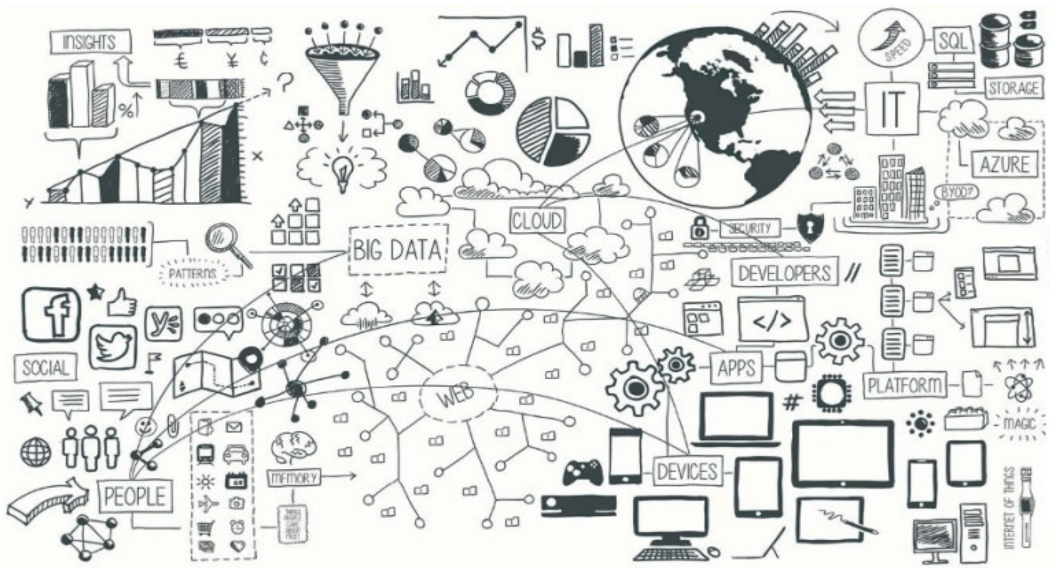
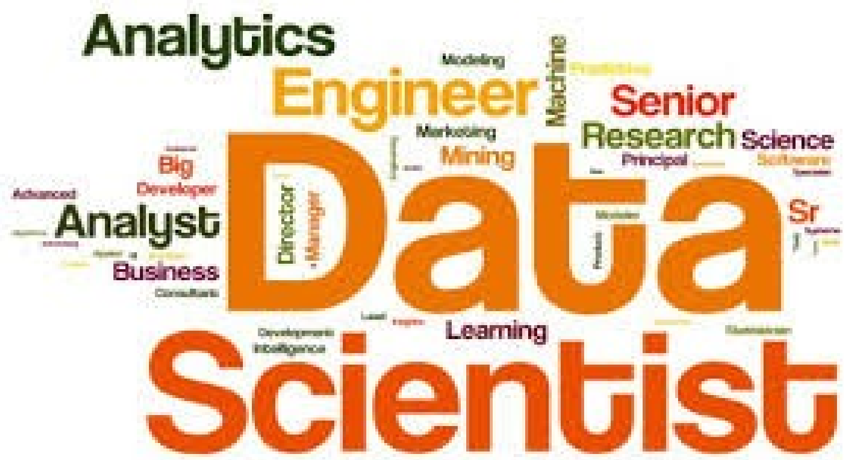
M1970 – Machine Learning II
Redes Probabilísticas Discretas





Sixto Herrera (sixto.herrera@unican.es) y Mikel Legasa

Grupo de Meteorología

Univ. de Cantabria – CSIC
MACC / IFCA

M1970 – Machine Learning (L 15:30-17:30; X 15:30-17:30)

Feb	28	L	Redes Probabilísticas Discretas (2h-T)
Mar	2	X	Redes Bayesianas: Creación e Inferencia (2h-L)
	7	L	Clasificacidores Bayesianos. Naive Bayes (2h-L)
	9	X	Redes Bayesianas: Aprendizaje Estructural (2h-T)
	14	L	Redes Bayesianas: Aprendizaje Paramétrico (2h-LT)
	16	X	Redes Bayesianas: Aprendizaje (2h-L)
	21	L	Evaluación (2h)

NOTA: Las líneas de código de R en esta presentación se muestran sobre un fondo gris.

7. MÉTODOS DE LA EVALUACIÓN				
Descripción	Tipología	Eval. Final	Recuper.	%
Valoración de informes y trabajos escritos	Actividad de evaluación con soporte virtual	Sí	Sí	60,00

Calif. mínima	3,00
Duración	
Fecha realización	Durante el periodo de impartición de la asignatura.
Condiciones recuperación	
Observaciones	Evaluación de los trabajos de grupo e individuales entregados por el alumno.

Descripción	Tipología	Eval. Final	Recuper.	%
Examen (escrito, oral y/o práctico en el aula de computación)	Actividad de evaluación con soporte virtual	Sí	Sí	40,00

Calif. mínima	0,00
Duración	Un máximo de dos horas
Fecha realización	Durante el periodo de impartición de la asignatura.
Condiciones recuperación	
Observaciones	

TOTAL	100,00

Observaciones

Si la nota final del alumno fuese menor que 5 sobre 10, entonces la recuperación consistirá en la realización de cada una de las tareas en las que hubiera obtenido una calificación menor que 5 sobre 10. El procedimiento de evaluación de una actividad recuperable será equivalente al de la actividad original.

Observaciones para alumnos a tiempo parcial

Examen tipo test a desarrollar en el aula a través de la plataforma Moodle

El examen incluirá cuestiones de ambas partes de la asignatura. 21 de Marzo

7. MÉTODOS DE LA EVALUACIÓN

Descripción	Tipología	Eval. Final	Recuper.	%
Valoración de informes y trabajos escritos	Actividad de evaluación con soporte virtual	Sí	Sí	60,00

Calif. mínima	3,00
Duración	
Fecha realización	Durante el periodo de impartición de la asignatura.
Condiciones recuperación	
Observaciones	Evaluación de los trabajos de grupo e individuales entregados por el alumno.

Descripción	Tipología	Eval. Final	Recuper.	%
Examen (escrito, oral y/o práctico en el aula de computación)	Actividad de evaluación con soporte virtual	Sí	Sí	40,00

Calif. mínima	0,00
Duración	Un máximo de dos horas
Fecha realización	Durante el periodo de impartición de la asignatura.
Condiciones recuperación	
Observaciones	

TOTAL	100,00

T01 – Redes Bayesianas

Observaciones

Si la nota final del alumno fuese menor que 5 sobre 10, entonces la recuperación consistirá en la realización de cada una de las tareas en las que hubiera obtenido una calificación menor que 5 sobre 10. El procedimiento de evaluación de una actividad recuperable será equivalente al de la actividad original.

Observaciones para alumnos a tiempo parcial

A nivel global, el valor de esta tarea se corresponde con el 30% de la nota final

M1970 – Machine Learning (L 15:30-17:30; X 15:30-17:30)

Feb	28	L	Redes Probabilísticas Discretas (2h-T)
Mar	2	X	Redes Bayesianas: Creación e Inferencia (2h-L)
	7	L	Clasificacidores Bayesianos. Naive Bayes (2h-L)
	9	X	Redes Bayesianas: Aprendizaje Estructural (2h-T)
	14	L	Redes Bayesianas: Aprendizaje Paramétrico (2h-LT)
	16	X	Redes Bayesianas: Aprendizaje (2h-L)
	21	L	Evaluación (2h)

T01 – Redes Bayesianas

NOTA: Las líneas de código de R en esta presentación se muestran sobre un fondo gris.

Con carácter obligatorio todas las tareas se realizarán o entregarán usando la plataforma virtual de la asignatura. Por tanto es responsabilidad del alumno, asegurarse de que puede acceder a la plataforma virtual de la asignatura, antes del comienzo de las sesiones en las que se realicen las pruebas.
Todas las entregas deberán incluir el **nombre y apellidos** del alumno que realiza la entrega.

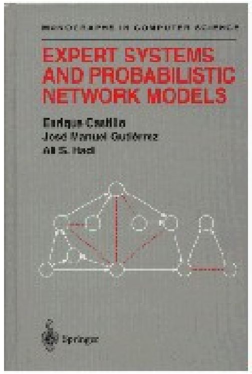
La plataforma usada es Moodle y podéis acceder a ella usando el Aula Virtual de la Universidad de Cantabria. Para ello es imprescindible vuestro usuario y contraseña.

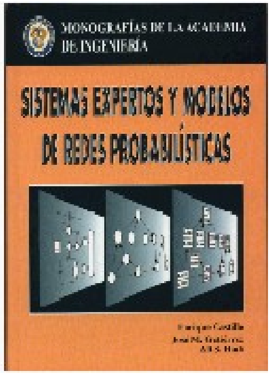
Si fuese necesario comunicarse mediante correo electrónico con el profesorado, es obligatorio usar el correo con el cual se ha inscrito en Moodle.

Para cualquier problema con vuestro correo, poneros en contacto con el Servicio de Informática de la Universidad de Cantabria.

Cualquier duda, anuncio o comentario sobre el desarrollo y actividades de la asignatura se hará mediante el Foro disponible en el Moodle.
Muchas veces las dudas también las puede tener otro compañero y de este modo toda la información relevante sobre el desarrollo de la asignatura estará a disposición de vuestros compañeros.
El Foro no se usa para evaluaros. Es una herramienta que os permite compartir información con vuestros compañeros.
El Foro no está moderado, pero sí supervisado por el profesorado, de tal forma que se resuelvan o aclaren dudas si ningún compañero vuestro lo hace.

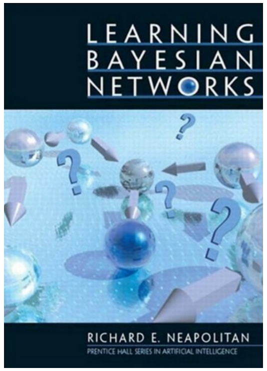
Sólo se permite el uso de mensajes personales o correo electrónico para situaciones personales muy excepcionales.



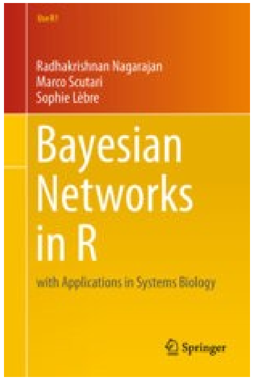


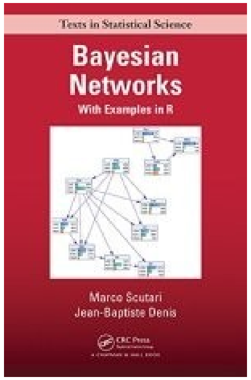
Expert Systems and Probabilistic Network Models.
E. Castillo, J.M. Gutiérrez, y A.S. Hadi
Springer-Verlag, New York.

Monografías de la Academia Española de Ingeniería



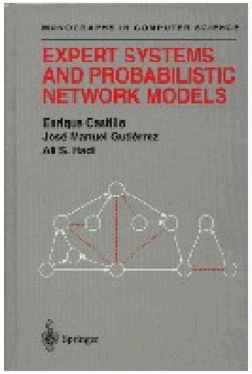
Richard E. Neapolitan
Learning Bayesian Networks
Prentice Hall Series in Artificial Intelligence
Pearson Prentice Hall, 2004 ~700

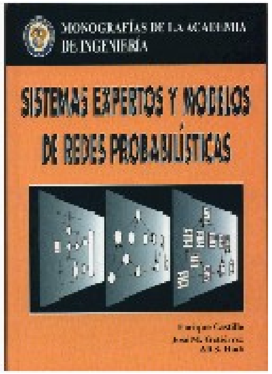




Marco Scutari: *Bayesian networks in R & Bayesian networks with examples in R*

http://www.bnlearn.com/





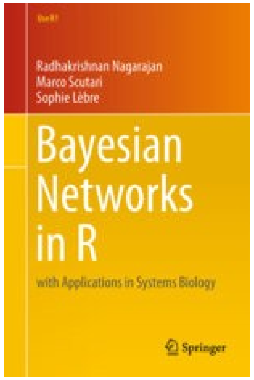
Expert Systems and Probabilistic Network Models.
E. Castillo, J.M. Gutiérrez, y A.S. Hadi
Springer-Verlag, New York.

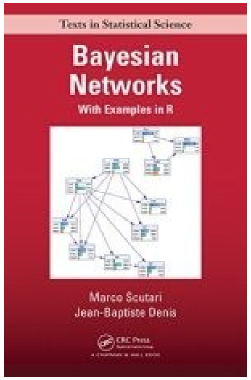
Monografías de la Academia Española de Ingeniería

Una aplicación en Ciencias Atmosféricas

LIBRO

J.M. Gutiérrez, R. Cano, A.S. Cofiño, and C. Sordo
Redes Probabilísticas y Neuronales en las Ciencias Atmosféricas
Ministerio de Medio Ambiente (*Monografías del Instituto Nacional de Meteorología*), Madrid. 350 páginas, 2004





Marco Scutari: *Bayesian networks in R & Bayesian networks with examples in R*

http://www.bnlearn.com/

1.1 Dataset de ejemplo: 'survey'

A partir de los datos de campo recogidos por la encuesta, se investigará la selección de medios de transporte por distintos perfiles de usuarios, y particularmente a la preferencia de tren o coche. Este tipo de análisis se utilizan con frecuencia en la planificación de infraestructuras. Para cada individuo encuestado, se han recopilado datos referenctes a 6 variables discretas. Las abreviaturas de dichas variables se muestran entre paréntesis, y se utilizarán a lo largo de la práctica para referirse a los nodos de la red creada. Tanto las abreviaturas como los nombres de las variables preservan la nomenclatura original del dataset en inglés.

- Edad (`A`): Edad del encuestado, agrupado en los siguientes estados: joven (`young` , < 30 años), adulto (`adult` , 30 < edad <= 60) y anciano (`old` , edad > 60).
- Sexo (`S`): Sexo del encuestado, con sus dos posibles estado: masculino (`M`) y femenino (`F`).
- Educación (`E`): Nivel más alto de educación alcanzado. Hasta educación secundaria (`high`) o título universitario (`uni`).
- Ocupación (`O`): Considera dos estados: trabajador por cuenta ajena (`emp`) o autónomo (`self`).
- Residencia (`R`): El tamaño de la población de residencia del individuo. Estados posibles: `big` y `small` .
- Transporte (`T`): El medio de transporte más utilizado por el encuestado para acudir al trabajo, diferenciando 3 posibles estados: `car` , `train` y `other` .

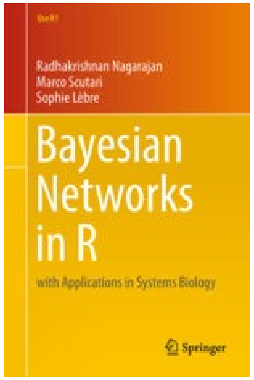
CATEGÓRICO

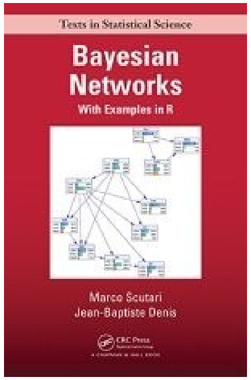
1.1 Dataset de ejemplo: 'survey'

A partir de los datos de campo recogidos por la encuesta, se investigará la selección de medios de transporte por distintos perfiles de usuarios, y particularmente a la preferencia de tren o coche. Este tipo de análisis se utilizan con frecuencia en la planificación de infraestructuras. Para cada individuo encuestado, se han recopilado datos referenctes a 6 variables discretas. Las abreviaturas de dichas variables se muestran entre paréntesis, y se utilizarán a lo largo de la práctica para referirse a los nodos de la red creada. Tanto las abreviaturas como los nombres de las variables preservan la nomenclatura original del dataset en inglés.

- Edad (`A`): Edad del encuestado, agrupado en los siguientes estados: joven (`young` , < 30 años), adulto (`adult` , 30 < edad <= 60) y anciano (`old` , edad > 60).
- Sexo (`S`): Sexo del encuestado, con sus dos posibles estado: masculino (`M`) y femenino (`F`).
- Educación (`E`): Nivel más alto de educación alcanzado. Hasta educación secundaria (`high`) o título universitario (`uni`).
- Ocupación (`O`): Considera dos estados: trabajador por cuenta ajena (`emp`) o autónomo (`self`).
- Residencia (`R`): El tamaño de la población de residencia del individuo. Estados posibles: `big` y `small` .
- Transporte (`T`): El medio de transporte más utilizado por el encuestado para acudir al trabajo, diferenciando 3 posibles estados: `car` , `train` y `other` .

