



IIC 2433 Minería de Datos

<https://github.com/marcelomendoza/IIC2433>

- Redes Bayesianas -

Repaso de conceptos: Distribución conjunta

- Intelligence (I) ← 2
 - i^0 (low), i^1 (high),
 - Difficulty (D) ← 2
 - d^0 (easy), d^1 (hard)
 - Grade (G) ← 3
 - g^1 (A), g^2 (B), g^3 (C)
- parameters*
 $2 \times 2 \times 3 = 12$
independent params
!!

$P(I, D, G)$

I	D	G	Prob.
i^0	d^0	g^1	0.126
i^0	d^0	g^2	0.168
i^0	d^0	g^3	0.126
i^0	d^1	g^1	0.009
i^0	d^1	g^2	0.045
i^0	d^1	g^3	0.126
i^1	d^0	g^1	0.252
i^1	d^0	g^2	0.0224
i^1	d^0	g^3	0.0056
i^1	d^1	g^1	0.06
i^1	d^1	g^2	0.036
i^1	d^1	g^3	0.024

1

Repaso de conceptos: Probabilidad condicional

condition on g^1


I	D	G	Prob.
i^0	d^0	g^1	0.126
i^0	d^0	g^2	0.168
i^0	d^0	g^3	0.126
i^0	d^1	g^1	0.009
i^0	d^1	g^2	0.045
i^0	d^1	g^3	0.126
i^1	d^0	g^1	0.252
i^1	d^0	g^2	0.0224
i^1	d^0	g^3	0.0056
i^1	d^1	g^1	0.06
i^1	d^1	g^2	0.036
i^1	d^1	g^3	0.024

Repaso de conceptos: Marginalización de variable

$P(I, D)$ Marginalize I

I	D	Prob.
i^0	d^0	0.282
i^0	d^1	0.02
i^1	d^0	0.564
i^1	d^1	0.134

D	Prob.
d^0	0.846
d^1	0.154



Repaso de conceptos: Distribución de probabilidad condicional

Supongamos que G tiene tres valores posibles (A , B y C).

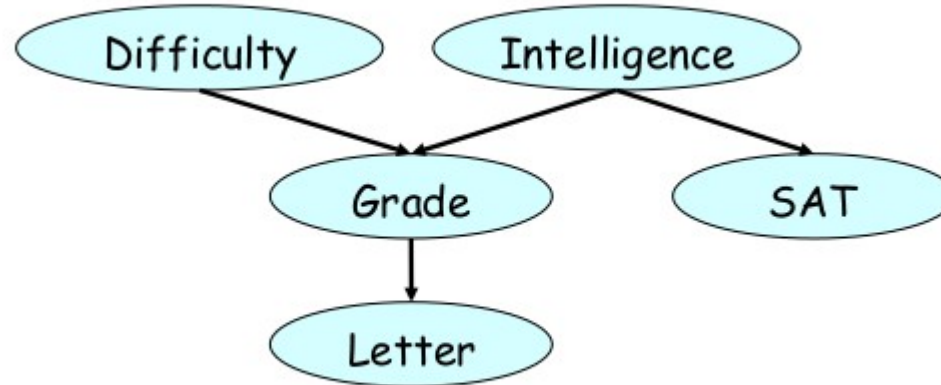
$P(G \mid \underline{I, D})$

	g^1	g^2	g^3
i^0, d^0	0.3	0.4	0.3
i^0, d^1	0.05	0.25	0.7
i^1, d^0	0.9	0.08	0.02
i^1, d^1	0.5	0.3	0.2
	<u>A</u>	<u>B</u>	<u>C</u>

