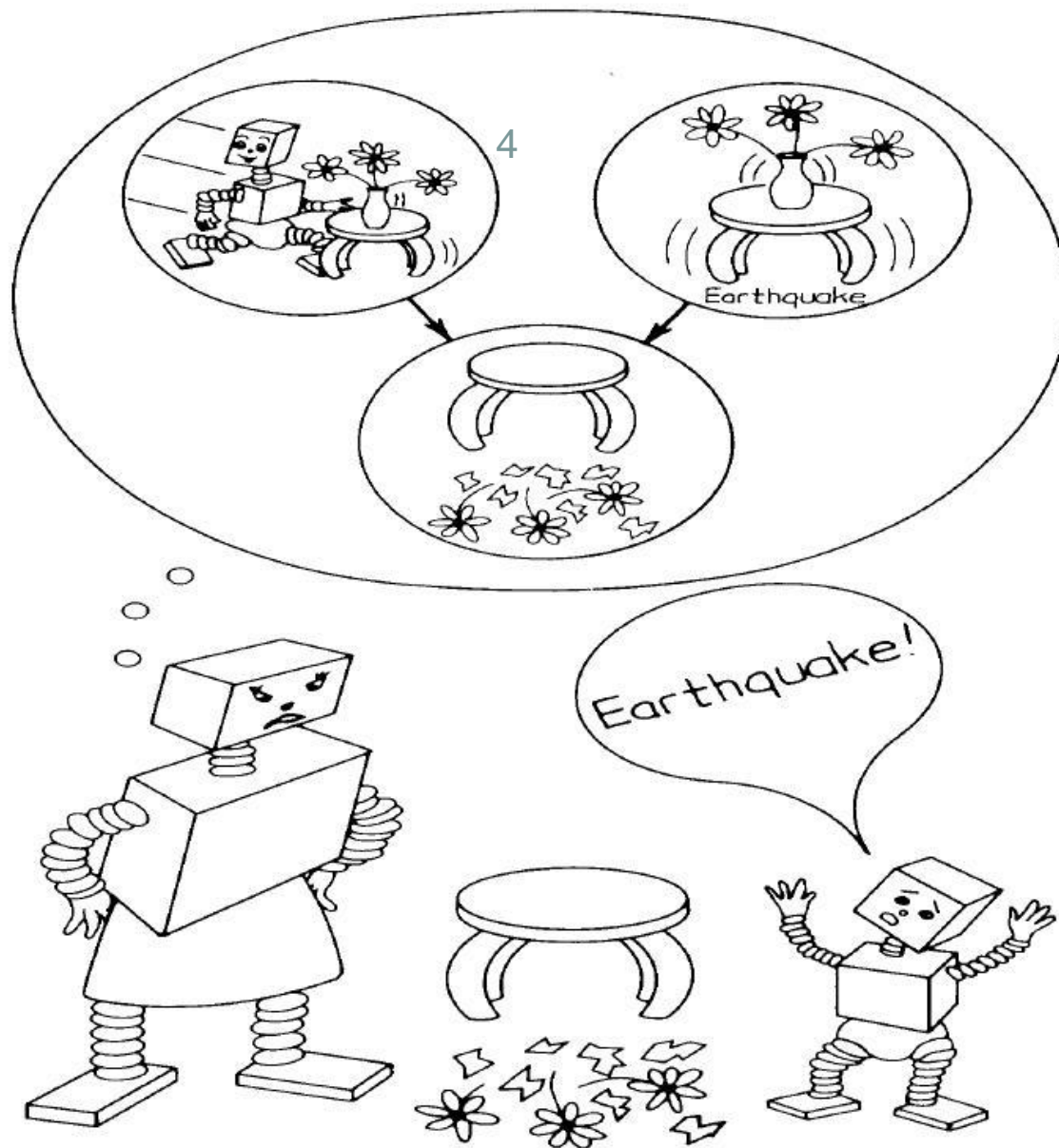


- Una red bayesiana consiste de
  - grafo dirigido acíclico conexo
  - nodos A, B, C, ... con estados asociados  $a_1, a_2, \dots, b_1, b_2, \dots$ , etc
  - probabilidades condicionales asociadas a las aristas que unen nodos
  - $P(C|A)$  puede ser una matriz con entradas  $P(c_i|a_j)$