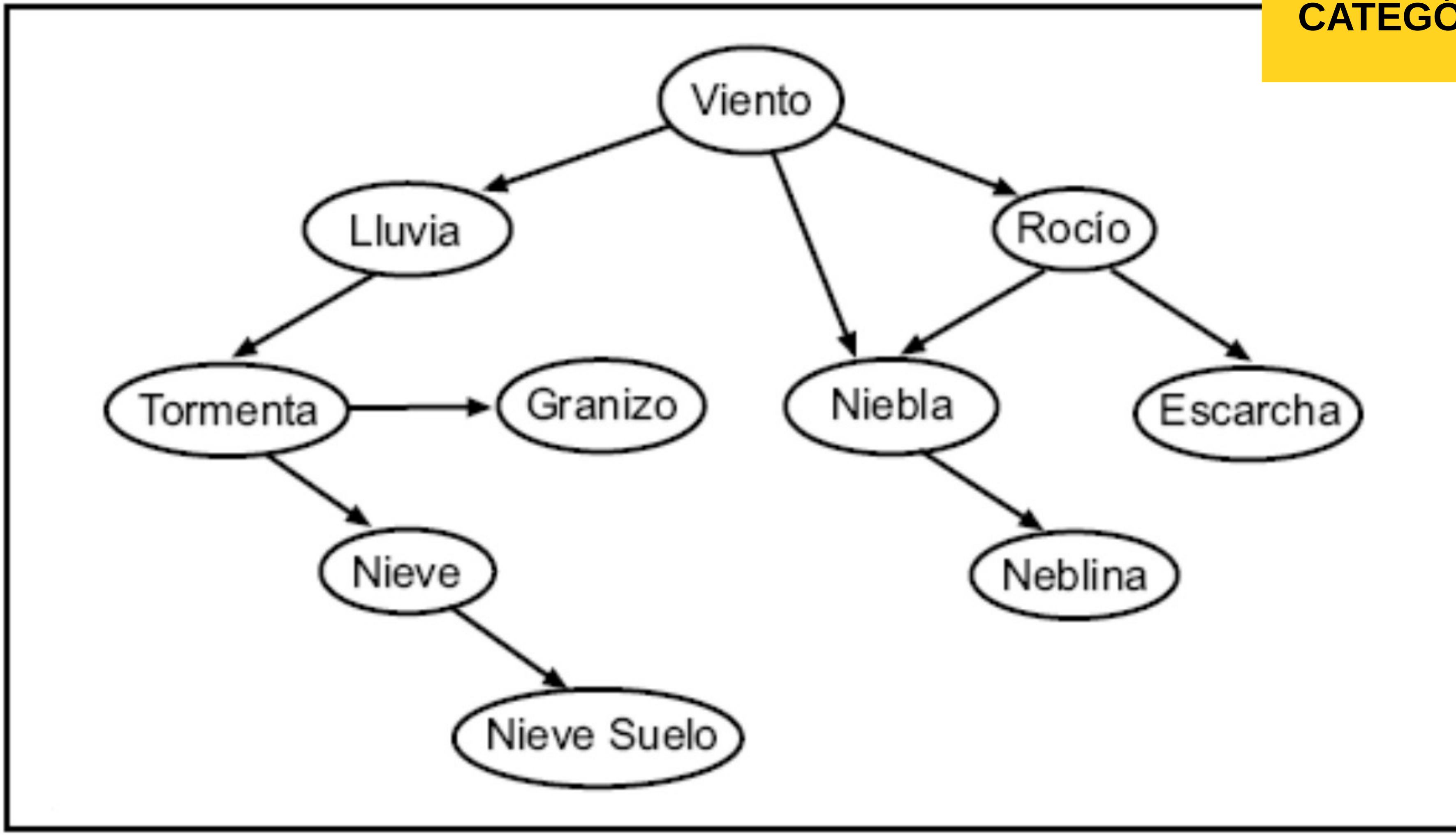
MIXTO



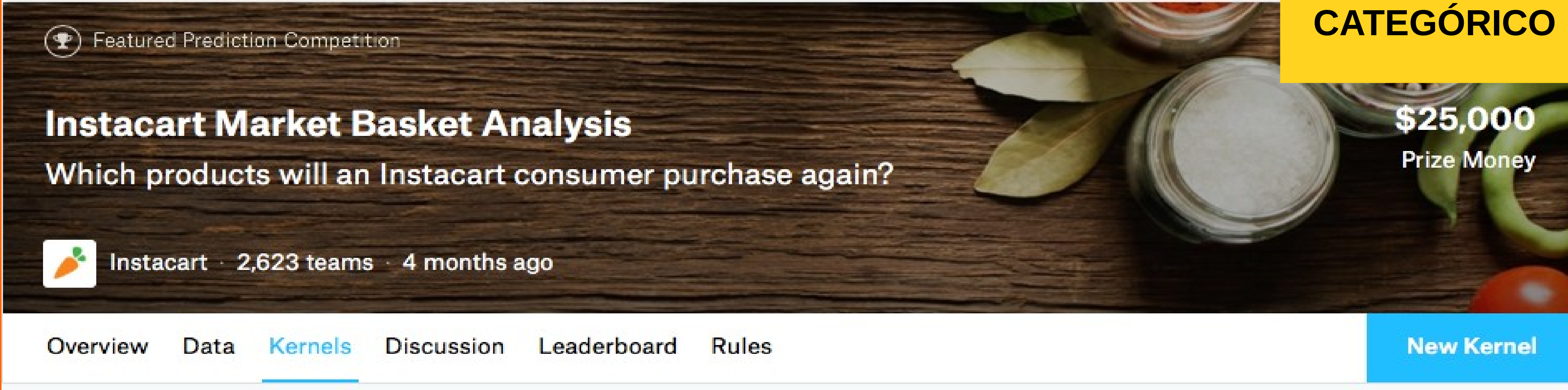


Marco Scutari: *Bayesian networks in R & Bayesian networks with examples in R*

http://www.bnlearn.com/




```
Lluvia nieve granizo tormenta niebla rocio escarcha nieveSuelo neblina viento
s     n     n     n     n     n     n     n     n     s
s     n     n     n     n     n     n     n     n     s
s     n     n     s     n     n     n     n     n     s
s     n     n     n     n     n     n     n     n     s
```



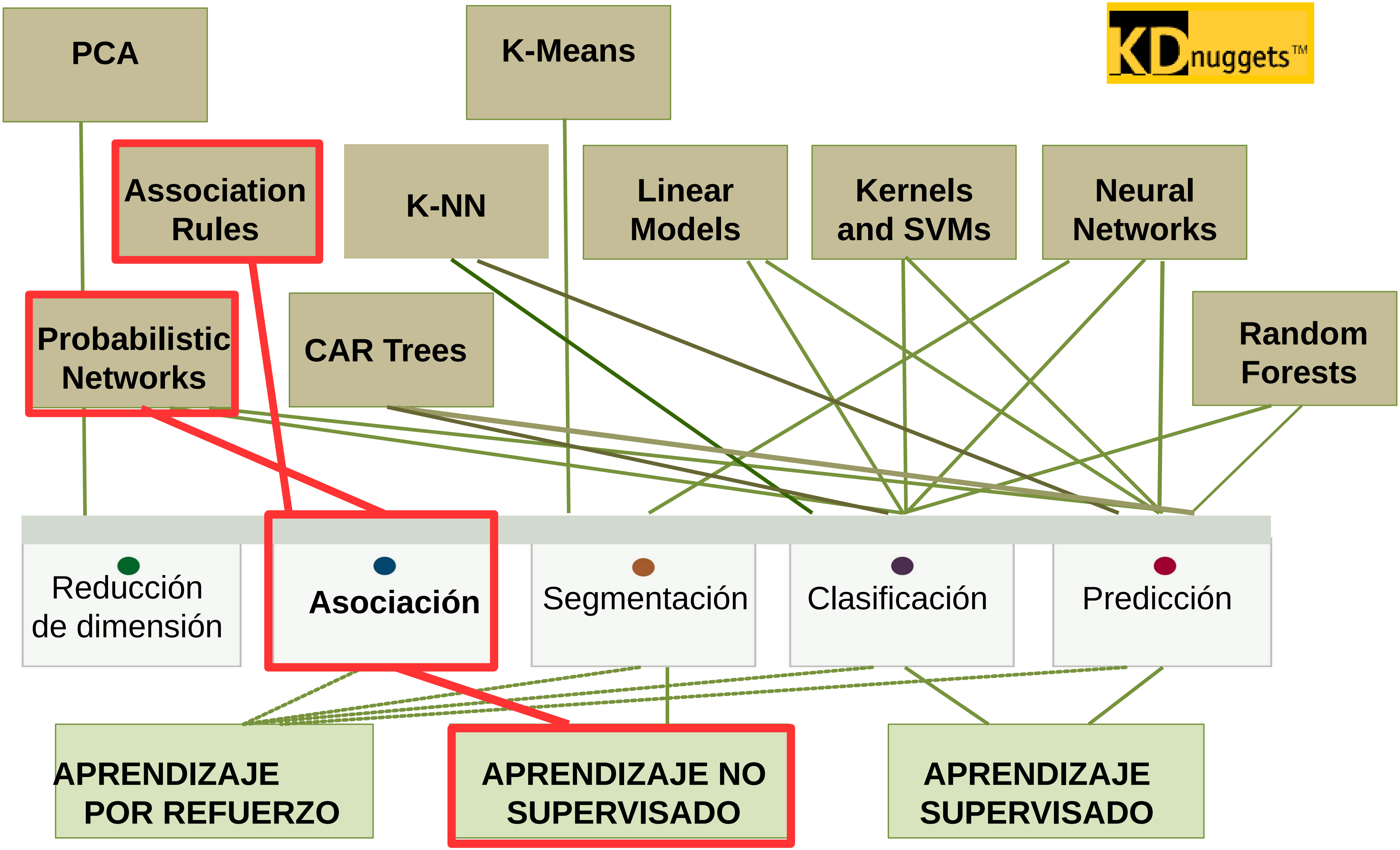
https://www.kaggle.com/philippsp/exploratory-analysis-instacart

En el curso utilizaremos un dataset más pequeño, “Groceries”, disponible en el paquete de R **arulesViz**.

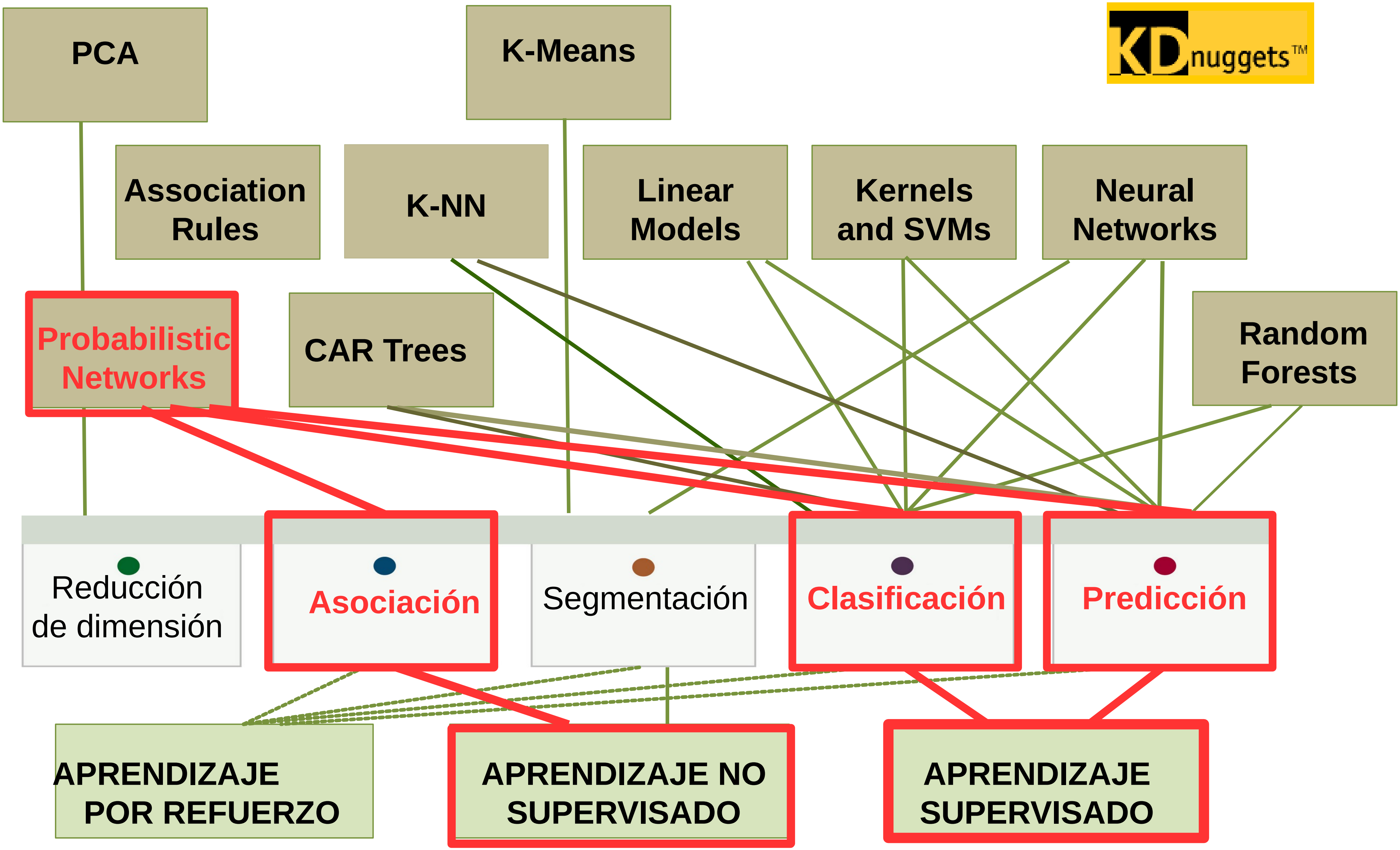
Attribute characteristics	Categorical
Number of instances	9835
Number of attributes	169

```
install.packages("arulesViz")
data(“Groceries”)
```

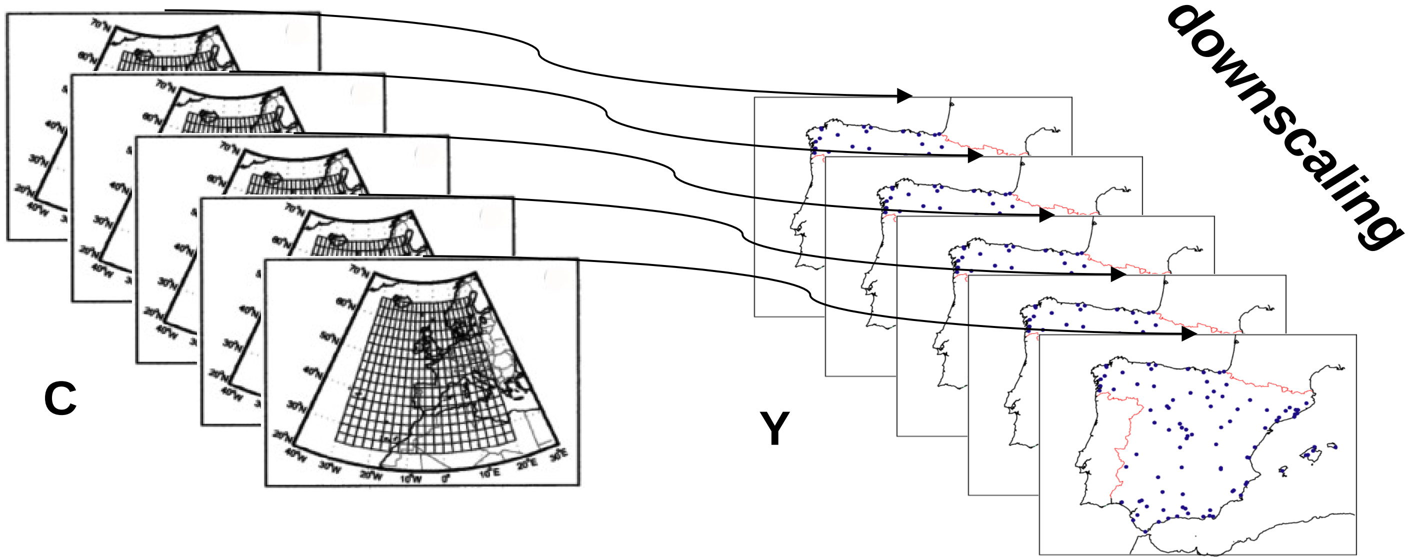
https://www.kdnuggets.com/2017/10/top-10-machine-learning-algorithms-beginners.html



https://www.kdnuggets.com/2017/10/top-10-machine-learning-algorithms-beginners.html



Clasificación
Predicción



Probabilistic Networks

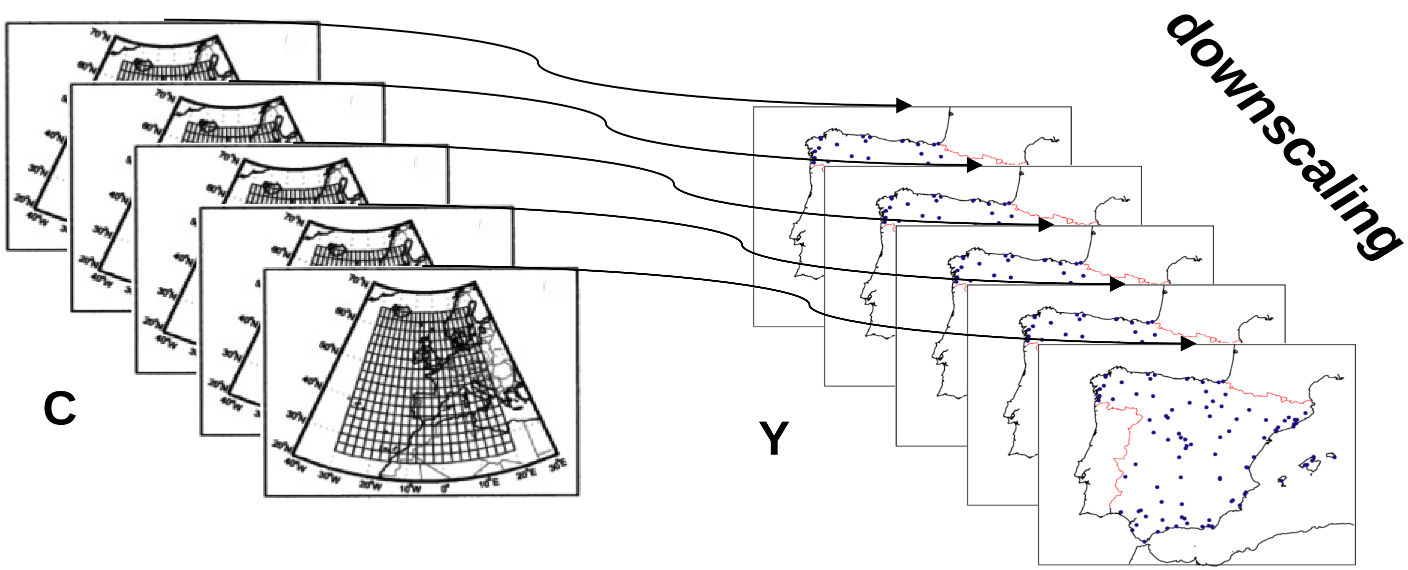
$$P(\mathbf{y}|\mathbf{c}) = P(y_1, \ldots, y_n \mid c_1, \ldots, c_m)$$