total →

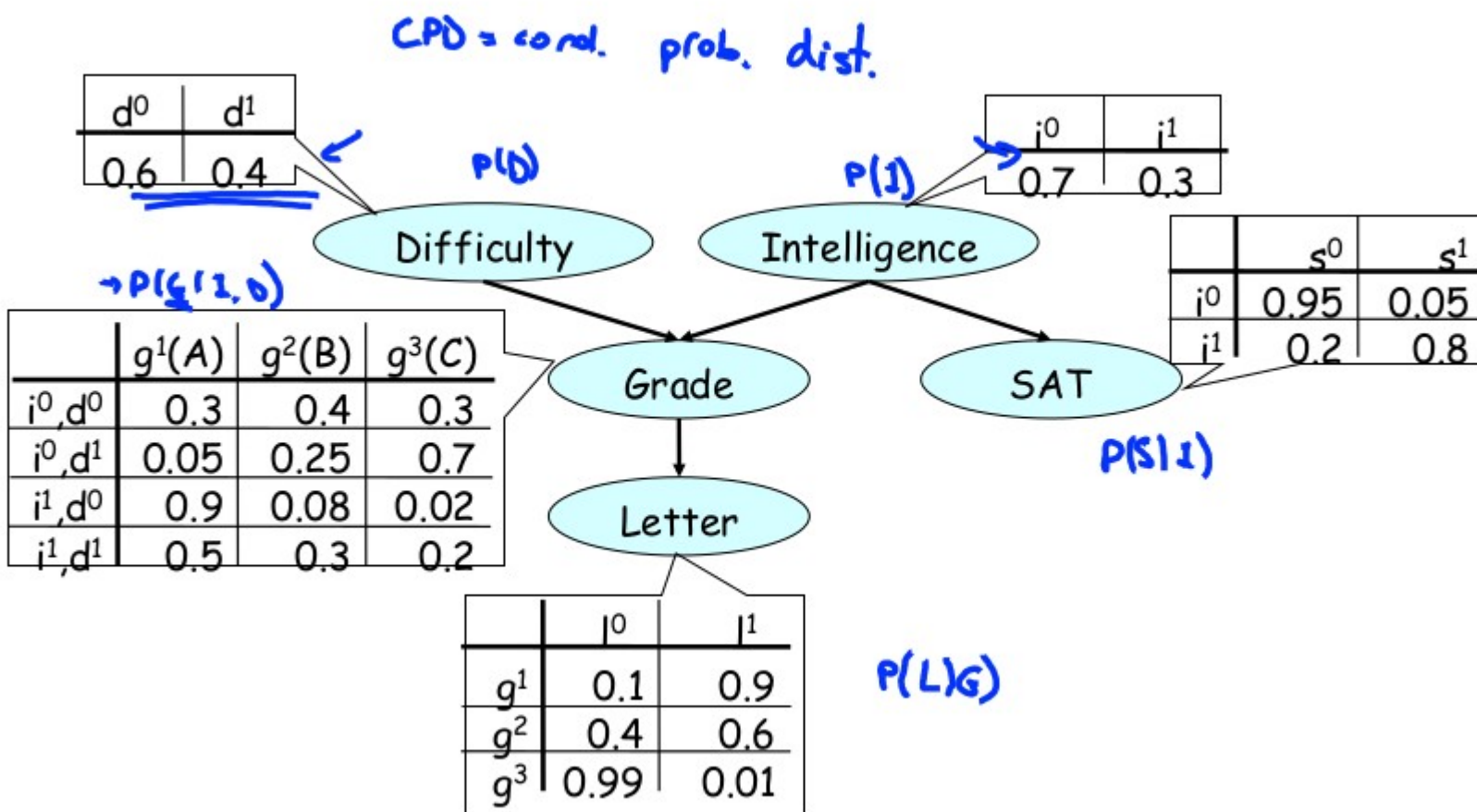
Red Bayesiana

Una red Bayesiana es un modelo que representa las relaciones de dependencia entre variables categóricas.