Definicion: una red bayesiana, junto a la distribución de probabilidad condicional sobre sus variables

3



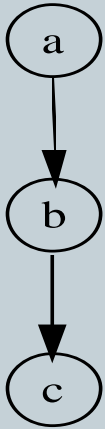


Probabilistic Reasoning in a Causal Network

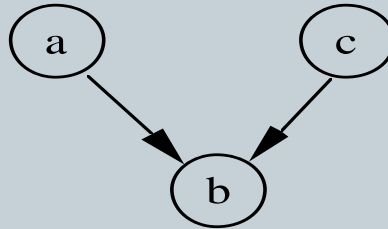
# Tipos de conexiones

5

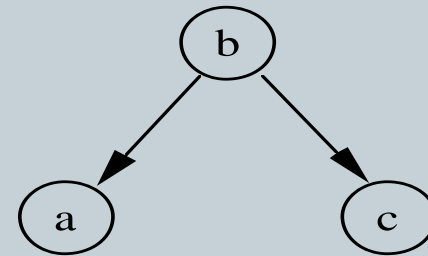
## Linear



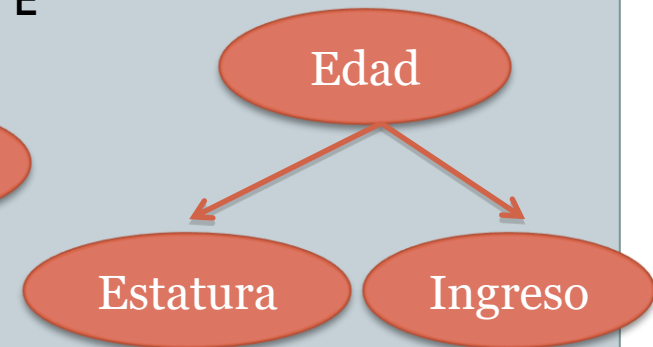
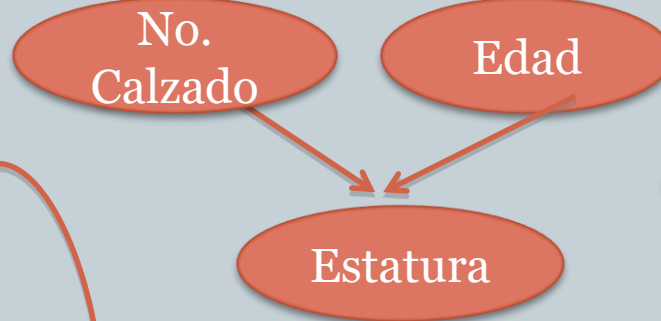
## Converging