Reducción de dimensión	**Asociación**	Segmentación	**Clasificación**	**Predicción**

APRENDIZAJE POR REFUERZO

APRENDIZAJE NO SUPERVISADO

APRENDIZAJE SUPERVISADO



Clasificación
Predicción

Probabilistic Networks

$$P(\mathbf{y}|\mathbf{c}) = P(y_1, \ldots, y_n \mid c_1, \ldots, c_m)$$

Article published in *Water Resources Research*. doi: 10.1029/2019WR026416

Predicción

Multisite Weather Generators using Bayesian Networks: An illustrative case study for precipitation occurrence

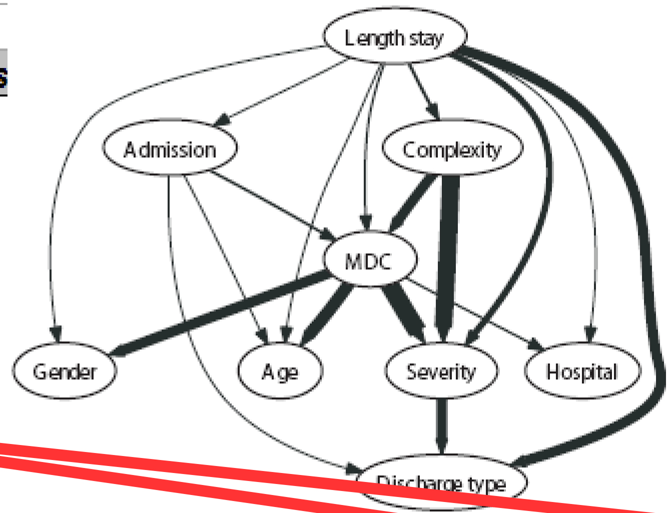
IZAJE
ISADO

M.N. Legasa[1], J.M. Gutiérrez[2]

[1]Meteorology Group. Dpto. de Matemática Aplicada y Ciencias de la Computación, Universidad de Cantabria, Santander, Spain
[2]Meteorology Group. Instituto de Física de Cantabria (IFCA, CSIC-UC). Santander, Spain

Healthcare management (ICMBD)

1. Estancia Media
2. Tasa de Mortalidad
3. Tasa de Reingresos (a 30 días)
4. Tasa de Infección Nosocomial
5. Estancia Media Preoperatoria
6. Tasa de Cesáreas
Conjunto de indicadores



RELACIONADAS CON LA ENFERMEDAD
1. Complejidad (medido a través del peso español de GRD-AP v 18)
2. Severidad (medido a través de GRD refinados)
3. Categoría Diagnostica Mayor de GRD-AP v 18
4. Tipo de GRD: medico, quirúrgico, indeterminado (Solo aplicado al indicador de infección nosocomial)
RELACIONADAS CON EL PACIENTE, O CON EL FUNCIONAMIENTO HOSPITALARIO
1. Edad
2. Sexo
3. Tipo de ingreso
4. Tipo de alta
5. Tipo de hospital
Variables de influencia propuestas para el análisis

Probabilistic Networks

Reducción de dimensión

Asociación

Segmentación

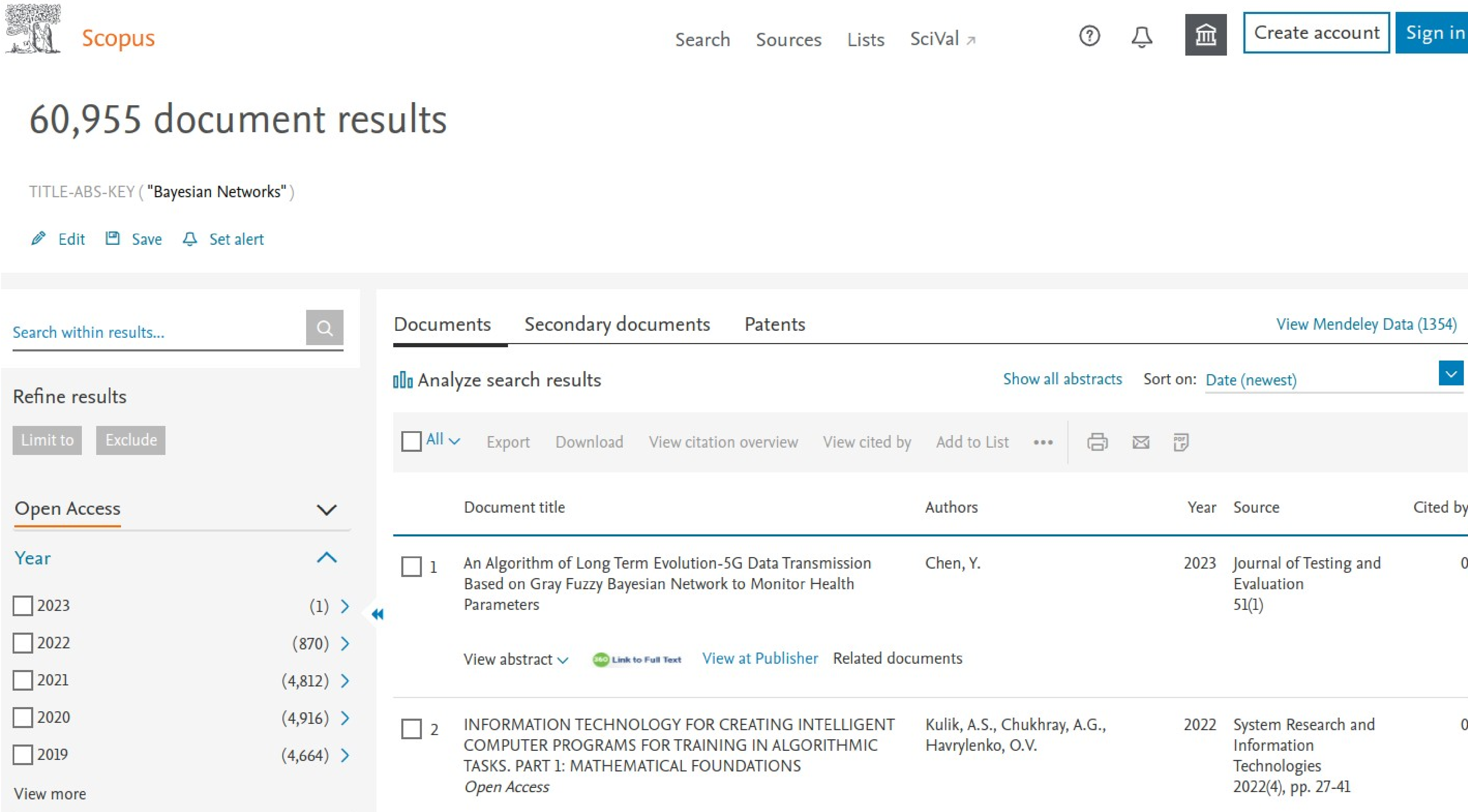
Clasificación

Predicción

APRENDIZAJE POR REFUERZO

APRENDIZAJE NO SUPERVISADO

APRENDIZAJE SUPERVISADO



APRENDIZAJE POR REFUERZO

APRENDIZAJE NO SUPERVISADO

APRENDIZAJE SUPERVISADO

$$P: \mathbf{M} \longrightarrow [0,1]$$
$$A \longmapsto a$$

	Anual		Invierno		Primavera		Verano		Otoño	
	S	Ll	S	Ll	S	Ll	S	Ll	S	Ll
NE	1014	516	190	99	287	166	360	162	177	89
SE	64	57	24	18	6	4	1	9	33	26
SW	225	661	98	223	18	119	15	71	94	248
NW	288	825	49	150	95	277	108	251	36	147
Total	1591	2059	361	490	406	566	484	493	340	510

$$P(X) \in [0,1], X \subseteq M$$

$$P(\,) = 0 \wedge P(M) = 1$$

$$P(X_1 \cup X_2) = P(X_1) + P(X_2) - P(X_1 \cap X_2)$$

$$P(X_1 \cap X_2) = P(X_1) * P(X_2) \Rightarrow X_1 \wedge X_2 \, independent$$

```
## States of the variables:
estados.Wind < - c("NE","SE","SW","NW")
estados.Season < - c("Anual","Invierno","Primavera","Verano","Otono")
estados.Precip < - c("Seco","Lluvioso")

## Table of Absolute frequencies:
table.freq < - array(c(1014, 64, 225, 288, 190, 24, 98, 49, 287, 6, 18, 95, 360, 1, 15, 108,
                        177, 33, 94, 36, 516, 57, 661, 825, 99, 18, 223, 150,
                          166, 4, 119, 277, 162, 9, 71, 251, 89, 26, 248, 147), dim = c(4,5,2),
                   dimnames = list(W=estados.Wind, S=estados.Season, P = estados.Precip))
```

$$P: \mathbf{M} \longrightarrow [0,1]$$
$$A \longrightarrow a$$

	Anual		Invierno		Primavera		Verano		Otoño	
	S	Ll	S	Ll	S	Ll	S	Ll	S	Ll
NE	1014	516	190	99	287	166	360	162	177	89
SE	64	57	24	18	6	4	1	9	33	26
SW	225	661	98	223	18	119	15	71	94	248
NW	288	825	49	150	95	277	108	251	36	147
Total	1591	2059	361	490	406	566	484	493	340	510

$$P(X) \in [0,1], X \subseteq M$$

$$P()=0 \wedge P(M)=1$$

$$P(X_1 \cup X_2)=P(X_1)+P(X_2)-P(X_1 \cap X_2)$$

$$P(X_1 \cap X_2)=P(X_1) * P(X_2) \Rightarrow X_1 \wedge X_2 \, independent$$

$$X=\{Inv, \mathrm{Ll}, NW\} \Rightarrow P(X)=\frac{freq(Inv, \mathrm{Ll}, NW)}{N}=\frac{150}{3650}=0.041$$

```
## States of the variables:
estados.Wind < - c("NE","SE","SW","NW")
estados.Season < - c("Anual","Invierno","Primavera","Verano","Otono")
estados.Precip < - c("Seco","Lluvioso")

## Table of Absolute frequencies:
table.freq < - array(c(1014, 64, 225, 288, 190, 24, 98, 49, 287, 6, 18, 95, 360, 1, 15, 108,
                       177, 33, 94, 36, 516, 57, 661, 825, 99, 18, 223, 150,
                        166, 4, 119, 277, 162, 9, 71, 251, 89, 26, 248, 147), dim = c(4,5,2),
                    dimnames = list(W=estados.Wind, S=estados.Season, P = estados.Precip))
## Obtain the probability:
table.freq["NW","Invierno","Lluvioso"]/sum(table.freq[,"Anual",])
```

$$P: \mathbf{M} \longrightarrow [0,1]$$
$$A \longrightarrow a$$

	Anual		Invierno		Primavera		Verano		Otoño	
	S	Ll	S	Ll	S	Ll	S	Ll	S	Ll
NE	1014	516	190	99	287	166	360	162	177	89
SE	64	57	24	18	6	4	1	9	33	26
SW	225	661	98	223	18	119	15	71	94	248
NW	288	825	49	150	95	277	108	251	36	147
Total	1591	2059	361	490	406	566	484	493	340	510

$$P(X) \in [0,1], X \subseteq M$$
$$P(\,) = 0 \wedge P(M) = 1$$
$$P(X_1 \cup X_2) = P(X_1) + P(X_2) - P(X_1 \cap X_2)$$
$$P(X_1 \cap X_2) = P(X_1) * P(X_2) \Rightarrow X_1 \wedge X_2 \, independent$$

$$X = \{Inv, \mathrm{Ll}, NW\} \Rightarrow P(X) = \frac{freq(Inv, \mathrm{Ll}, NW)}{N} = \frac{150}{3650} = 0.041$$

$$\boxed{X = \{Inv\}} \Rightarrow P(X) = \frac{freq(Inv)}{N} = \frac{\sum_{p \in Pr} \sum_{v \in Vi} freq(Inv, p, v)}{N} = 0.233$$

```
## Obtain the probability:
sum(table.freq[,"Invierno",])/sum(table.freq[,"Anual",])
```

$$P: \mathbf{M} \longrightarrow [0,1]$$
$$A \longrightarrow a$$

	Anual		Invierno		Primavera		Verano		Otoño	
	S	Ll	S	Ll	S	Ll	S	Ll	S	Ll
NE	1014	516	190	99	287	166	360	162	177	89
SE	64	57	24	18	6	4	1	9	33	26
SW	225	661	98	223	18	119	15	71	94	248
NW	288	825	49	150	95	277	108	251	36	147
Total	1591	2059	361	490	406	566	484	493	340	510

$$P(X) \in [0,1], X \subseteq M$$
$$P() = 0 \wedge P(M) = 1$$
$$P(X_1 \cup X_2) = P(X_1) + P(X_2) - P(X_1 \cap X_2)$$
$$P(X_1 \cap X_2) = P(X_1) * P(X_2) \Rightarrow X_1 \wedge X_2 \, independent$$

$$X = \{Inv, \mathrm{Ll}, NW\} \Rightarrow P(X) = \frac{freq(Inv, \mathrm{Ll}, NW)}{N} = \frac{150}{3650} = 0.041$$

$$X = \{Inv\} \Rightarrow P(X) = \frac{freq(Inv)}{N} = \frac{\sum_{p \in Pr} \sum_{v \in Vi} freq(Inv, p, v)}{N} = 0.233$$

$$\boxed{X = \{NW\} \Rightarrow P(Y \vee X) = \frac{P(Y, X)}{P(X)}}$$

$$Y = \{Inv\} \Rightarrow P(Y \vee X) = \frac{freq(Y, X)}{freq(X)} = \frac{199}{1113} = 0.179$$

$$P: \mathbf{M} \longrightarrow [0,1]$$
$$A \longrightarrow a$$

	Anual		Invierno		Primavera		Verano		Otoño	
	S	Ll	S	Ll	S	Ll	S	Ll	S	Ll
NE	1014	516	190	99	287	166	360	162	177	89
SE	64	57	24	18	6	4	1	9	33	26
SW	225	661	98	223	18	119	15	71	94	248
NW	288	825	49	150	95	277	108	251	36	147
Total	1591	2059	361	490	406	566	484	493	340	510

$$P(X) \in [0,1], X \subseteq M$$
$$P(\,) = 0 \wedge P(M) = 1$$
$$P(X_1 \cup X_2) = P(X_1) + P(X_2) - P(X_1 \cap X_2)$$
$$P(X_1 \cap X_2) = P(X_1) * P(X_2) \Rightarrow X_1 \wedge X_2 \, independent$$

$$X = \{Inv, \mathrm{Ll}, NW\} \Rightarrow P(X) = \frac{freq(Inv, \mathrm{Ll}, NW)}{N} = \frac{150}{3650} = 0.041$$

$$X = \{Inv\} \Rightarrow P(X) = \frac{freq(Inv)}{N} = \frac{\sum_{p \in Pr} \sum_{v \in Vi} freq(Inv, p, v)}{N} = 0.233$$

$$\boxed{X = \{NW\} \Rightarrow P(Y \vee X) = \frac{P(Y, X)}{P(X)}}$$

$$Y = \{Inv\} \Rightarrow P(Y \vee X) = \frac{freq(Y, X)}{freq(X)} = \frac{199}{1113} = 0.179$$

```
## New probability-space:
cond.table.freq <- table.freq["NW",,]
print(cond.table.freq)
```

$$P: \mathbf{M} \longrightarrow [0,1]$$
$$A \longrightarrow a$$

	Anual		Invierno		Primavera		Verano		Otoño	
	S	Ll	S	Ll	S	Ll	S	Ll	S	Ll
NE	1014	516	190	99	287	166	360	162	177	89
SE	64	57	24	18	6	4	1	9	33	26
SW	225	661	98	223	18	119	15	71	94	248
NW	288	825	49	150	95	277	108	251	36	147
Total	1591	2059	361	490	406	566	484	493	340	510

$$P(X) \in [0,1], X \subseteq M$$

$$P(\,) = 0 \wedge P(M) = 1$$

$$P(X_1 \cup X_2) = P(X_1) + P(X_2) - P(X_1 \cap X_2)$$

$$P(X_1 \cap X_2) = P(X_1) * P(X_2) \Rightarrow X_1 \wedge X_2 \, independent$$

$$X = \{Inv, \mathrm{Ll}, NW\} \Rightarrow P(X) = \frac{freq(Inv, \mathrm{Ll}, NW)}{N} = \frac{150}{3650} = 0.041$$

$$X = \{Inv\} \Rightarrow P(X) = \frac{freq(Inv)}{N} = \frac{\sum_{p \in Pr} \sum_{v \in Vi} freq(Inv, p, v)}{N} = 0.233$$

$$\boxed{X = \{NW\} \Rightarrow P(Y \vee X) = \frac{P(Y, X)}{P(X)}}$$

$$Y = \{Inv\} \Rightarrow P(Y \vee X) = \frac{freq(Y, X)}{freq(X)} = \frac{199}{1113} = 0.179$$

```
## Obtain the probability:
sum(cond.table.freq["Invierno",])/sum(cond.table.freq["Anual",])
```

$$P: \mathbf{M} \longrightarrow [0,1]$$
$$A \longmapsto a$$

	Anual		Invierno		Primavera		Verano		Otoño	
	S	Ll	S	Ll	S	Ll	S	Ll	S	Ll
NE	1014	516	190	99	287	166	360	162	177	89
SE	64	57	24	18	6	4	1	9	33	26
SW	225	661	98	223	18	119	15	71	94	248
NW	288	825	49	150	95	277	108	251	36	147
Total	1591	2059	361	490	406	566	484	493	340	510

$$P(X) \in [0,1], X \subseteq M$$

$$P(\,) = 0 \wedge P(M) = 1$$

$$P(X_1 \cup X_2) = P(X_1) + P(X_2) - P(X_1 \cap X_2)$$

$$P(X_1 \cap X_2) = P(X_1) * P(X_2) \Rightarrow X_1 \wedge X_2 independent \Leftarrow P(X_1 \vee X_2) = P(X_1) \wedge P(X_2 \vee X_1) = P(X_2)$$