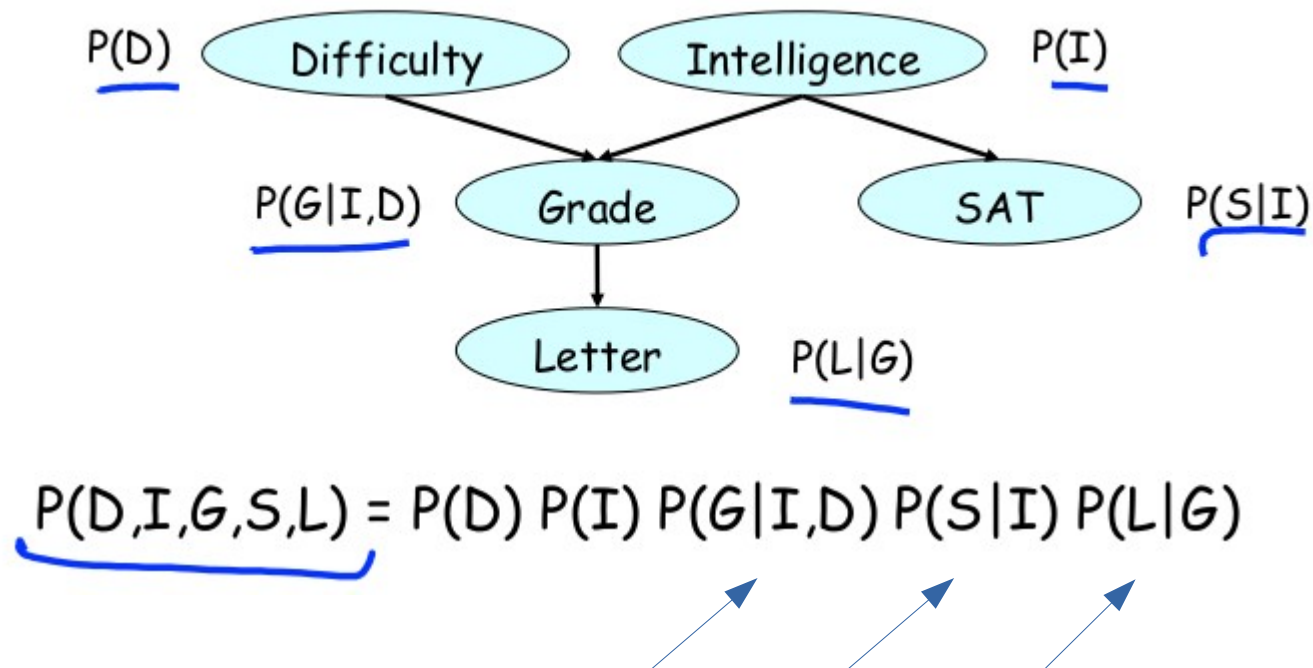
Red Bayesiana

La red incluirá las probabilidades condicionales.



Red Bayesiana

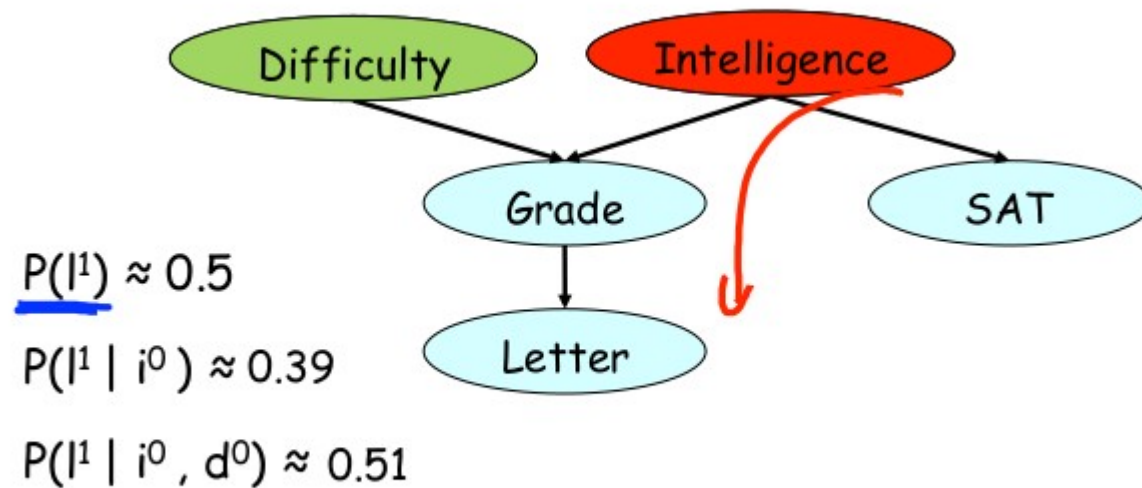
Finalmente el modelo es una representación gráfica de la distribución conjunta.



Cada CPD corresponde a un factor de la red.