## Diverging

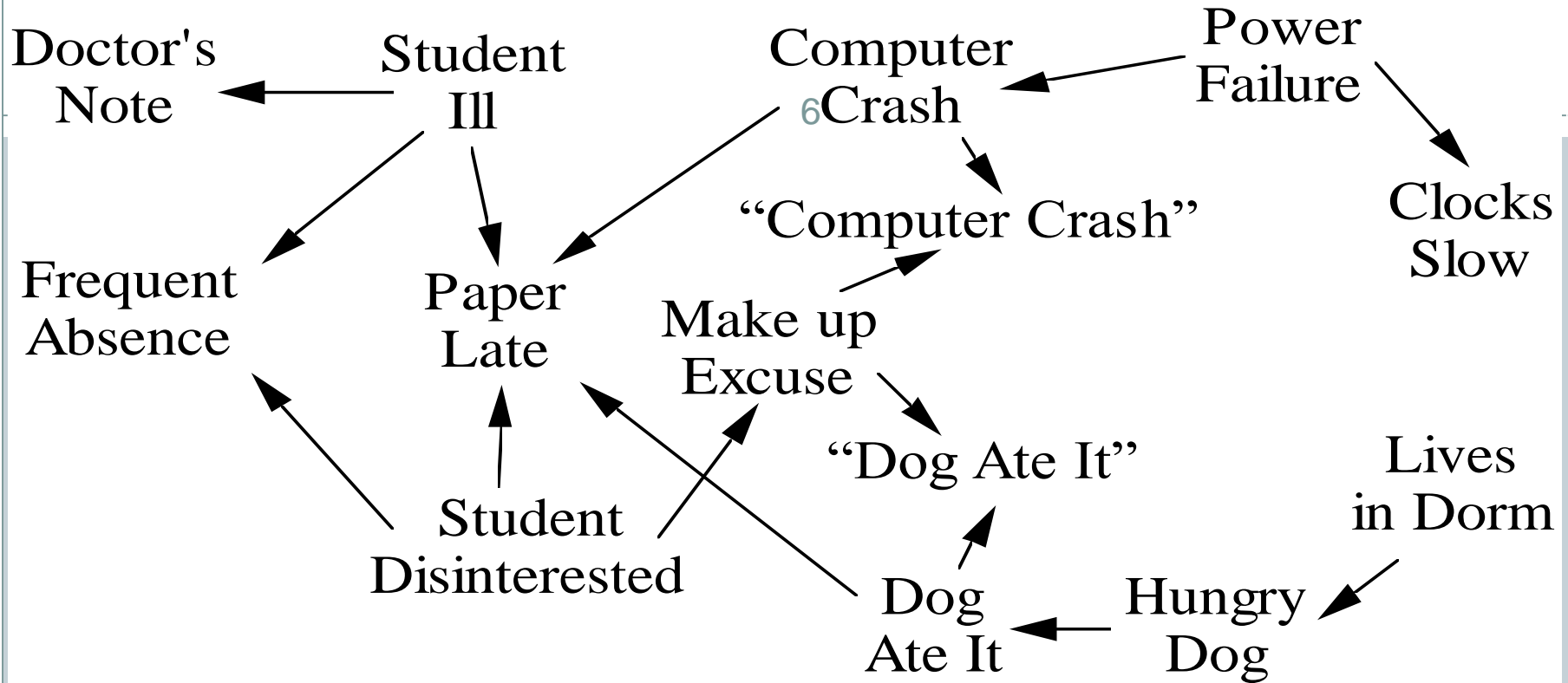


## Dependientes dado E



Independientes

Independientes



Diagnósticos?  
Predicciones?

