



# IIC 2433 Minería de Datos

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

## Cierre de la clase 7 – Asociaciones



### **Asociaciones:**

¿Cuál es la diferencia entre soporte y confianza de una regla?

¿Qué es lo que hace que una regla sea interesante?

### **Algoritmo Apriori:**

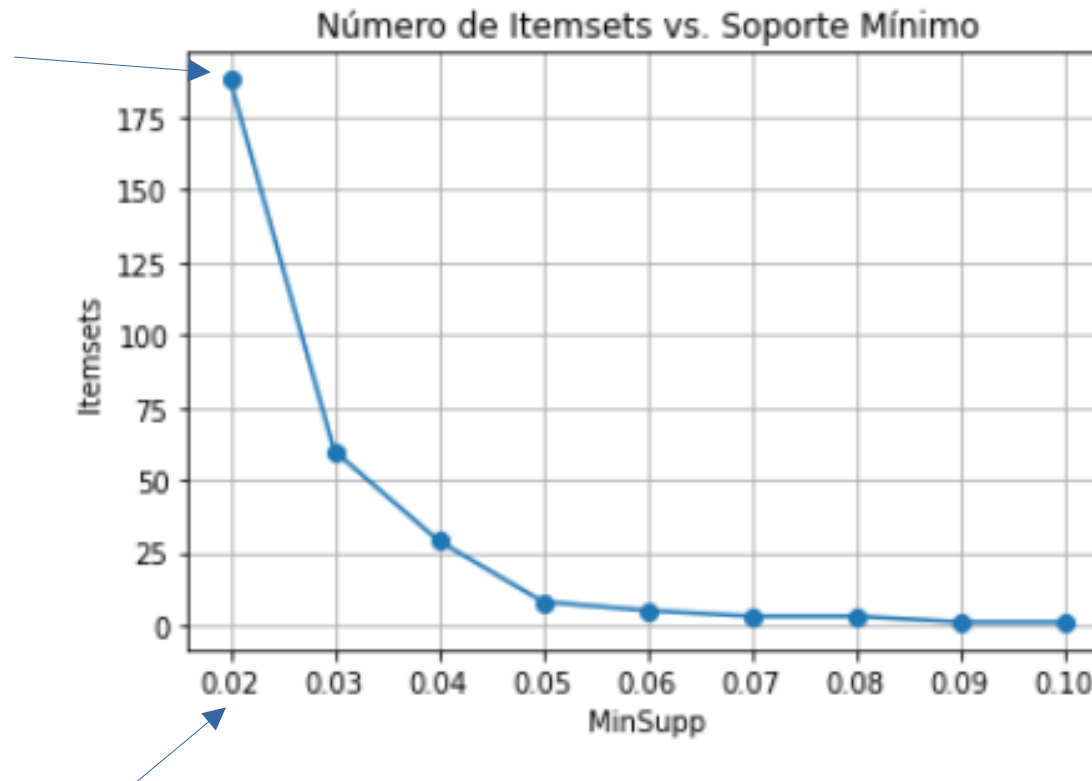
¿Qué indica el principio de mononicidad?

¿Cómo lo aplicamos a la búsqueda de itemsets frecuentes?

¿Cómo lo aplicamos a la búsqueda de reglas interesantes?

## Cierre de la clase 7 – Actividad formativa

¿Cómo encontramos reglas interesantes?



Para encontrar reglas interesantes es necesario **identificar un conjunto amplio de itemsets frecuentes** donde buscar.

Luego, en ese conjunto, aplicamos una condición de confianza alta, por ejemplo, confianza  $> 0.5$ .

## Cierre de la clase 7 – Actividad formativa

¿Cómo encontramos reglas interesantes?

Para encontrar reglas interesantes es necesario **identificar un conjunto amplio de itemsets frecuentes** donde buscar.