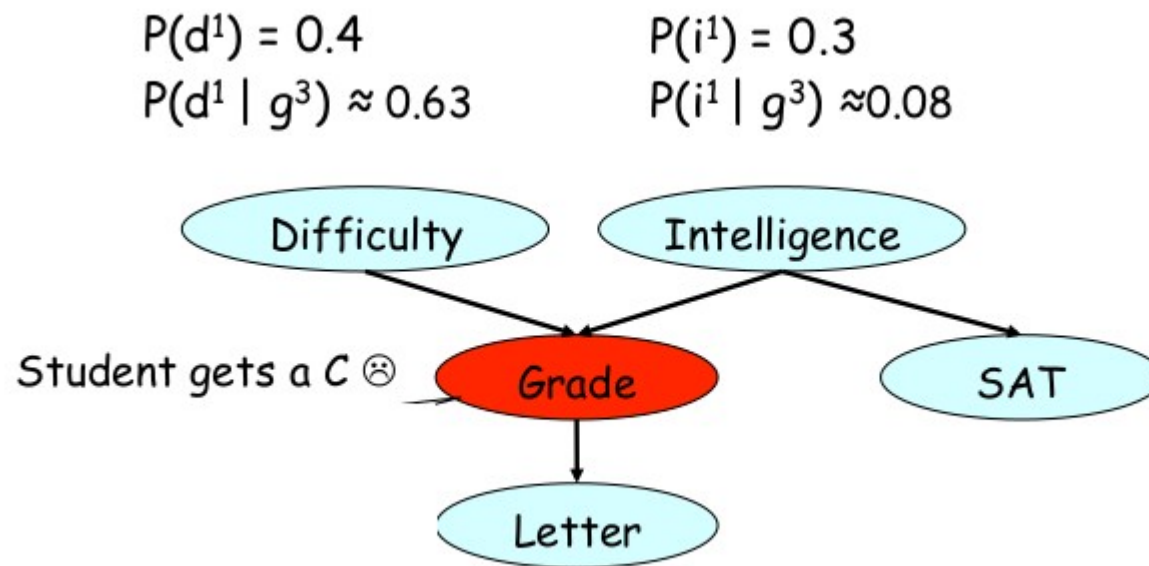
Red Bayesiana

El modelo es útil para construir una explicación causal.