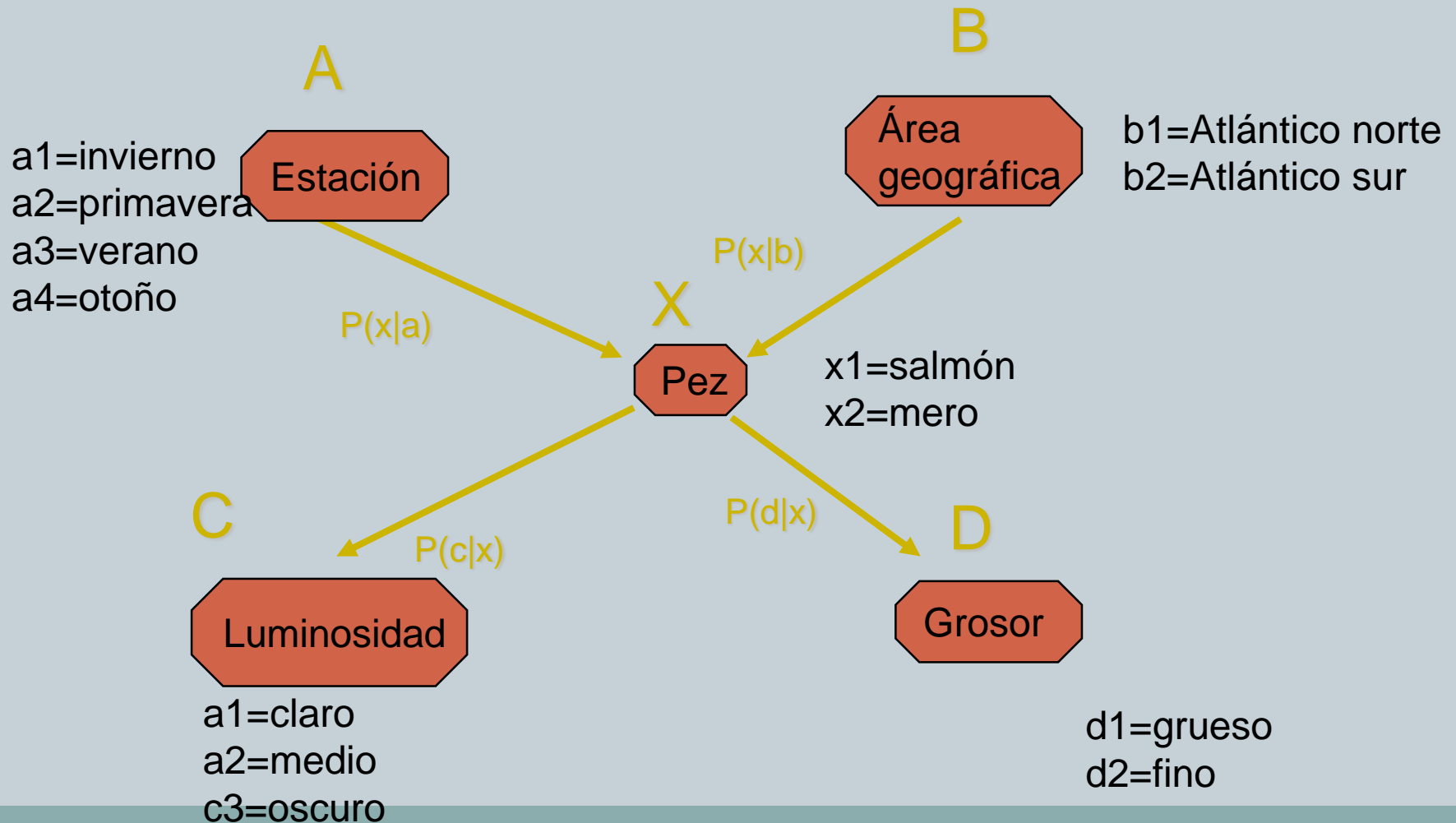
# Clasificación del mero y el salmon

7

- A representa estación, con cuatro valores equiprobables  $a_1$ =invierno,  $a_2$ =primavera,  $a_3$ =verano,  $a_4$ =otoño
- El nodo B representa el área geográfica donde el pez ha sido capturado:  $b_1$ =Atlántico norte,  $b_2$  Atlántico Sur, y son equiprobables.
- A y B son padres del nodo X, el cual representa el pez capturado, y tiene dos posibles valores, salmón o mero.
- Los nodos hijos representan luminosidad, C, con  $c_1$ =claro,  $c_2$ =medio,  $c_3$ =oscuro, así como D representa la grosor, con  $d_1$ =grueso y  $d_2$ =fino.

# Red Bayesiana para el problema del salmon

8



- La estación y el área geográfica de pesca no dependen del pez, pero si dependen las variables luminosidad y grosor.
- $P(a_j) = 0.25$ ,  $P(b_j) = 0.5$
- Un experto genera las siguientes probabilidades:

$P(x_i a_j)$	salmon	mero
Invierno	.9	.1
Primavera	.3	.7
Verano	.4	.6
Otoño	.8	.2

$P(x_i b_j)$	salmon	mero
Norte	.65	.35
Primavera	.95	.05

- El experto tambien genera probabilidades condicionales para los eventos conjuntos y las variables hijos

10

$P(x_i a_i b_j)$	Norte	Sur
Invierno	.9	.8
Primavera	.3	.2
Verano	.4	.1
Otoño	.7	.6

$P(c_i x_j)$	salmon	mero
Claro	.33	.8
Medio	.33	.1
Oscuro	.34	.1

$P(d_i x_j)$	salmon	mero
grueso	.4	.95
fino	.6	.05

# Probabilidades conjuntas

11

- Con estas tablas, la probabilidad de que un pez oscuro y fino recogido en verano en el Atlántico norte sea mero es

$$\begin{aligned} P(x_2, a_3, b_1, c_3, d_2) &= P(a_3)P(b_1)P(x_2|a_3, b_1)P(c_3|x_2)P(d_2|x_2) \\ &= (0.25)(0.6)(0.6)(0.5)(0.4) = 0.018 \end{aligned}$$