Luego, en ese conjunto, aplicamos una condición de confianza alta, por ejemplo, confianza  $> 0.5$ .

```
itemsets = apriori(df, min_support = 0.02, use_colnames = True, verbose = 1)
```

Processing 705 combinations | Sampling itemset size 3 2

```
df_ar = association_rules(itemsets, metric = "confidence", min_threshold = 0.6)
df_ar[['antecedents', 'consequents', 'support', 'confidence']]
```

	antecedents	consequents	support	confidence
0	(22356)	(20724)	0.02162	0.676048
1	(21231)	(21232)	0.02264	0.740353
2	(21733)	(85123A)	0.02486	0.697140
3	(22386)	(85099B)	0.02860	0.633304
4	(22699)	(22697)	0.02080	0.724739
5	(22697)	(22699)	0.02080	0.756364
6	(85099F)	(85099B)	0.02378	0.643050

## - Redes Bayesianas -

*Un modelo basado en dependencias entre variables*

# 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
<del><math>i^0</math></del>	<del><math>d^0</math></del>	<del><math>g^2</math></del>	<del>0.168</del>
<del><math>i^0</math></del>	<del><math>d^0</math></del>	<del><math>g^3</math></del>	<del>0.126</del>
$i^0$	$d^1$	$g^1$	0.009
<del><math>i^0</math></del>	<del><math>d^1</math></del>	<del><math>g^2</math></del>	<del>0.045</del>
<del><math>i^0</math></del>	<del><math>d^1</math></del>	<del><math>g^3</math></del>	<del>0.126</del>
$i^1$	$d^0$	$g^1$	0.252
<del><math>i^1</math></del>	<del><math>d^0</math></del>	<del><math>g^2</math></del>	<del>0.0224</del>
<del><math>i^1</math></del>	<del><math>d^0</math></del>	<del><math>g^3</math></del>	<del>0.0056</del>
$i^1$	$d^1$	$g^1$	0.06
<del><math>i^1</math></del>	<del><math>d^1</math></del>	<del><math>g^2</math></del>	<del>0.036</del>
<del><math>i^1</math></del>	<del><math>d^1</math></del>	<del><math>g^3</math></del>	<del>0.024</del>

## 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).  
Si la fila representa em condicionante, debe sumar 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

$= 0.42$

$0.126 / 0.42$

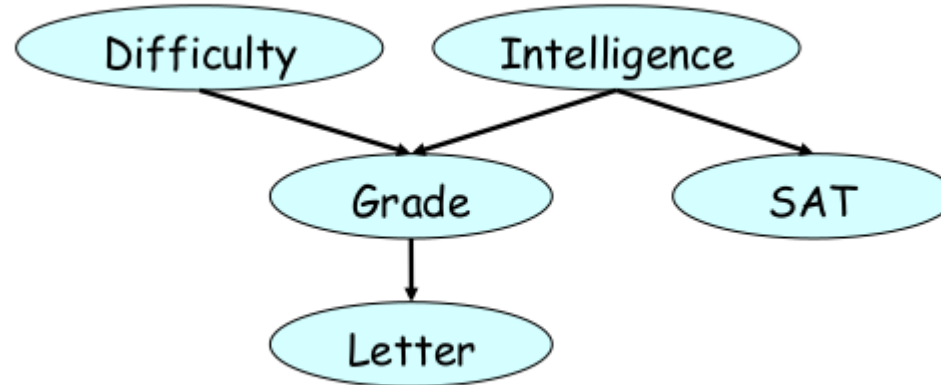
$P(G \mid I, D)$

	$g^1$	$g^2$	$g^3$
$i^0, d^0$	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