Red Bayesiana

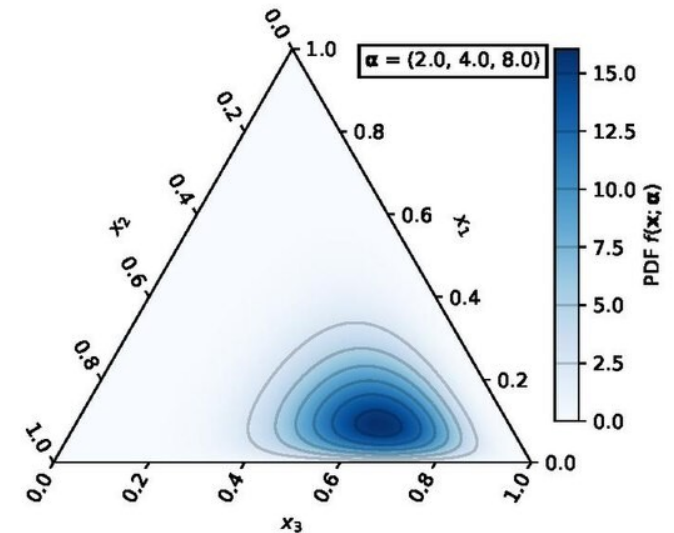
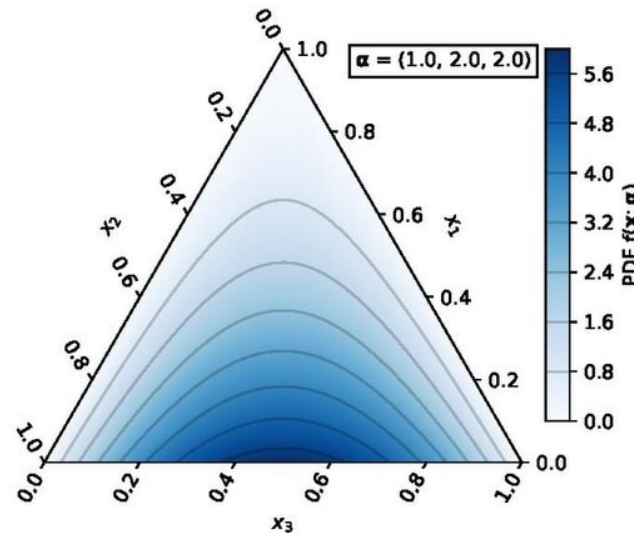
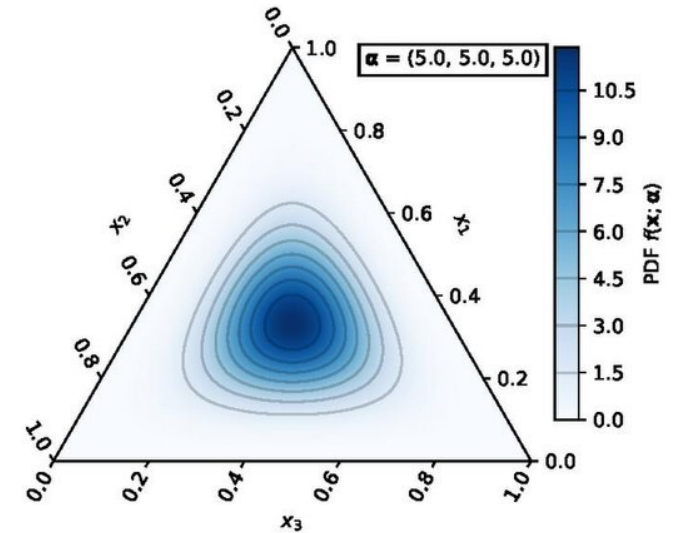
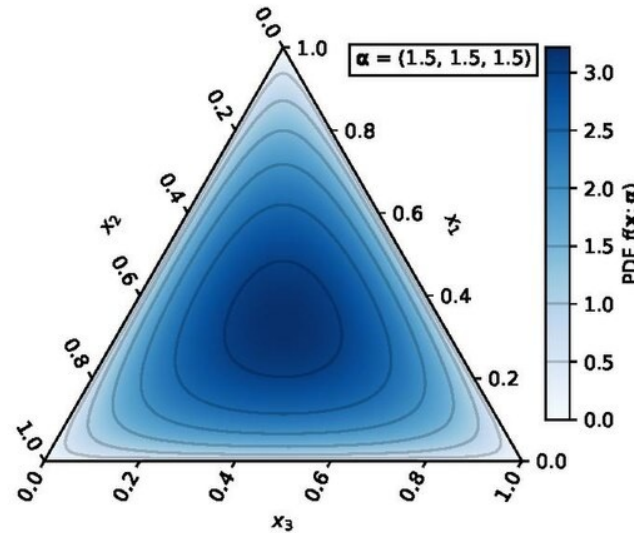
El modelo es útil para razonamiento basado en evidencia.



Inferencia de parámetros en base a MLE

Si hay pocos datos, usaremos priors de Dirichlet.

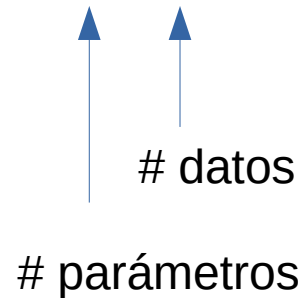
Card=3 para distintos
parámetros de Dirichlet.



Selección de modelos con BIC

El Bayesian Information Criterion (BIC) permite seleccionar modelos de redes Bayesianas. Se calcula según:

$$\text{BIC} = k \ln(n) - 2 \ln(\text{MLE})$$



Mientras más cerca estemos del 0, mejor.

Structural learning

Si ajustamos una red Bayesiana desde datos, no sabremos las dependencias entre las variables. Necesitamos entonces seleccionar modelos en base a las posibles relaciones de dependencia:

Buscaremos según:

- Búsqueda exhaustiva.
- Heurística de búsqueda (Hill climbing).

Seleccionamos según BIC.