Bayes' Theorem (Predictands vs. Predictors), Factorization, etc.

$$\{X_1, \ldots, X_n : X_1 \cup \ldots \cup X_n = M \wedge X_i \cap X_j = \varnothing \; \forall i \neq j\} \Rightarrow P(X_i \vee B) = \frac{j}{\,}$$

$$X = \{NW\} \Rightarrow P(Y \vee X) = \frac{P(Y, X)}{P(X)}$$

$$Y = \{Inv\} \Rightarrow P(Y \vee X) = \frac{freq(Y, X)}{freq(X)} = \frac{199}{1113} = 0.179$$

$$P: \mathbf{M} \longrightarrow [0,1]$$
$$A \longmapsto a$$

	Anual		Invierno		Primavera		Verano		Otoño	
	S	Ll	S	Ll	S	Ll	S	Ll	S	Ll
NE	1014	516	190	99	287	166	360	162	177	89
SE	64	57	24	18	6	4	1	9	33	26
SW	225	661	98	223	18	119	15	71	94	248
NW	288	825	49	150	95	277	108	251	36	147
Total	1591	2059	361	490	406	566	484	493	340	510

$$P(X) \in [0,1], X \subseteq M$$

$$P(\,) = 0 \wedge P(M) = 1$$

$$P(X_1 \cup X_2) = P(X_1) + P(X_2) - P(X_1 \cap X_2)$$

$$P(X_1 \cap X_2) = P(X_1) * P(X_2) \Rightarrow X_1 \wedge X_2 \, independent \Leftarrow P(X_1 \vee X_2) = P(X_1) \wedge P(X_2 \vee X_1) = P(X_2)$$

Bayes' Theorem (Predictands vs. Predictors), Factorization, etc.

$$\{X_1, \ldots, X_n : X_1 \cup \ldots \cup X_n = M \wedge X_i \cap X_j = \varnothing \; \forall i \neq j\} \Rightarrow P(X_i \vee B) = \frac{P(B \vee X_i)\, P(X_i)}{P(B)}$$

$$X = \{NW\} \Rightarrow P(Y \vee X) = \frac{P(Y, X)}{P(X)}$$

$$Y = \{Inv\} \Rightarrow P(Y \vee X) = \frac{freq(Y, X)}{freq(X)} = \frac{199}{1113} = 0.179$$

$$P:\ \mathbf{M} \longrightarrow [0,1]$$
$$A \longmapsto a$$

	Anual		Invierno		Primavera		Verano		Otoño	
	S	Ll	S	Ll	S	Ll	S	Ll	S	Ll
NE	1014	516	190	99	287	166	360	162	177	89
SE	64	57	24	18	6	4	1	9	33	26
SW	225	661	98	223	18	119	15	71	94	248
NW	288	825	49	150	95	277	108	251	36	147
Total	1591	2059	361	490	406	566	484	493	340	510

$$P(X) \in [0,1], X \subseteq M$$
$$P(\,)=0 \wedge P(M)=1$$
$$P(X_1 \cup X_2)=P(X_1)+P(X_2)-P(X_1 \cap X_2)$$
$$P(X_1 \cap X_2)=P(X_1) * P(X_2) \Rightarrow X_1 \wedge X_2 independent \Leftarrow P(X_1 \vee X_2)=P(X_1) \wedge P(X_2 \vee X_1)=P(X_2)$$

Bayes' Theorem (Predictands vs. Predictors), Factorization, etc.

$$\{X_1, \ldots, X_n : X_1 \cup \ldots \cup X_n = M \wedge X_i \cap X_j = \varnothing\ \forall i \neq j\} \Rightarrow P\boxed{(X_i \vee B)=}\frac{P(B \vee X_i)\boxed{P(X_i)}}{P(B)}$$

Probability "*a posteriori*" **Probability "*a priori*"**

$$X=\{NW\} \Rightarrow P(Y \vee X)=\frac{P(Y,X)}{P(X)} \quad \Rightarrow \quad Y=\{Inv\} \Rightarrow P(Y \vee X)=\frac{freq(Y,X)}{freq(X)}=\frac{199}{1113}=0.179$$

$$\begin{array}{llll} P: & \mathbf{M} & \longrightarrow & [0,1] \\ & A & \longmapsto & a \end{array}$$

	Anual		Invierno		Primavera		Verano		Otoño	
	S	Ll	S	Ll	S	Ll	S	Ll	S	Ll
NE	1014	516	190	99	287	166	360	162	177	89
SE	64	57	24	18	6	4	1	9	33	26
SW	225	661	98	223	18	119	15	71	94	248
NW	288	825	49	150	95	277	108	251	36	147
Total	1591	2059	361	490	406	566	484	493	340	510

$$P(X) \in [0,1], X \subseteq M$$

$$P(\,)=0 \wedge P(M)=1$$

$$P(X_1 \cup X_2)=P(X_1)+P(X_2)-P(X_1 \cap X_2)$$

$$P(X_1 \cap X_2)=P(X_1) * P(X_2) \Rightarrow X_1 \wedge X_2 \, independent \Leftarrow P(X_1 \vee X_2)=P(X_1) \wedge P(X_2 \vee X_1)=P(X_2)$$

[Bayes' Theorem (Predictands vs. Predictors), Factorization, etc.]

Verosimilitud

$$\{X_1,\dots,X_n : X_1 \cup \dots \cup X_n = M \wedge X_i \cap X_j = \varnothing \; \forall i \neq j\} \Rightarrow P(X_i \vee B)=\frac{\boxed{P(B \vee X_i)}\,P(X_i)}{P(B)}$$

$$X=\{NW\} \Rightarrow P(Y \vee X)=\frac{P(Y,X)}{P(X)}$$

$$\Rightarrow$$

$$Y=\{Inv\} \Rightarrow P(Y \vee X)=\frac{freq(Y,X)}{freq(X)}=\frac{199}{1113}=0.179$$

$$P: \mathbf{M} \longrightarrow [0,1]$$
$$A \longrightarrow a$$

	Anual		Invierno		Primavera		Verano		Otoño	
	S	Ll	S	Ll	S	Ll	S	Ll	S	Ll
NE	1014	516	190	99	287	166	360	162	177	89
SE	64	57	24	18	6	4	1	9	33	26
SW	225	661	98	223	18	119	15	71	94	248
NW	288	825	49	150	95	277	108	251	36	147
Total	1591	2059	361	490	406	566	484	493	340	510

$$P(X) \in [0,1], X \subseteq M$$

$$P()=0 \wedge P(M)=1$$

$$P(X_1 \cup X_2)=P(X_1)+P(X_2)-P(X_1 \cap X_2)$$

$$P(X_1 \cap X_2)=P(X_1) * P(X_2) \Rightarrow X_1 \wedge X_2 \, independent \Leftarrow P(X_1 \vee X_2)=P(X_1) \wedge P(X_2 \vee X_1)=P(X_2)$$

Bayes' Theorem (Predictands vs. Predictors), Factorization, etc.

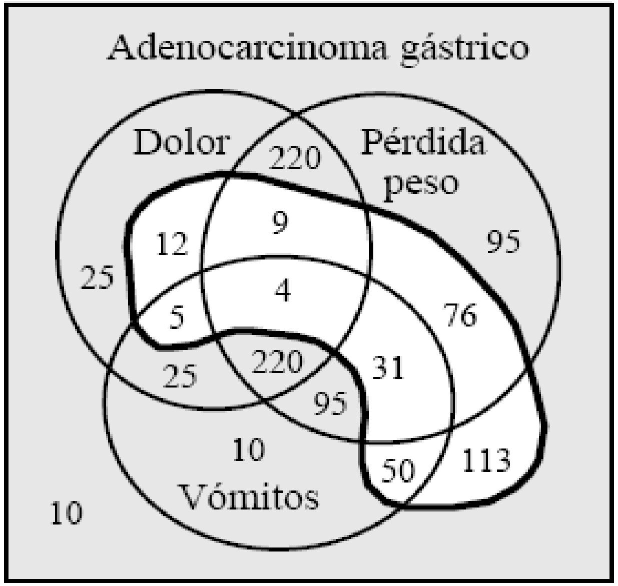
$$\{X_1, \ldots, X_n : X_1 \cup \ldots \cup X_n = M \wedge X_i \cap X_j = \varnothing \, \forall i \neq j\} \Rightarrow P(X_i \vee B) = \frac{P(B \vee X_i) P(X_i)}{\boxed{\sum_{j=1}^{n} P(B \vee X_j) P(X_j)}}$$

Partición del espacio muestral

$$X=\{NW\} \Rightarrow P(Y \vee X)=\frac{P(Y,X)}{P(X)}$$

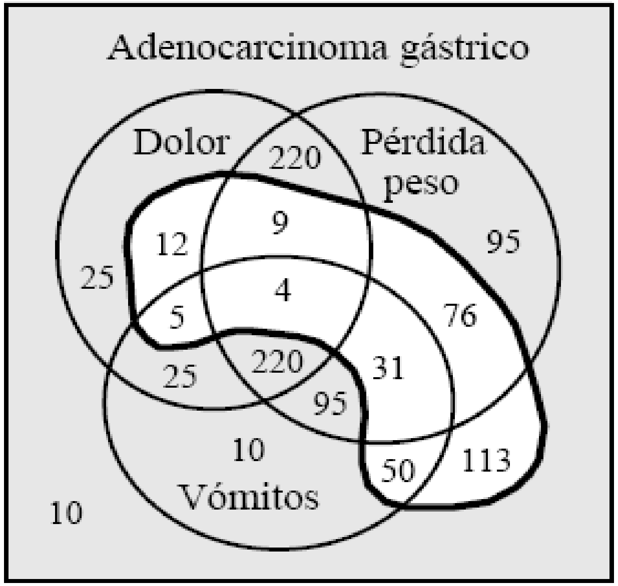
$$Y=\{Inv\} \Rightarrow P(Y \vee X)=\frac{freq(Y,X)}{freq(X)}=\frac{199}{1113}=0.179$$

Could we predict the probability of a disease based on the symptoms?



$$p(e_i|s_1,\ldots,s_k) = \frac{p(s_1,\ldots,s_k|e_i)p(e_i)}{\sum_{e_i} p(s_1,\ldots,s_k|e_i)p(e_i)}.$$

- La probabilidad $p(e_i)$ se llama probabilidad *marginal*, *prior*, *"a priori"* o *inicial* de la enfermedad $E = e_i$ puesto que puede ser obtenida *antes* de conocer los síntomas.
- La probabilidad $p(e_i|s_1,\ldots,s_k)$ es la probabilidad *posterior*, *"a posteriori"* o *condicional* de la enfermedad $E = e_i$, puesto que se calcula *después* de conocer los síntomas $S_1 = s_1,\ldots,S_k = s_k$.
- La probabilidad $p(s_1,\ldots,s_k|e_i)$ se conoce por el nombre de *verosimilitud* de que un paciente con la enfermedad $E = e_i$ tenga los síntomas $S_1 = s_1,\ldots,S_k = s_k$.



Initial Probabilities:

Gray → Adenocarcinoma

White → Not Adenocarcinoma

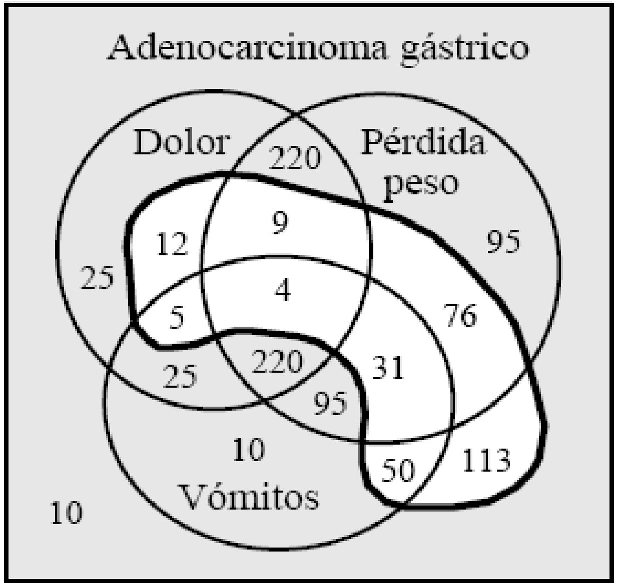
$$P(g)=\frac{700}{700+300}=\frac{700}{1000}=0.7$$

$$P(\neg g)=1-P(g)=1-0.7=0.3$$

Could we predict the probability of a disease based on the symptoms?

Bayes' Theorem (Predictands vs. Predictors), Factorization, etc.

$$\{X_1,\dots,X_n : X_1 \cup \dots \cup X_n = M \wedge X_i \cap X_j = \varnothing \; \forall i \neq j\} \Rightarrow P(X_i \vee B) = \frac{P(B \vee X_i)\,P(X_i)}{\sum_{j=1}^{n} P(B \vee X_j)\,P(X_j)}$$



Initial Probabilities:

Gray → Adenocarcinoma

White → Not Adenocarcinoma

$$P(g)=\frac{700}{700+300}=\frac{700}{1000}=0.7$$

$$P(\neg g)=1-P(g)=1-0.7=0.3$$

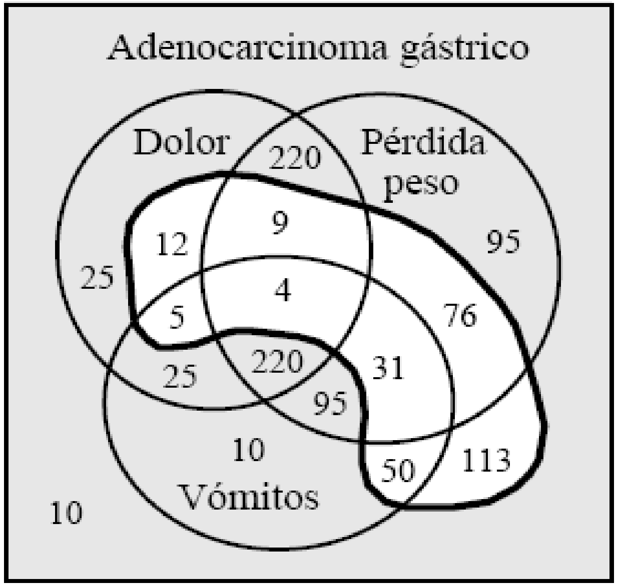
Could we predict the probability of a disease based on the symptoms?

Bayes' Theorem (Predictands vs. Predictors), Factorization, etc.

$$\{X_1,\dots,X_n : X_1 \cup \dots \cup X_n = M \wedge X_i \cap X_j = \varnothing \; \forall i \neq j\} \Rightarrow P(X_i \vee B)=\frac{P(B \vee X_i)P(X_i)}{\sum_{j=1}^{n} P(B \vee X_j)P(X_j)}$$

Patient has suffered threw up:

$$\{V=v\} \Rightarrow P(g \vee v)=\frac{P(g)P(v \vee g)}{P(g)P(v \vee g)+P(\neg g)P(v \vee \neg g)}=\frac{0.7 * 0.5}{0.7 * 0.5+0.3 * 0.3}=0.795$$



Initial Probabilities:

Gray → Adenocarcinoma

White → Not Adenocarcinoma

$$P(g)=\frac{700}{700+300}=\frac{700}{1000}=0.7$$

$$P(\neg g)=1-P(g)=1-0.7=0.3$$

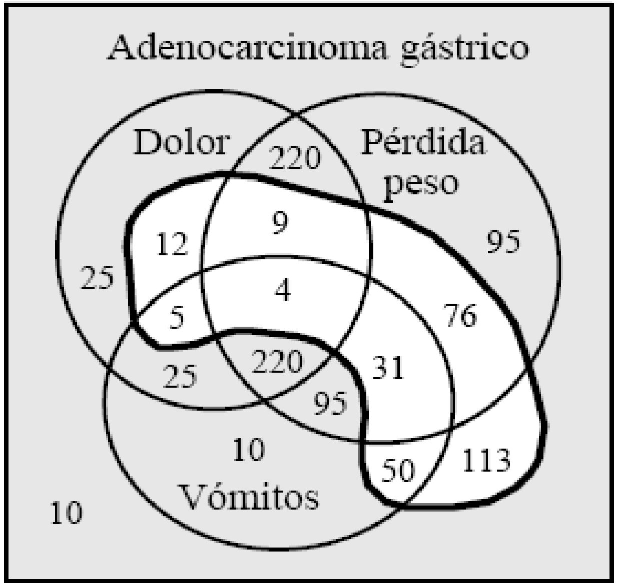
Could we predict the probability of a disease based on the symptoms?

Patient has suffered threw up:

$$\{V=v\}\Rightarrow P(g\vee v)=\frac{P(g)P(v\vee g)}{P(g)P(v\vee g)+P(\neg g)P(v\vee\neg g)}=\frac{0.7 * 0.5}{0.7 * 0.5+0.3 * 0.3}=0.795$$

Patient has suffered of weight loss and threw up:

$$\{P=p\wedge V=v\}\Rightarrow P(g\vee v,p)=\frac{P(g)P(v,p\vee g)}{P(g)P(v,p\vee g)+P(\neg g)P(v,p\vee\neg g)}=\frac{0.7 * 0.45}{0.7 * 0.45+0.3 * 0.12}=0.9$$



Initial Probabilities:

Gray → Adenocarcinoma

White → Not Adenocarcinoma

$$P(g)=\frac{700}{700+300}=\frac{700}{1000}=\boxed{0.7}$$

$$P(\neg g)=1-P(g)=1-0.7=0.3$$

Significant changes in the probabilities reflect the dependence between predictand and predictors.