A      B      C

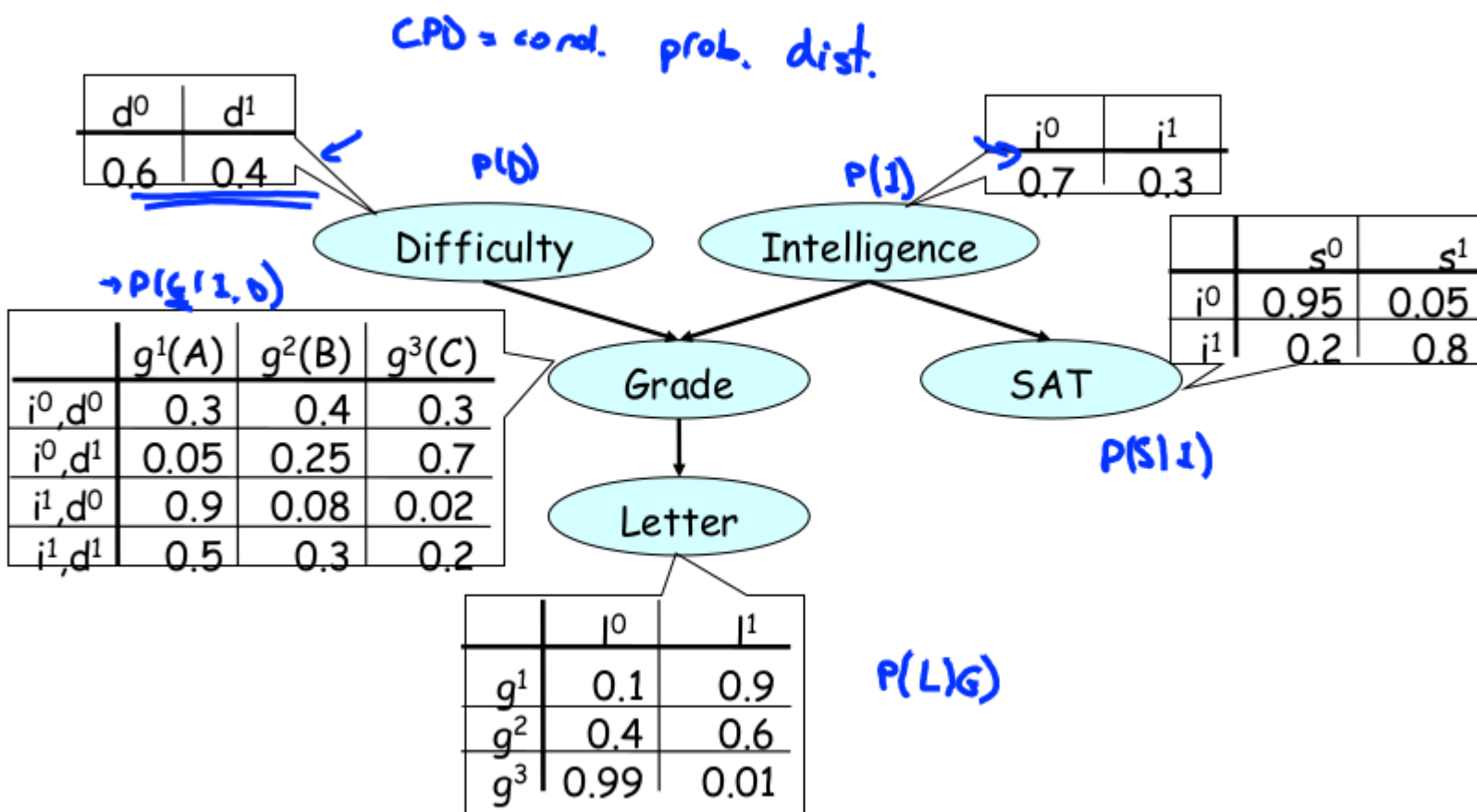
# Red Bayesiana

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



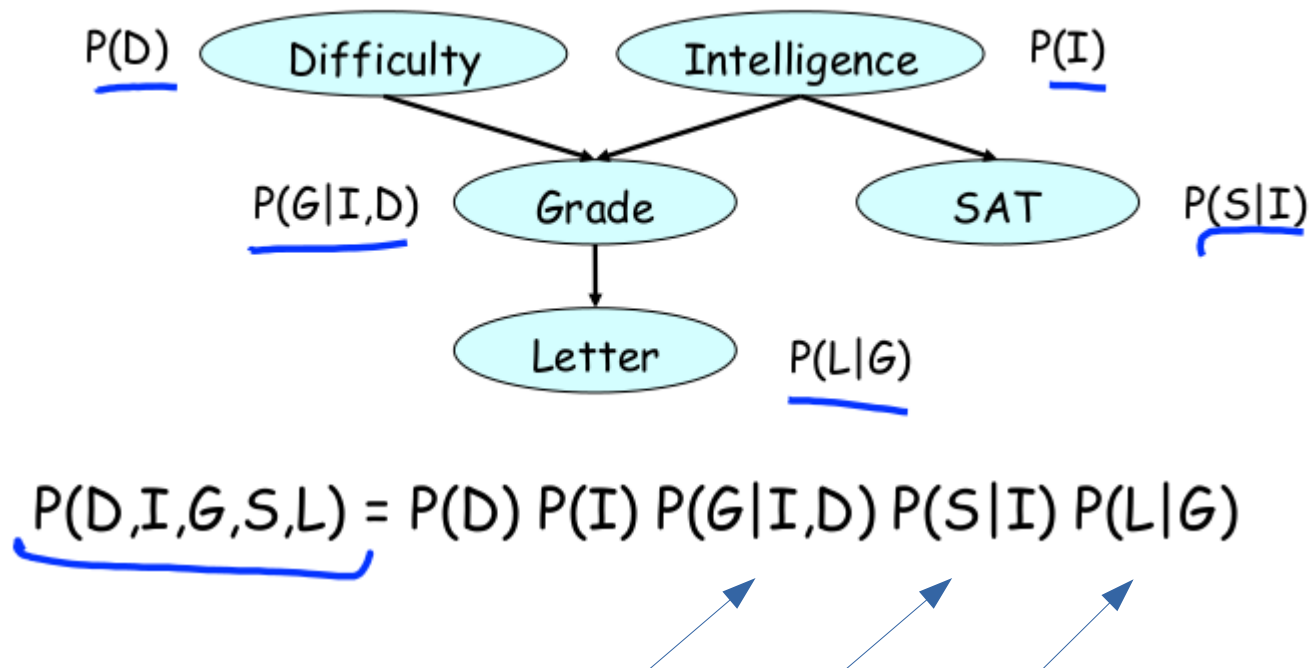
# Red Bayesiana

La red incluirá las probabilidades condicionales.



# Red Bayesiana

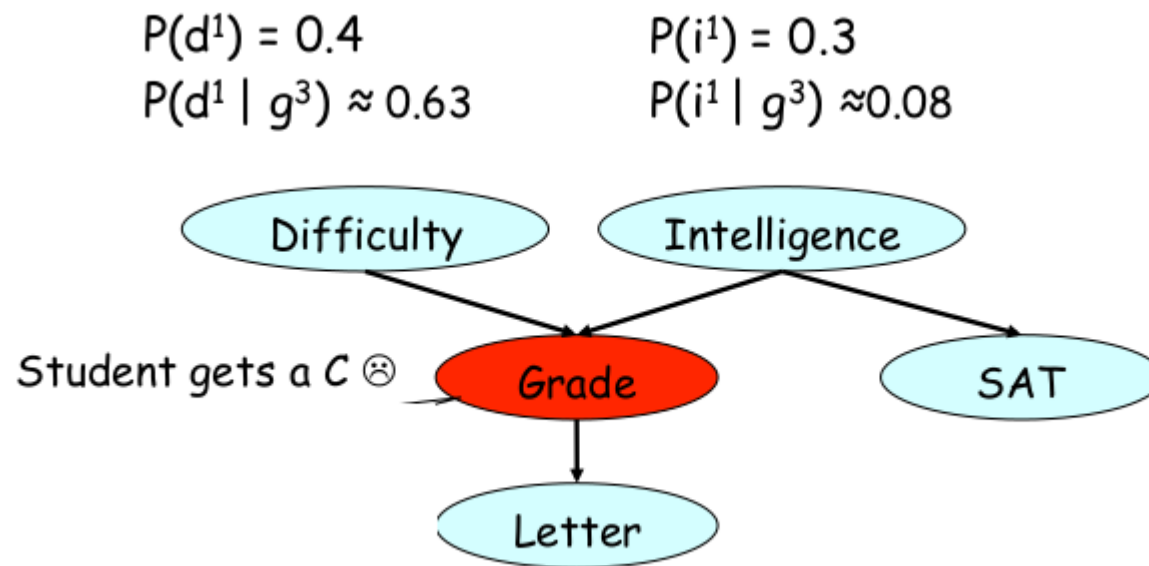
Finalmente el modelo es una representación gráfica de la distribución conjunta.



Cada CPD corresponde a un factor de la red.

# Red Bayesiana

El modelo es útil para razonamiento basado en evidencia.



## Aprendizaje estructural de la red

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

Buscaremos según:

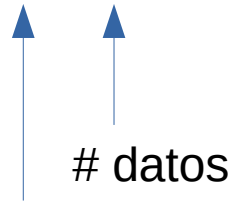
- Búsqueda exhaustiva.
- Heurística de búsqueda (Hill climbing).

Seleccionamos la red según un criterio.

# Selección de modelos con BIC

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

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



# datos

# parámetros (filas de las condicionales)

Mientras más cerca estemos del 0, mejor.