- Usando estas informaciones podríamos clasificar un nuevo pez con la clase que produzca el mínimo error de clasificación

# Max( $P(x_1)$ , $P(x_2)$ )

12

- Si uso la red para calcular las probabilidades del Mero y el Salmón resulta

$$\begin{aligned} P(x_1) &= \sum_{i,j,k,l} P(x_1, a_i, b_j, c_k, d_l) \\ &= \sum_{i,j,k,l} P(a_i)P(b_j)P(x_1 | a_i, b_j)P(c_k | x_1)P(d_l | x_1) \\ &= \sum_{i,j} P(a_i)P(b_j)P(x_1 | a_i, b_j) \\ &= (0.25)(0.5) \sum_{i,j} P(x_1 | a_i, b_j) \\ &= (0.25)(0.5)[0.9 + 0.3 + 0.4 + 0.7 + 0.8 + 0.2 + 0.1 + 0.6] \\ &= 0.5 \end{aligned}$$

# Necesito Evidencia

13

- Si recolectamos evidencia,  $\{e_A, e_B, e_C, e_D\}$ , asumiendo que son independientes unos de los otros entonces podemos clasificar el pez en la clase con mayor probabilidad dada la evidencia .
  - $Max(P(x_1|e) , P(x_2|e))$
- Usamos la red para calcular dichas probabilidades, sabiendo que pescamos en invierno un pez mas o menos claro, no estamos seguros si se pesco en el sur y no tenemos datos del ancho.

## Ejemplo $\{e_A, e_B, e_C, e_D\}$ ,

14

- Supongamos que es invierno, i.e,  $P(a_1|e_A) = 1$  y  $P(a_i|e_A) = 0$  para  $i = 2, 3, 4$ .
- Supongamos que no sabemos de que área vino el bote, pero que una cierta tripulación le gusta pescar al sur ; por lo cual asumimos que  $P(b_1|e_B) = 0.2$  y  $P(b_2|e_B) = 0.8$ .
- Medimos el pez y encontramos que es bastante claro y decidimos  $P(e_C|c_1) = 1$ ,  $P(e_C|c_2) = 0.5$ , y  $P(e_C|c_3) = 0$ .
- Supongamos que existe oclusión y que no podemos medir el ancho del pez, por lo cual  $P(e_D|d_1) = P(e_D|d_2)$ .

# Salmon dado evidencia

15

- $P(x_i|e) \propto P_P(x_i) P_c(x_i)$ 
  - $P_P(x_i)=P(x_i |e^P)$  la evidencia de los padres
  - $P_c(x_i)= P(e^C |x_i)$  la evidencia de los hijos
- $P(e^C |x_i)$  es simple de derivar, es el producto de las evidencias condicional a la etiqueta
- $P(x_i |e^P)$  es mucho mas complicada porque involucra todas las combinaciones de condicionantes.

# Padres e hijos

16

- $P(x_i | e^P)$  probabilidad de salmón producida por sus padres

$$\begin{aligned} P_P(x_i) &= P(x_i | e^P) \\ &\propto P(x_i | a_1, b_1)P(a_1 | e)P(b_1 | e) + P(x_i | a_1, b_2)P(a_1 | e)P(b_2 | e) \\ &\quad + P(x_i | a_2, b_1)P(a_2 | e)P(b_1 | e) + P(x_i | a_2, b_2)P(a_2 | e)P(b_2 | e) \\ &\quad + P(x_i | a_3, b_1)P(a_3 | e)P(b_1 | e) + P(x_i | a_3, b_2)P(a_3 | e)P(b_2 | e) \\ &\quad + P(x_i | a_4, b_1)P(a_4 | e)P(b_1 | e) + P(x_i | a_4, b_2)P(a_4 | e)P(b_2 | e) \\ &= P(x_i | a_1, b_1)P(a_1 | e)P(b_1 | e) + P(x_i | a_1, b_2)P(a_1 | e)P(b_2 | e) \end{aligned}$$