Predictability

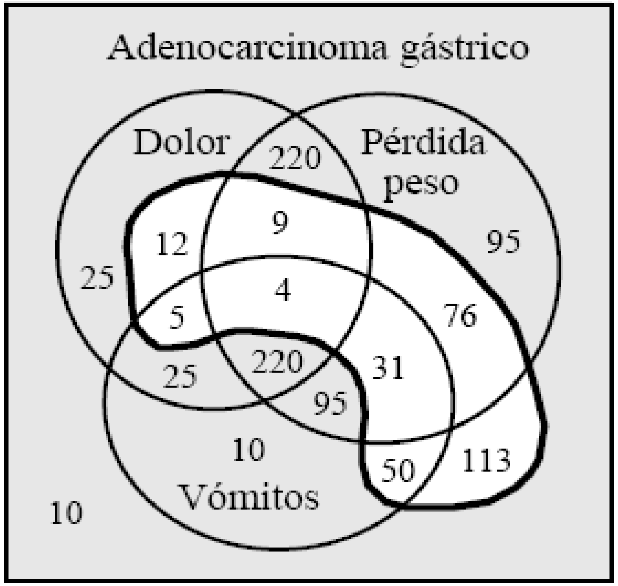
Could we predict the probability of a disease based on the symptoms?

Patient has suffered threw up:

$$\{V=v\}\Rightarrow P(g\vee v)=\frac{P(g)P(v\vee g)}{P(g)P(v\vee g)+P(\neg g)P(v\vee\neg g)}=\frac{0.7\,*\,0.5}{0.7\,*\,0.5+0.3\,*\,0.3}=\boxed{0.795}$$

Patient has suffered of weight loss and threw up:

$$\{P=p\wedge V=v\}\Rightarrow P(g\vee v,p)=\frac{P(g)P(v,p\vee g)}{P(g)P(v,p\vee g)+P(\neg g)P(v,p\vee\neg g)}=\frac{0.7\,*\,0.45}{0.7\,*\,0.45+0.3\,*\,0.12}=\boxed{0.9}$$



Initial Probabilities:

Gray → Adenocarcinoma

White → Not Adenocarcinoma

$$P(g) = \frac{700}{700+300} = \frac{700}{1000} = \boxed{0.7}$$

$$P(\neg g) = 1 - P(g) = 1 - 0.7 = 0.3$$

Significant changes in the probabilities reflect the dependence between predictand and predictors.

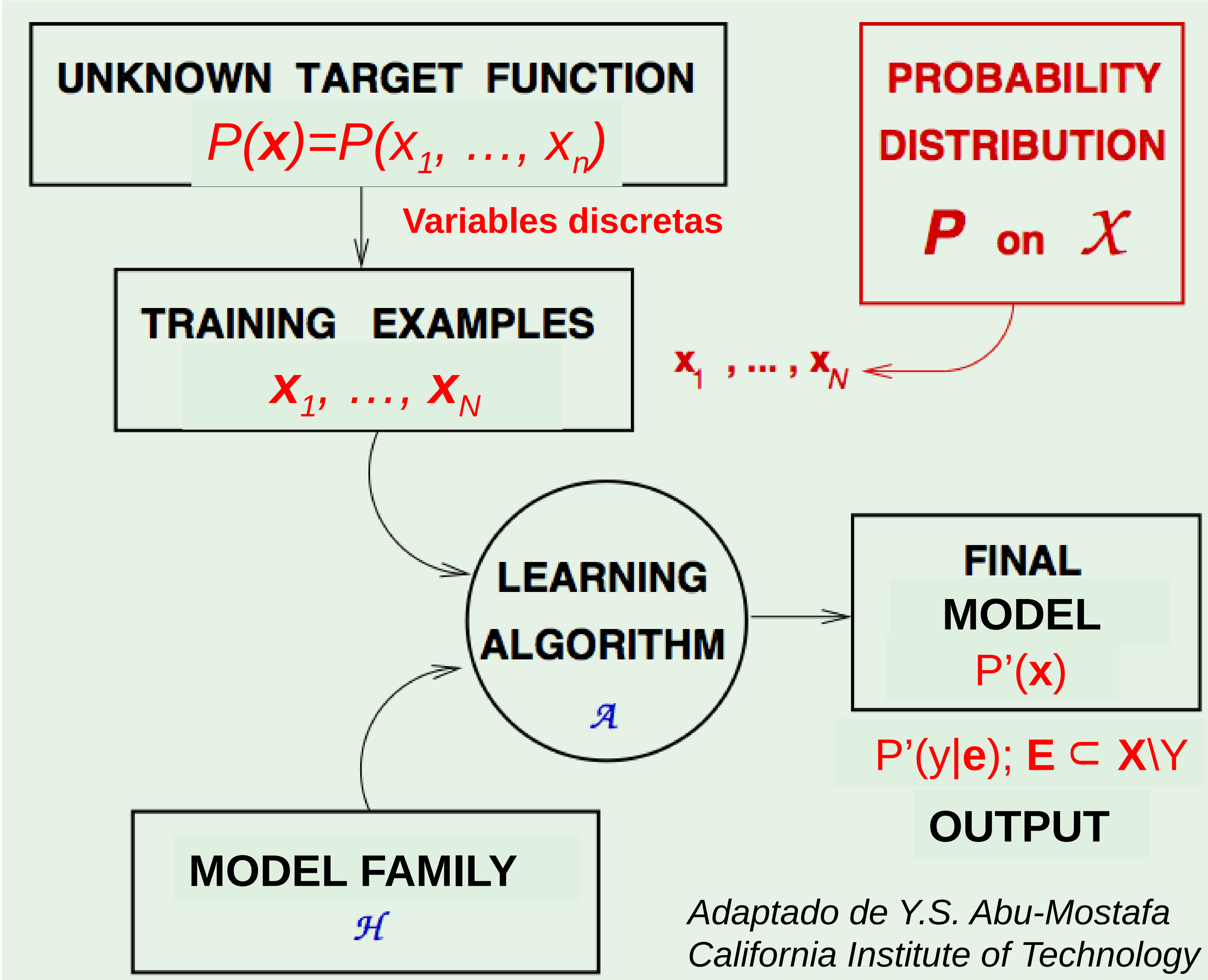
Predictability

Hypothesis Testing to Compare Two Population Proportions

$$Z = \frac{p_1 - p_2}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$p = \frac{x_1 + x_2}{n_1 + n_2}$$

$$If\ Z > N_{(0,1)}^{-1}(\alpha) \Rightarrow p_1 \neq p_2$$


UNKNOWN TARGET FUNCTION
$P(\mathbf{x})=P(x_1, \ldots, x_n)$
Variables discretas
PROBABILITY DISTRIBUTION
P on $\mathcal{X}$
TRAINING EXAMPLES
$\mathbf{x}_1, \ldots, \mathbf{x}_N$
$\mathbf{x}_1, \ldots, \mathbf{x}_N$
LEARNING ALGORITHM
$\mathcal{A}$
FINAL MODEL
$P'(\mathbf{x})$
$P'(y|\mathbf{e})$; $\mathbf{E} \subset \mathbf{X} \backslash Y$
OUTPUT
MODEL FAMILY
$\mathcal{H}$
Adaptado de Y.S. Abu-Mostafa
California Institute of Technology

x	y	z	$p(x,y,z)$
0	0	0	0.12
0	0	1	0.18
0	1	0	0.04
0	1	1	0.16
1	0	0	0.09
1	0	1	0.21
1	1	0	0.02
1	1	1	0.18

To explicitly define the joint probability function is not possible in most of the real cases due to the large amount of parameters (e.g. 10^{25} parameters for 100 variables).

x	y	z	$p(x,y,z)$
0	0	0	0.12
0	0	1	0.18
0	1	0	0.04
0	1	1	0.16
1	0	0	0.09
1	0	1	0.21
1	1	0	0.02
1	1	1	0.18

To explicitly define the joint probability function is not possible in most of the real cases due to the large amount of parameters (e.g. 10^{25} parameters for 100 variables).

Bayes's Theorem → Factorization

$$P(X_i \vee B) = \frac{P(B \vee X_i) P(X_i)}{\sum_{j=1}^{n} P(B \vee X_j) P(X_j)}$$

$$\underbrace{B \wedge X_i \, independent \Rightarrow P(X_i \vee B) = P(X_i) \wedge P(B \vee X_i) = P(B)}$$

Reduction of parameters

x	y	z	$p(x,y,z)$
0	0	0	0.12
0	0	1	0.18
0	1	0	0.04
0	1	1	0.16
1	0	0	0.09
1	0	1	0.21
1	1	0	0.02
1	1	1	0.18

To explicitly define the joint probability function is not possible in most of the real cases due to the large amount of parameters (e.g. 10^{25} parameters for 100 variables).

Bayes's Theorem → Factorization

$$P(X_i \vee B) = \frac{P(B \vee X_i) P(X_i)}{\sum_{j=1}^{n} P(B \vee X_j) P(X_j)}$$

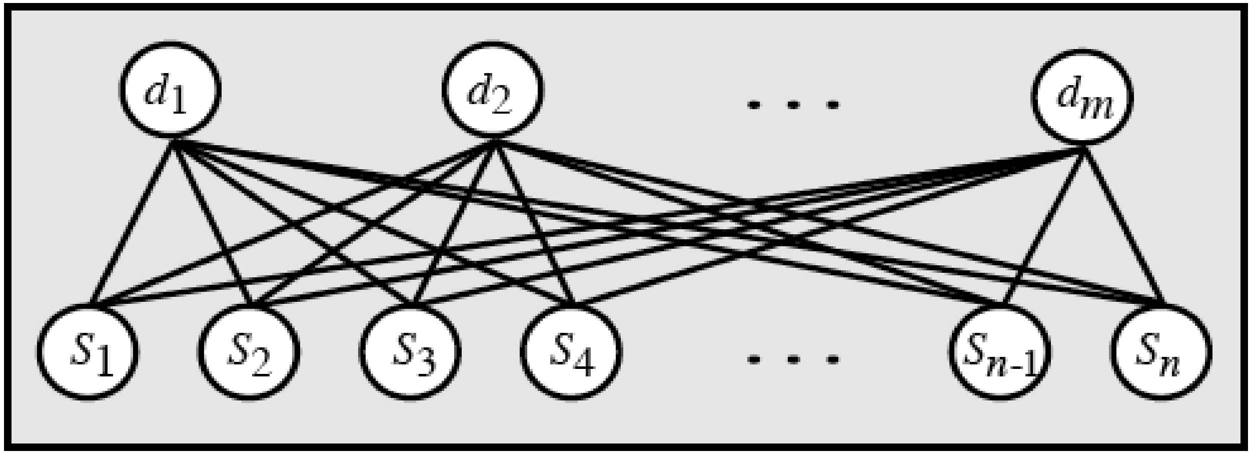
$$\underbrace{B \wedge X_i \mathit{independent} \Rightarrow P(X_i \vee B) = P(X_i) \wedge P(B \vee X_i) = P(B)}$$

Reduction of parameters

Can we build a model including some pre-defined independences?

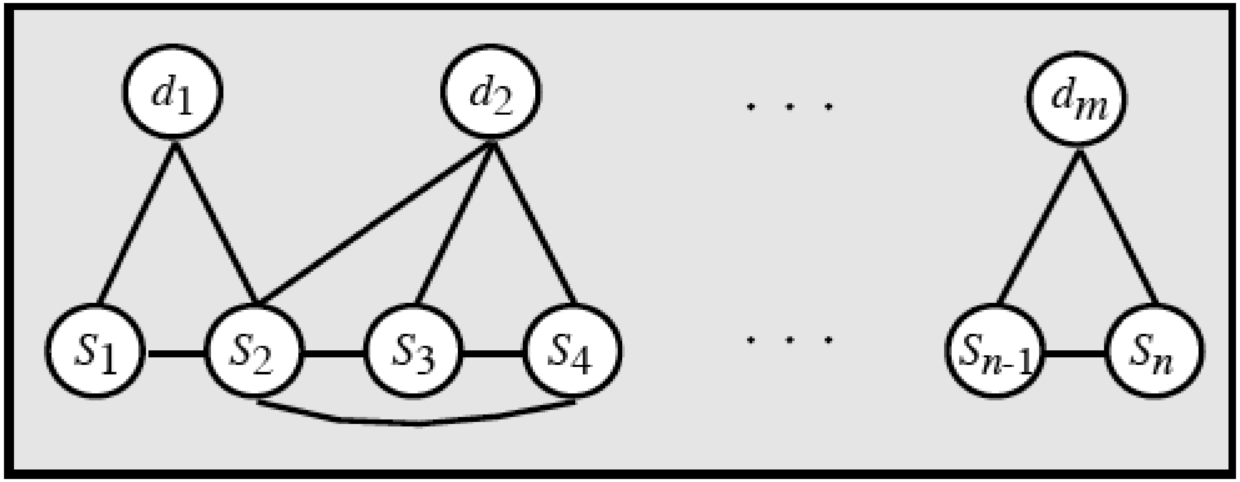
Firstly, unrealistic models "*ad-hoc*" were proposed.

$$P(s_1, \ldots, s_n, d_1, \ldots, d_m) = P(s_1, \ldots, s_n | d_1, \ldots, d_m)\, P(d_1, \ldots, d_m)$$



Independent symptoms model → Independent symptoms given a disease

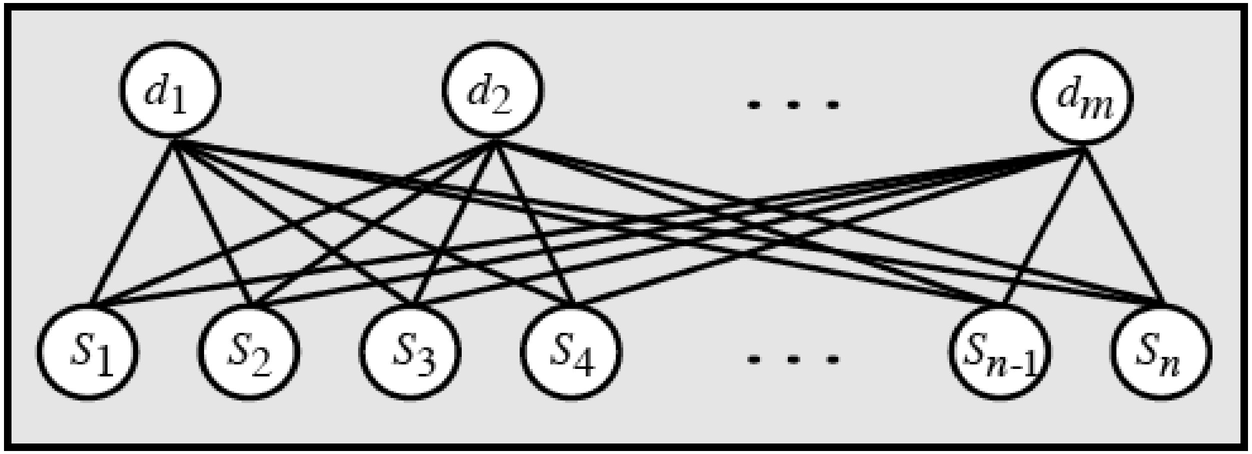
$$p(s_1, \ldots, s_n | d_i) = \prod_{j=1}^{n} p(s_j | d_i).$$



Syndrome model → for each disease there are a relevant subset of dependent symptoms.

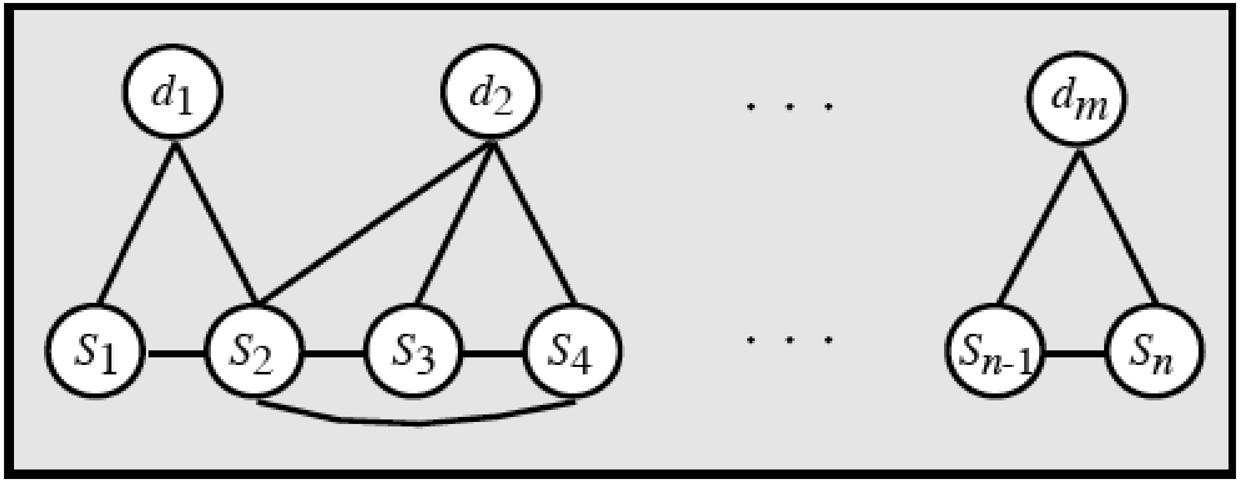
Firstly, unrealistic models “*ad-hoc*” were proposed.

$$P(s_1, \ldots, s_n, d_1, \ldots, d_m) = P(s_1, \ldots, s_n | d_1, \ldots, d_m)\, P(d_1, \ldots, d_m)$$



Independent symptoms model → Independent symptoms given a disease

$$p(s_1, \ldots, s_n | d_i) = \prod_{j=1}^{n} p(s_j | d_i).$$



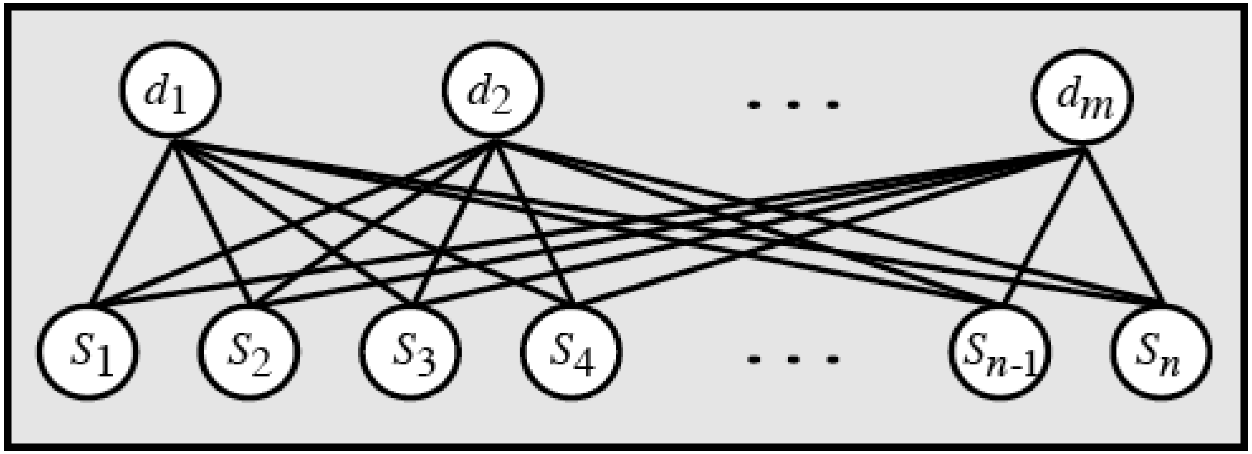
Syndrome model → for each disease there are a relevant subset of dependent symptoms.