$$P_P(x_1) = 0.82, P_P(x_2) = 1 - 0.82 = 0.18$$

# Calculamos ahora $P_C(x_1)$

17

- $P(e^c | x_i)$  probabilidad de salmón producida por sus hijos

$$\begin{aligned} P_c(x_1) &= P(e^c | x_1) = P(e_c | x_1)P(e_D | x_1) = \\ &= [P(e_c | c_1)P(c_1 | x_1) + P(e_c | c_2)P(c_2 | x_2) + P(e_c | c_3)P(c_3 | x_3)] \\ &\times [P(e_D | d_1)P(d_1 | x_1) + P(e_D | d_2)P(d_2 | x_1)] \\ &= [(1.0)(0.33 + (0.5)(0.33) + 0(.34))] \times [1.0(0.4) + (1.0)(0.6)] \\ &= 0.495 \end{aligned}$$

- Un calculo similar da  $P_C(x_2) \propto 0.85$ .

# Calculamos ahora $P(x_1|e)$

18

$$P(x_i|e) \propto P_c(x_i)P_P(x_i)$$

- Re-normalizando (i.e, dividiendo por su suma), resulta

$$P(x_1 | e) = \frac{(0.82)(0.495)}{(0.82)(0.495) + (0.18)(0.85)} = 0.726$$

$$P(x_2 | e) = \frac{(0.18)(0.85)}{(0.82)(0.495) + (0.18)(0.85)} = 0.274$$

# Otro ejemplo más simple

19

- Supongamos que un pez capturado en el Atlántico sur ( $b_2$ ) es claro ( $c_1$ ), pero no sabemos que tiempo del año fue capturado o que ancho tiene.
- Por lo cual podemos focalizar directamente en la evidencia que tenemos para clasificar

$$\begin{aligned} P(x_1|c_1, b_2) &= \frac{P(x_1, c_1, b_2)}{P(c_1, b_2)} = \alpha \left[ \sum_{a,d} P(a)P(b_2)P(x_1|a, b_2)P(c_1|x_1)P(d|x_1) \right] \\ &= \alpha P(b_2)P(c_1|x_1) \left[ \sum_a P(a)P(x_1|a, b_2) \right] \underbrace{\left[ \sum_d P(d|x_1) \right]}_{=1} = \\ &= \alpha P(b_2)P(c_1|x_1) [P(a_1)P(x_1|a_1, b_2) + P(a_2)P(x_1|a_2, b_2) + P(a_{13})P(x_1|a_3, b_2) \\ &\quad + P(a_4)P(x_1|a_4, b_2)] \\ &= \alpha(0.4)(0.6) [(0.25)(0.7) + (0.25)(0.8) + (0.25)(0.1) + (0.25)(0.3)]1.0 = \alpha(0.114) \end{aligned}$$

- Observemos que
  - $[\sum_d P(d| x_1)] = 1$  , esto es si no hay información sobre el nodo D la tabla de condicionales no afecta el resultado
- $P(x_2|c_1, b_2) = \alpha(0.042)$  realizando una cuenta parecida a la anterior.
- Normalizando (dividiendo por  $P(c_1, b_2)$ ) resulta  $P(x_1|c_1, b_2) = 0.73$   $P(x_2|c_1, b_2) = 0.27$  por lo cual dada esta evidencia el pez se clasifica como salmón.

# Dependencias desconocidas

21

- Cuando las dependencias son desconocidas, podemos suponer la menor de las hipótesis, la independencia condicional dada la categoría

$$p(\omega_k | x_1, x_2, \dots, x_n) \propto \prod_{i=1}^d p(x_i | \omega_k) p(\omega_k)$$

- Este es la llamada regla ingenua (naïve) de Bayes, que a menudo tiene mucho éxito en la práctica, y que puede se expresada como una red muy simple.
- Clasifica el punto  $x$  en la clase que maximiza la probabilidad anterior.

# Ejemplo naïve bayes

22

- Las probabilidades condicionales pueden ser discretas o continuas
  - Gaussian naïve Bayes
  - Multinomial naïve Bayes