Subjective/ad-hoc approach
Large amount of parameters

Modelo	Número de parámetros	
	Fórmula	Valor
DSM	$m2^n - 1$	$> 10^{62}$
ISM	$m(n+1) - 1$	20,099
IRSM	$m(r+1) + n - 1$	1,299
DRSM	$m2^r + n - 1$	102,599

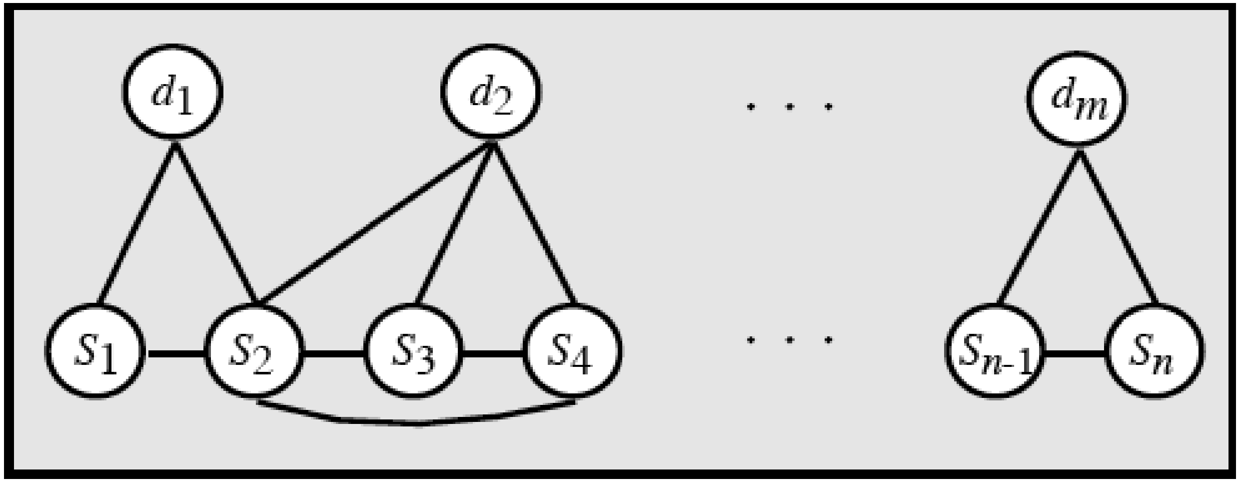
Firstly, unrealistic models "*ad-hoc*" were proposed.

$$P(s_1, \ldots, s_n, d_1, \ldots, d_m) = P(s_1, \ldots, s_n | d_1, \ldots, d_m)\, P(d_1, \ldots, d_m)$$



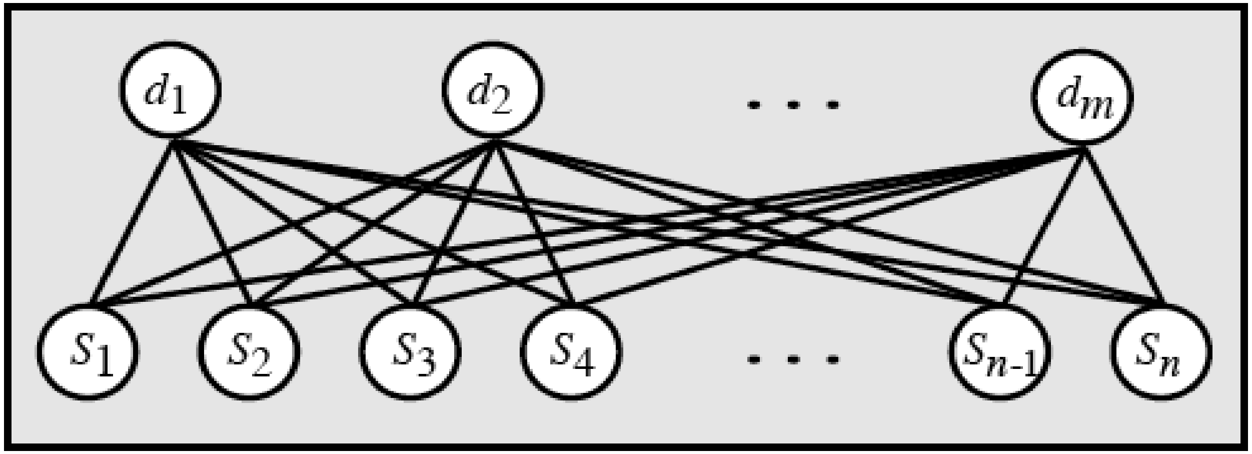
Is there any method to objectively define (in)dependences between the variables and reduce the number of parameters?

$$p(s_1, \ldots, s_n | d_i) = \prod_{i=1}^{n} p(s_j | d_i).$$



Firstly, unrealistic models "*ad-hoc*" were proposed.

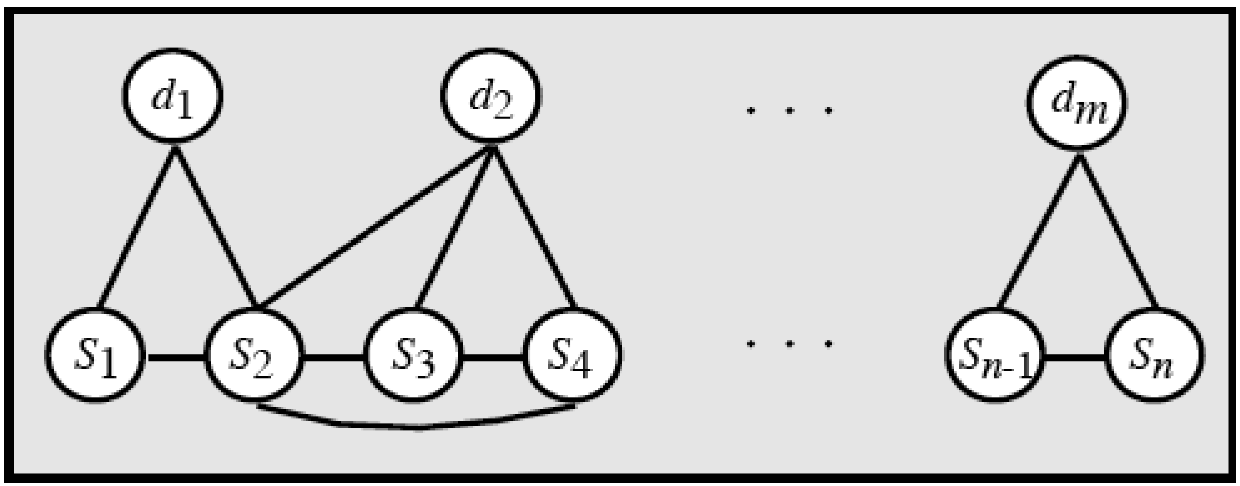
$$P(s_1, \ldots, s_n, d_1,\ldots, d_m) = P(s_1, \ldots, s_n|d_1,\ldots, d_m)\, P(d_1,\ldots, d_m)$$



Is there any method to objectively define (in)dependences between the variables and reduce the number of parameters?

$$p(s_1,\ldots,s_n|d_i) = \prod_{i=1}^{n} p(s_j|d_i).$$

YES



$X_1 \wedge X_2$ *independent*

$$P(X_1 \cap X_2) = P(X_1) * P(X_2)$$

$$P(X_1 \vee X_2) = P(X_1) \wedge P(X_2 \vee X_1) = P(X_2)$$

x	y	z	$p(x,y,z)$
0	0	0	0.12
0	0	1	0.18
0	1	0	0.04
0	1	1	0.16
1	0	0	0.09
1	0	1	0.21
1	1	0	0.02
1	1	1	0.18

To explicitly define the joint probability function is not possible in most of the real cases due to the large amount of parameters (e.g. 10^{25} parameters for 100 variables).

Bayes's Theorem → Factorization

$$P(X_i \vee B) = \frac{P(B \vee X_i) P(X_i)}{\sum_{j=1}^{n} P(B \vee X_j) P(X_j)}$$

$$\underbrace{B \wedge X_i \, independent \Rightarrow P(X_i \vee B) = P(X_i) \wedge P(B \vee X_i) = P(B)}$$

Can we build a model including some pre-defined independences? YES

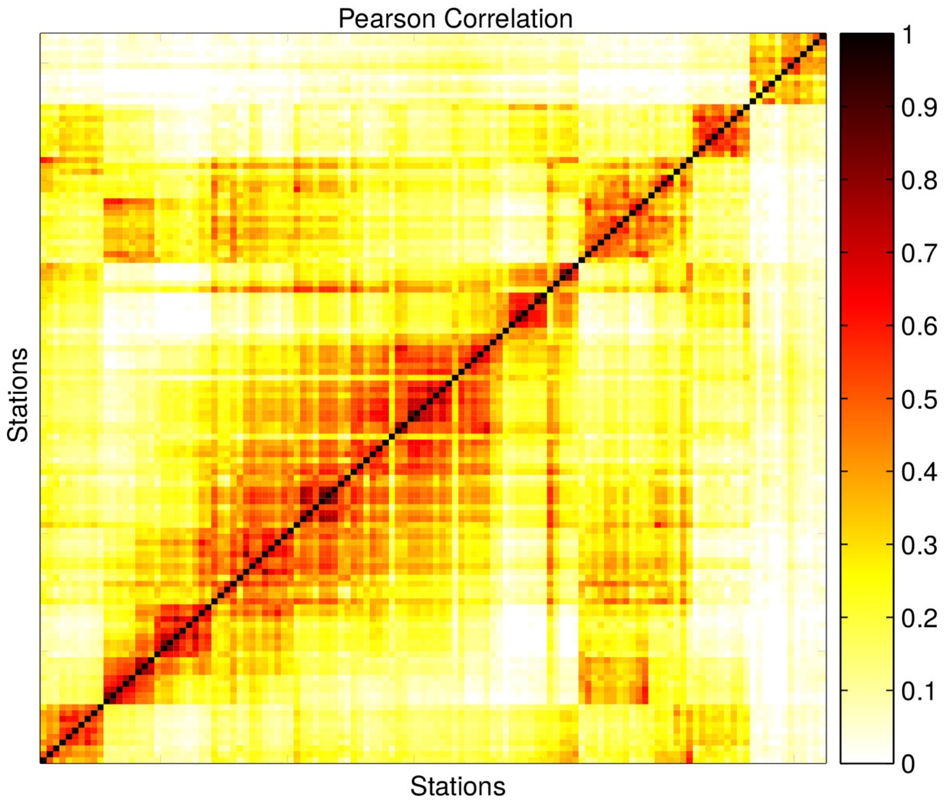
$$\boxed{I(X_3, X_1|X_2) \text{ and } I(X_4, \{X_1, X_3\}|X_2).} \Rightarrow \begin{cases} p(x_3|x_1, x_2) = p(x_3|x_2), \\ p(x_4|x_1, x_2, x_3) = p(x_4|x_2). \end{cases}$$

4 tables instead 16 → $\boxed{p(x_1, \ldots, x_4) = p(x_1)p(x_2|x_1)p(x_3|x_2)p(x_4|x_2).}$

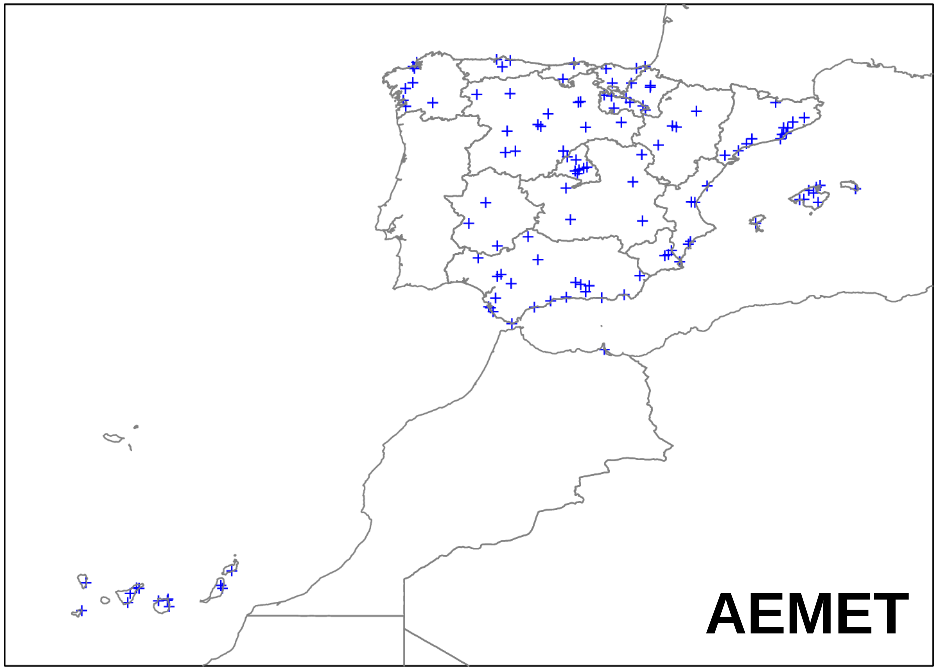
Information theory has defined several measures to evaluate de dependence between two variables:

Pearson Correlation

$$r(X,Y)=\frac{\sigma_{xy}}{\sigma_x \sigma_y}$$



Precip. Occurrence



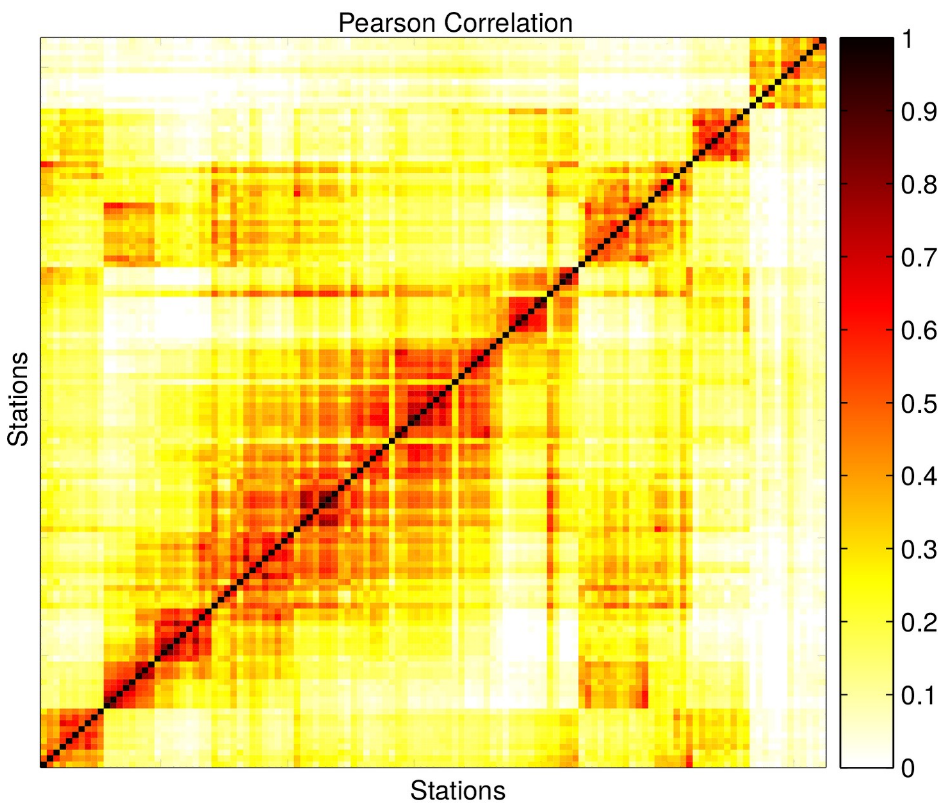
Information theory has defined several measures to evaluate de dependence between two variables:

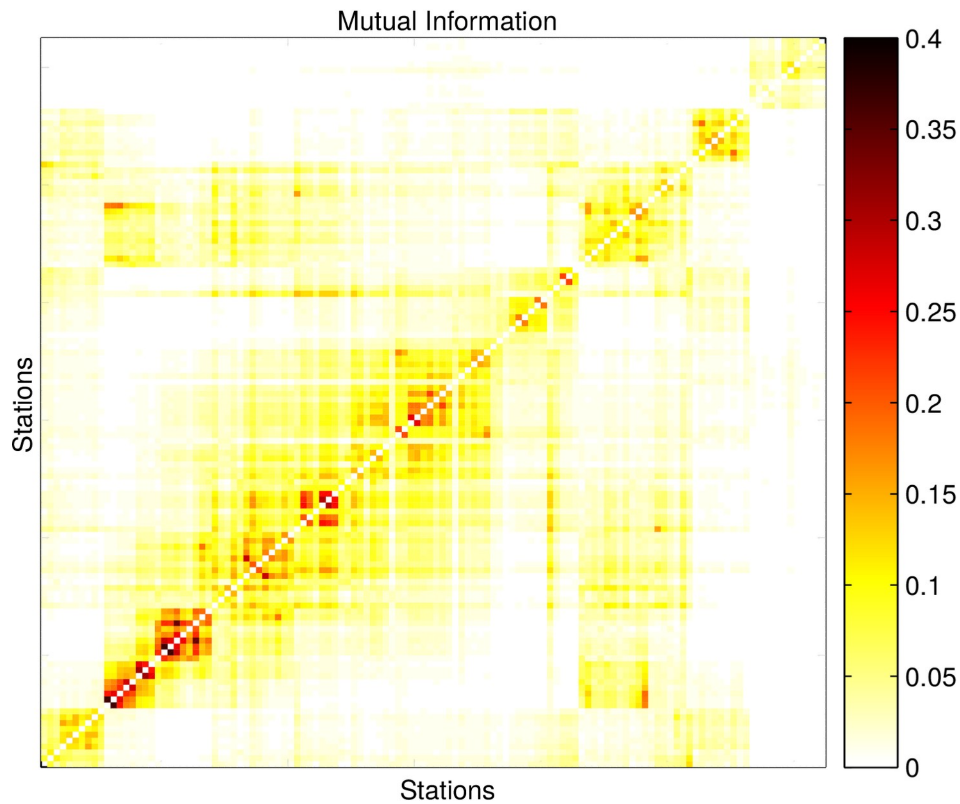
Pearson Correlation

$$r(X,Y)=\frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

Mutual Information

$$I(X,Y)=\sum_{x\in X}\sum_{y\in Y} P(x,y)\log_b\left(\frac{P(x,y)}{(P(x)P(y))}\right)$$





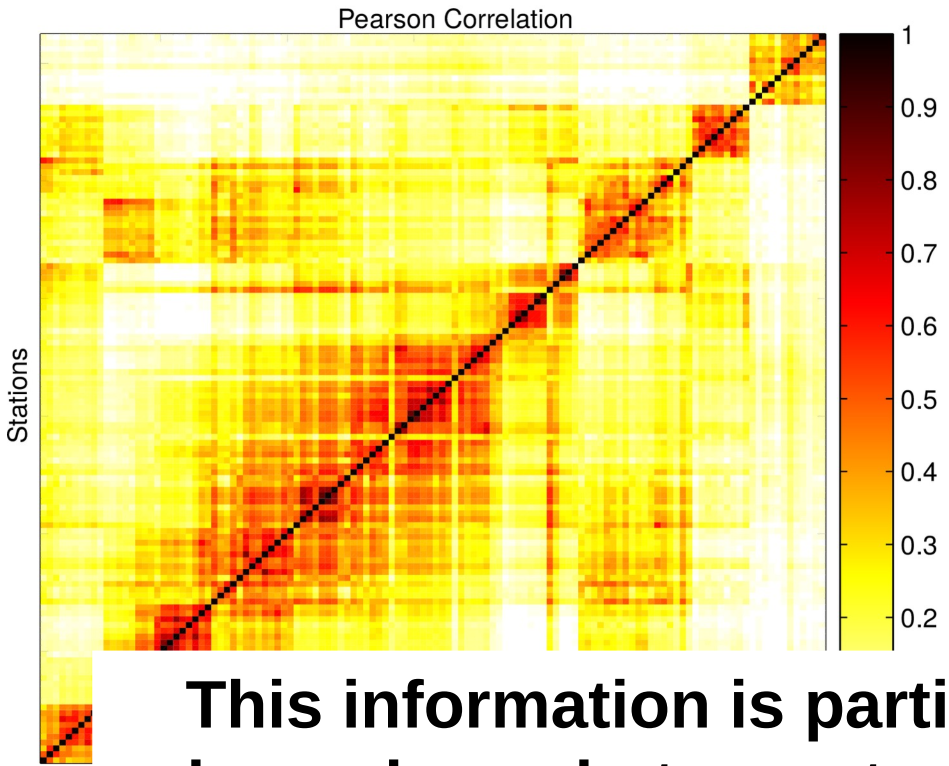
Information theory has defined several measures to evaluate de dependence between two variables:

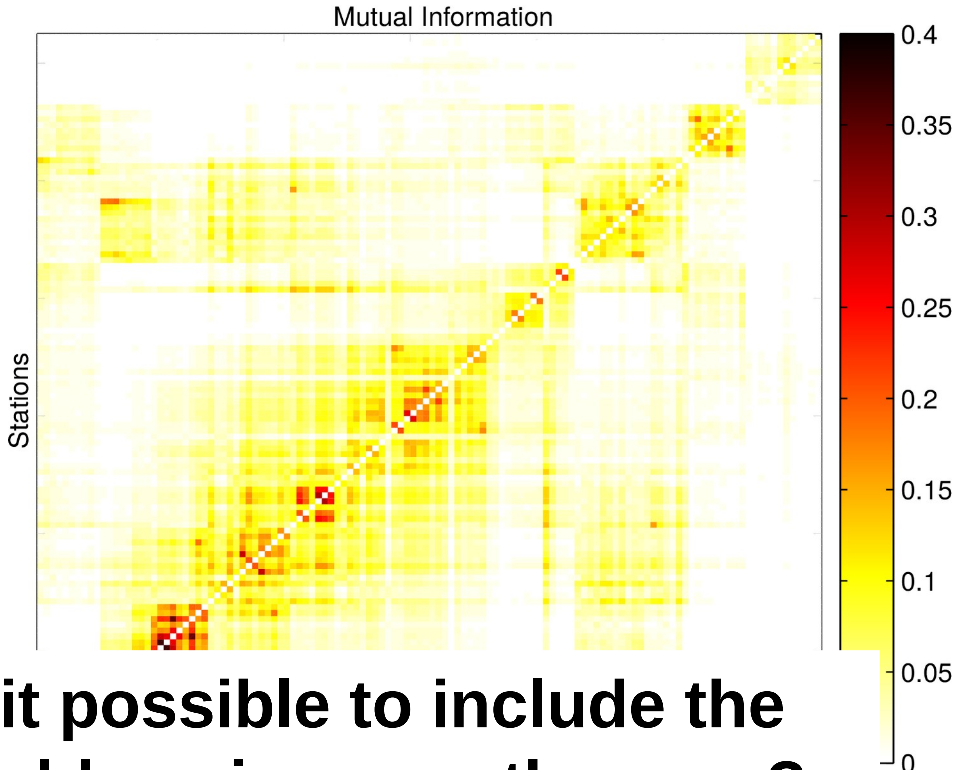
Pearson Correlation

$$r(X,Y)=\frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

Mutual Information

$$I(X,Y)=\sum_{x\in X}\sum_{y\in Y} P(x,y)\log_b\left(\frac{P(x,y)}{(P(x)P(y))}\right)$$





This information is partial, is it possible to include the dependence between two variables given another one?

Information theory has defined several measures to evaluate de dependence between two variables:

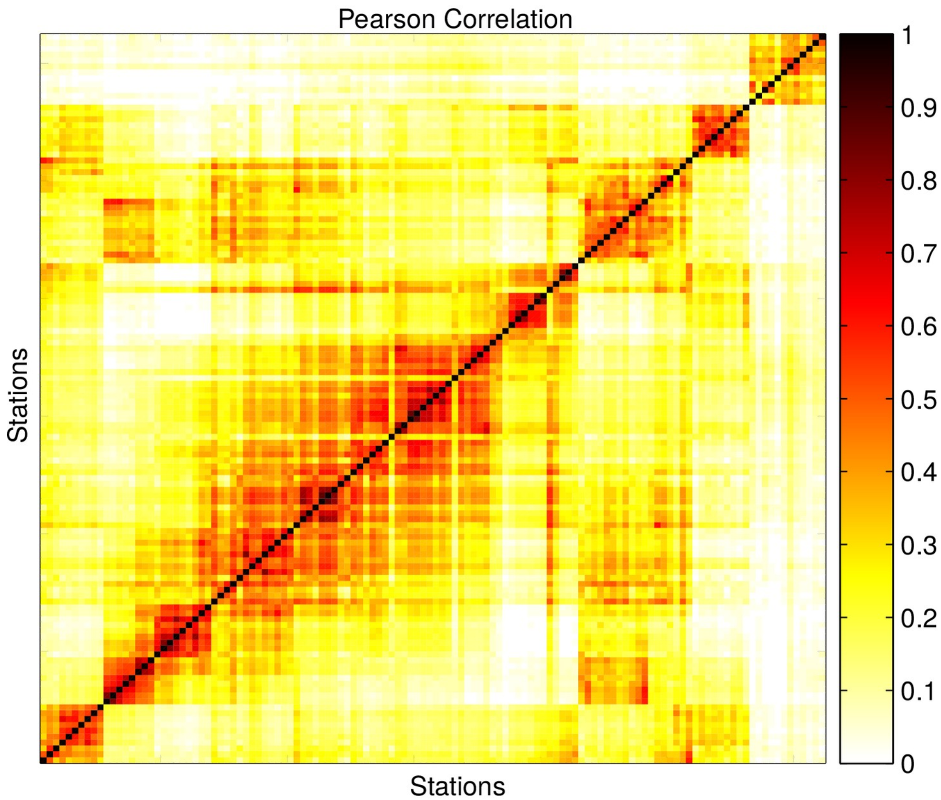
Pearson Correlation

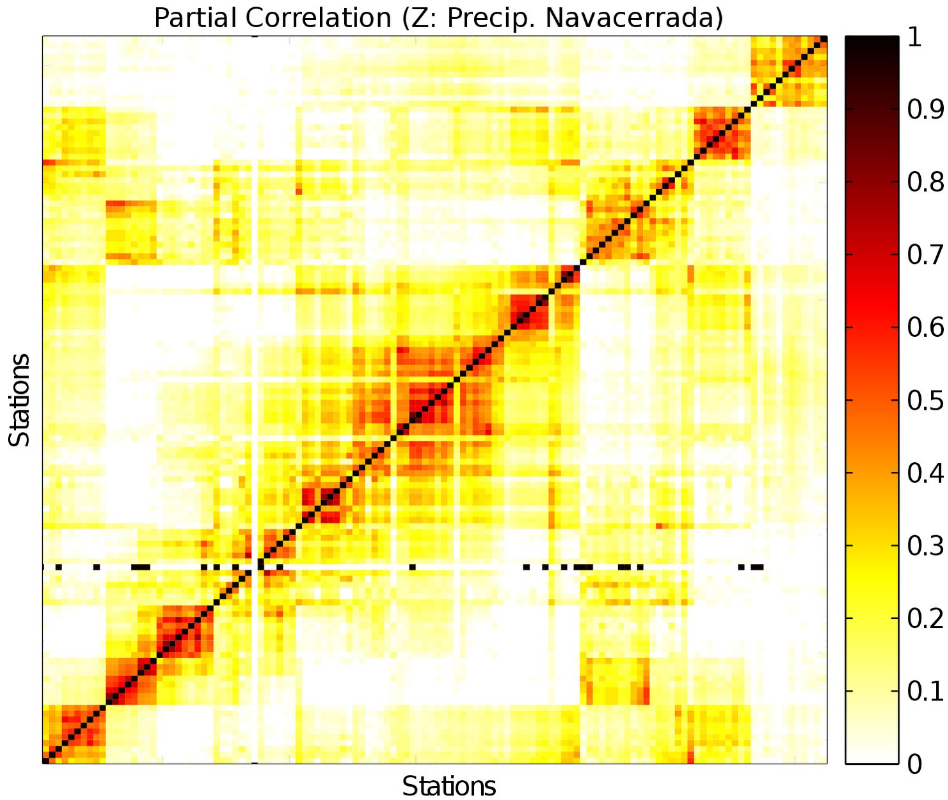
$$r(X,Y)=\frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

Partial Correlation

$$r(X,Y \vee Z)=\frac{r(X,Y)-r(X,Z)*r(X,Z)}{\sqrt{1-r(X,Z)^2}*\sqrt{1-r(Y,Z)^2}}$$

Recursively applied to Z involving more than one variable





Information theory has defined several measures to evaluate de dependence between two variables:

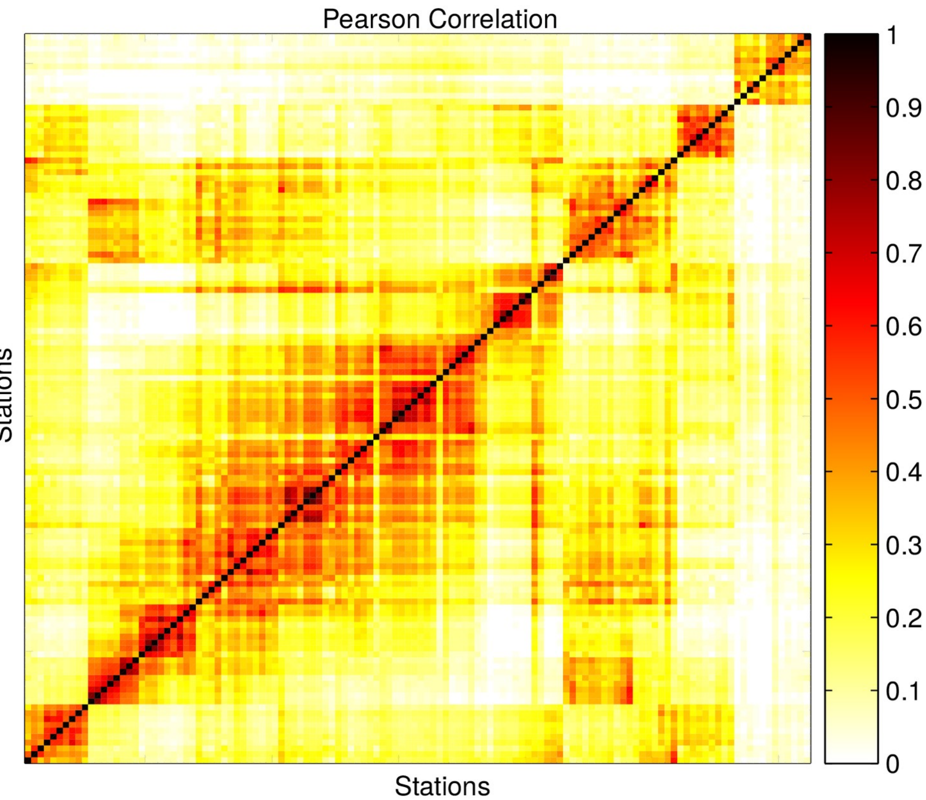
Pearson Correlation

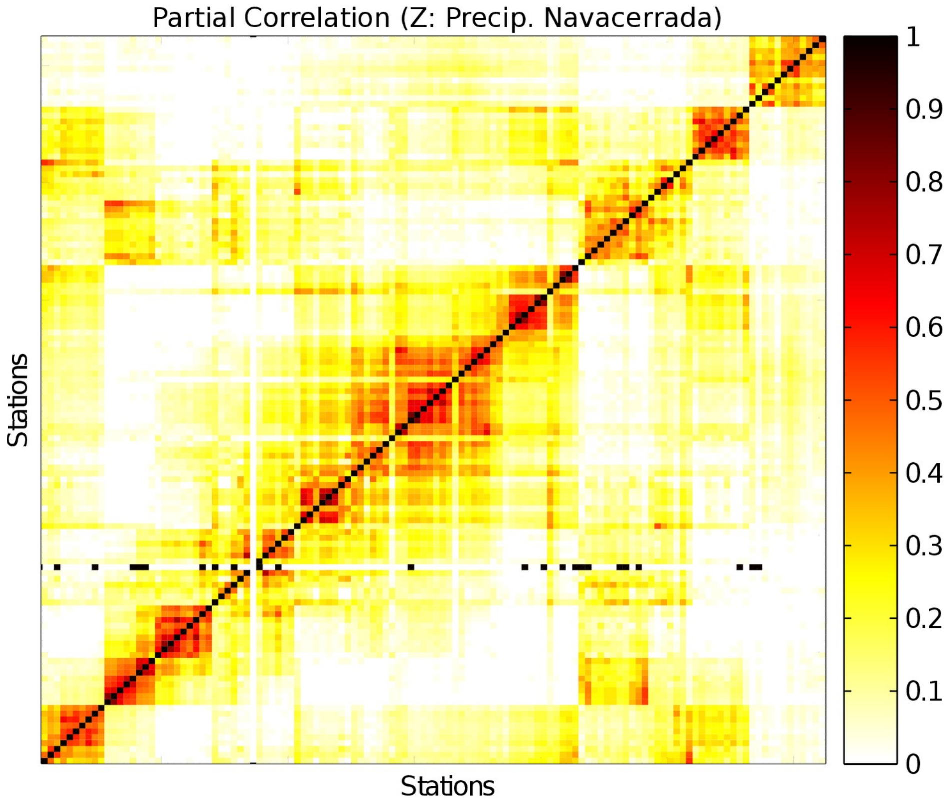
$$r(X,Y)=\frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

Partial Correlation

$$r(X,Y \vee Z)=\frac{r(X,Y)-r(X,Z) * r(X,Z)}{\sqrt{1-r(X,Z)^2} * \sqrt{1-r(Y,Z)^2}}$$

How to condition to a particular event (e.g. pr > 1mm in Navacerrada)?





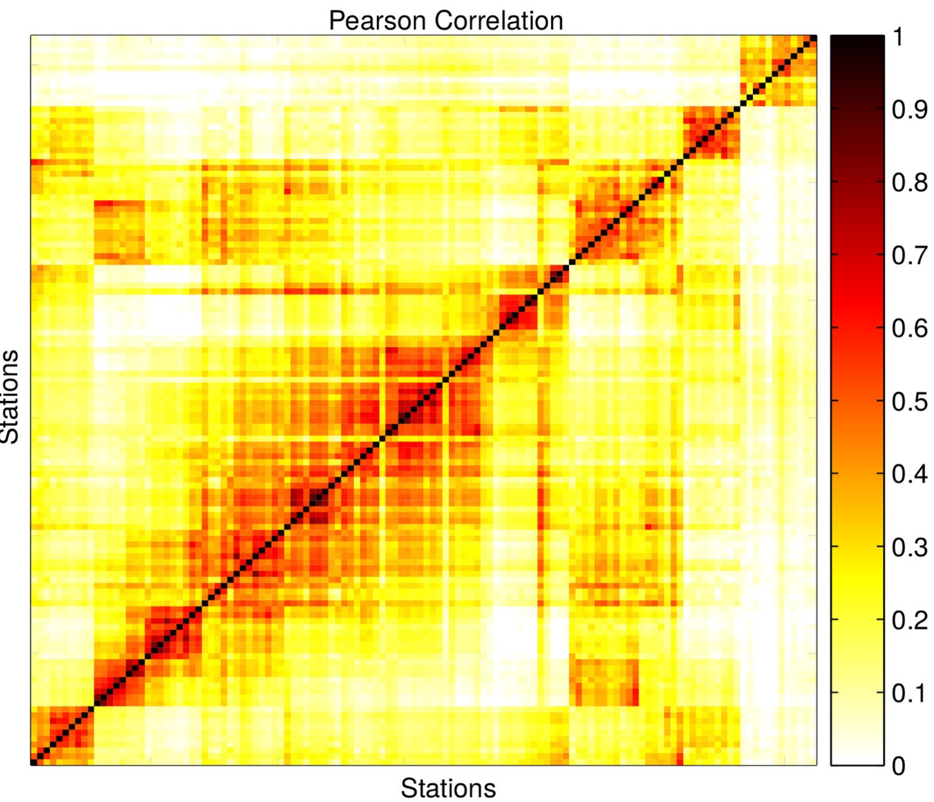
Information theory has defined several measures to evaluate de dependence between two variables:

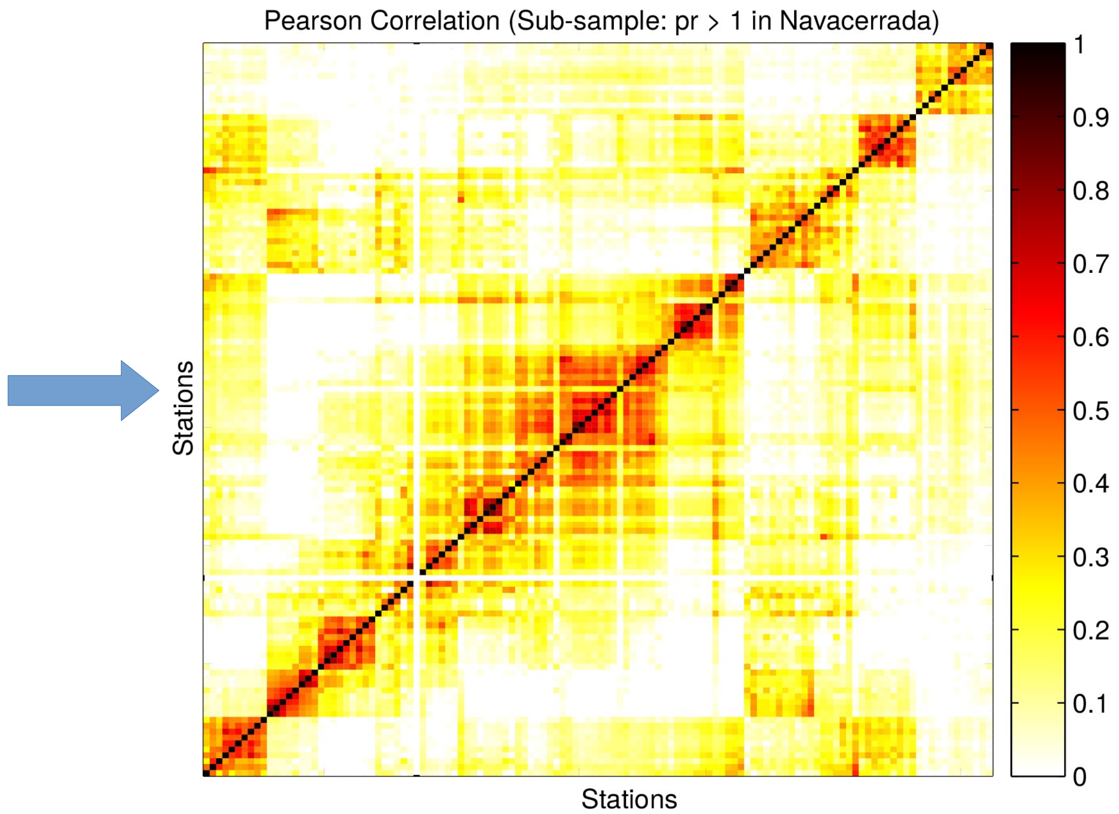
Pearson Correlation

$$r(X, Y) = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

Filtering by the occurrence of the target event

How to condition to a particular event (e.g. pr > 1mm in Navacerrada)?





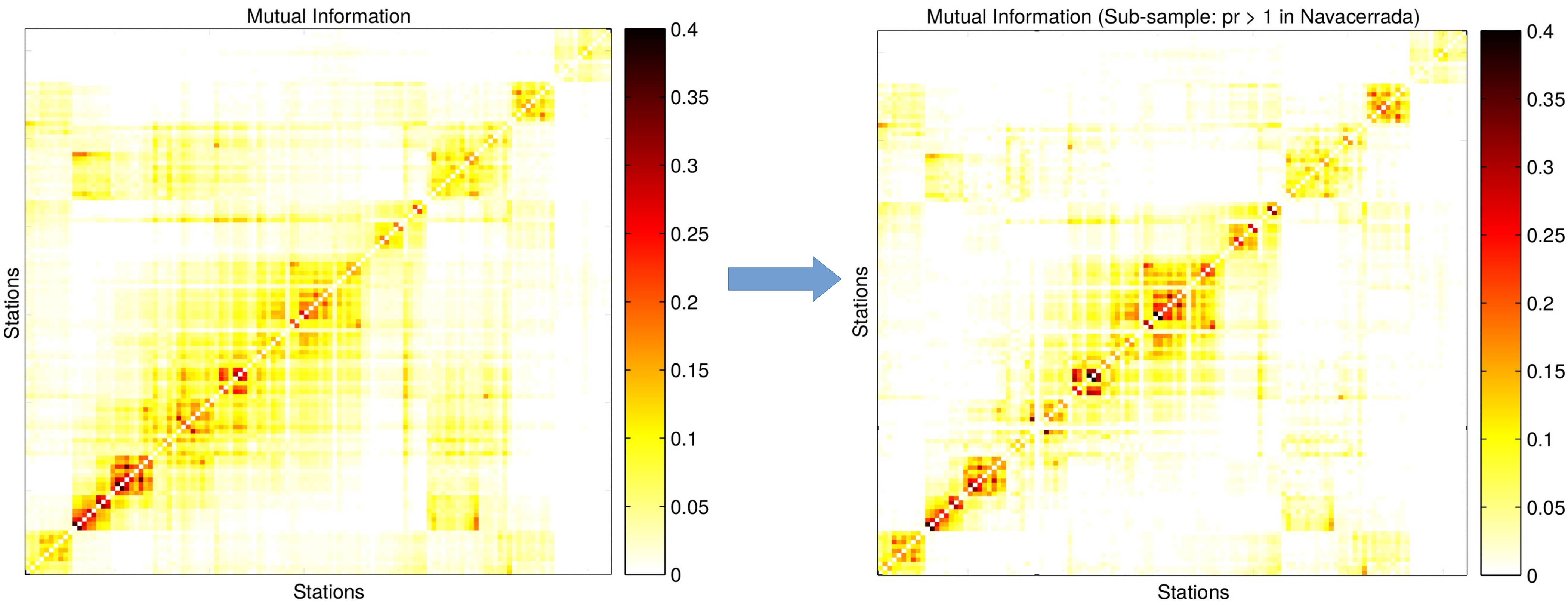
Information theory has defined several measures to evaluate de dependence between two variables:

Mutual Information

$$I(X,Y)=\sum_{x\in X}\sum_{y\in Y}P(x,y)\log_b\left(\frac{P(x,y)}{(P(x)P(y))}\right)$$

Filtering by the occurrence of the target event

How to condition to a particular event (e.g. pr > 1mm in Navacerrada)?



Information theory has defined several measures to evaluate de dependence between two variables:

Mutual Information

$$I(X,Y)=\sum_{x\in X}\sum_{y\in Y}P(x,y)\log_b\left(\frac{P(x,y)}{(P(x)P(y))}\right)$$

Filtering by the occurrence of the target event

Remember the definition of condioned probability

$$X=\{NW\}\Rightarrow P(Y\vee X)=\frac{P(Y,X)}{P(X)}$$

	Anual		Invierno		Primavera		Verano		Otoño	
	S	Ll	S	Ll	S	Ll	S	Ll	S	Ll
NE	1014	516	190	99	287	166	360	162	177	89
SE	64	57	24	18	6	4	1	9	33	26
SW	225	661	98	223	18	119	15	71	94	248
NW	288	825	49	150	95	277	108	251	36	147
Total	1591	2059	361	490	406	566	484	493	340	510

Information theory has defined several measures to evaluate de dependence between two variables:

Pearson Correlation

$$r(X,Y)=\frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

Mutual Information

$$I(X,Y)=\sum_{x\in X}\sum_{y\in Y} P(x,y)\log_b\left(\frac{P(x,y)}{(P(x)P(y))}\right)$$

This information is partial and does not include the dependence between two variables given another one!!!!

A definition of conditional (in)dependence surges naturally from the conditional probability definition

$$P_Z(Y \vee X)=P(Y \vee X,Z)=P(Y \vee Z)=P_Z(Y) \Rightarrow I(X,Y \vee Z)$$

x	y	z	$p(x,y,z)$
0	0	0	0.12
0	0	1	0.18
0	1	0	0.04
0	1	1	0.16
1	0	0	0.09
1	0	1	0.21
1	1	0	0.02
1	1	1	0.18

To explicitly define the joint probability function is not possible in most of the real cases due to the large amount of parameters (e.g. 10^{25} parameters for 100 variables).

Bayes's Theorem → Factorization

$$P(X_i \vee B) = \frac{P(B \vee X_i) P(X_i)}{\sum_{j=1}^{n} P(B \vee X_j) P(X_j)}$$

$$\underbrace{B \wedge X_i \, independent \Rightarrow P(X_i \vee B) = P(X_i) \wedge P(B \vee X_i) = P(B)}$$

Can we build a model including some pre-defined independences? **YES**

$$\boxed{I(X_3, X_1 | X_2) \text{ and } I(X_4, \{X_1, X_3\} | X_2).}$$

$$\left\{ \begin{array}{l} p(x_3|x_1, x_2) = p(x_3|x_2), \\ p(x_4|x_1, x_2, x_3) = p(x_4|x_2). \end{array} \right.$$

4 tables instead 16 →

$$\boxed{p(x_1, \ldots, x_4) = p(x_1) p(x_2|x_1) p(x_3|x_2) p(x_4|x_2).}$$

$$P_Z(Y \vee X) = P(Y \vee X, Z) = P(Y \vee Z) = P_Z(Y) \Rightarrow I(X, Y \vee Z)$$

	Anual		Invierno		Primavera		Verano		Otoño	
	S	Ll	S	Ll	S	Ll	S	Ll	S	Ll
NE	1014	516	190	99	287	166	360	162	177	89
SE	64	57	24	18	6	4	1	9	33	26
SW	225	661	98	223	18	119	15	71	94	248
NW	288	825	49	150	95	277	108	251	36	147
Total	1591	2059	361	490	406	566	484	493	340	510

P (Ll / Primavera) = ?
P (Ll / Invierno) = ?

P (Ll) = ?

```
## States of the variables:
estados.Moon <- c("Anual","Llena","Menguante","Creciente","Nueva")

## Table of Absolute frequencies:
table2.freq <- array(c(1014, 64, 225, 288, 255, 12, 59, 51, 208, 16, 65, 77,
                       297, 22, 58, 82, 254, 14, 43, 78, 516, 57, 661, 825, 137, 12, 165, 192,
                       106, 16, 166, 231, 132, 12, 175, 225, 141, 17, 155, 177), dim = c(4,5,2),
              dimnames = list(W=estados.Wind, S=estados.Moon, P = estados.Precip))

## Obtain the probabilities:
```

	Anual		Llena		C. Menguante		C. Creciente		Nueva	
	S	Ll	S	Ll	S	Ll	S	Ll	S	Ll
NE	1014	516	255	137	208	106	297	132	254	141
SE	64	57	12	12	16	16	22	12	14	17
SW	225	661	59	165	65	166	58	175	43	155
NW	288	825	51	192	77	231	82	225	78	177
Total	1591	2059	377	506	366	519	459	544	389	490

P (Ll) = ?
P (Ll / Cc) = ?
P (Ll / Ln) = ?
P (Ll / Cm) = ?
P (Ll / LL) = ?

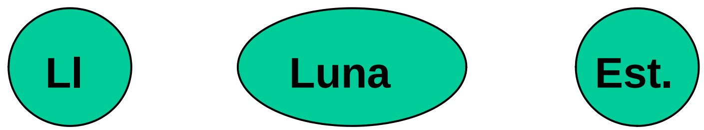
$$P_Z(Y \vee X) = P(Y \vee X, Z) = P(Y \vee Z) = P_Z(Y) \Rightarrow I(X, Y \vee Z)$$

	Anual		Invierno		Primavera		Verano		Otoño	
	S	Ll	S	Ll	S	Ll	S	Ll	S	Ll
NE	1014	516	190	99	287	166	360	162	177	89
SE	64	57	24	18	6	4	1	9	33	26
SW	225	661	98	223	18	119	15	71	94	248
NW	288	825	49	150	95	277	108	251	36	147
Total	1591	2059	361	490	406	566	484	493	340	510

P (Ll / Primavera) = 0.576
P (Ll / Invierno) = 0.582

Direct independence variables → Involve only two variables

P (Ll) = 0.564



	Anual		Llena		C. Menguante		C. Creciente		Nueva	
	S	Ll	S	Ll	S	Ll	S	Ll	S	Ll
NE	1014	516	255	137	208	106	297	132	254	141
SE	64	57	12	12	16	16	22	12	14	17
SW	225	661	59	165	65	166	58	175	43	155
NW	288	825	51	192	77	231	82	225	78	177
Total	1591	2059	377	506	366	519	459	544	389	490

P (Ll) = 0.564
P (Ll / Cc) = 0.557
P (Ll / Ln) = 0.542
P (Ll / Cm) = 0.586
P (Ll / LL) = 0.573

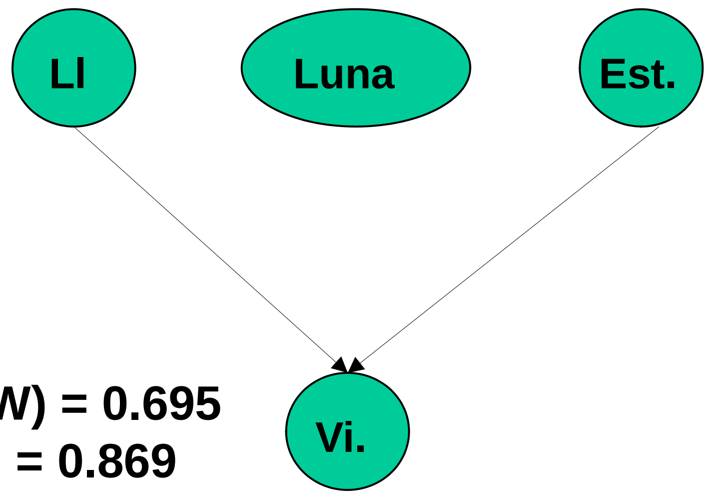
$$P_Z(Y \vee X) = P(Y \vee X, Z) = P(Y \vee Z) = P_Z(Y) \Rightarrow I(X, Y \vee Z)$$

	Anual		Invierno		Primavera		Verano		Otoño	
	S	Ll	S	Ll	S	Ll	S	Ll	S	Ll
NE	1014	516	190	99	287	166	360	162	177	89
SE	64	57	24	18	6	4	1	9	33	26
SW	225	661	98	223	18	119	15	71	94	248
NW	288	825	49	150	95	277	108	251	36	147
Total	1591	2059	361	490	406	566	484	493	340	510

P (Ll / Primavera) = 0.576
P (Ll / Invierno) = 0.582

Direct independence variables → Involve only two variables

P (Ll) = 0.564



P (Ll / Primavera, SW) = 0.695
P (Ll / Invierno, SW) = 0.869

Conditional dependence between rainfall and season, given the wind

P (Ll) = 0.564
P (Ll / Cc) = 0.557
P (Ll / Ln) = 0.542
P (Ll / Cm) = 0.586
P (Ll / LL) = 0.573

x	y	z	$p(x, y, z)$
0	0	0	0.12
0	0	1	0.18
0	1	0	0.04
0	1	1	0.16
1	0	0	0.09
1	0	1	0.21
1	1	0	0.02
1	1	1	0.18

To explicitly define the joint probability function is not possible in most of the real cases due to the large amount of parameters (e.g. 10^{25} parameters for 100 variables).

Bayes's Theorem → Factorization

$$P(X_i \vee B) = \frac{P(B \vee X_i) P(X_i)}{\sum_{j=1}^{n} P(B \vee X_j) P(X_j)}$$

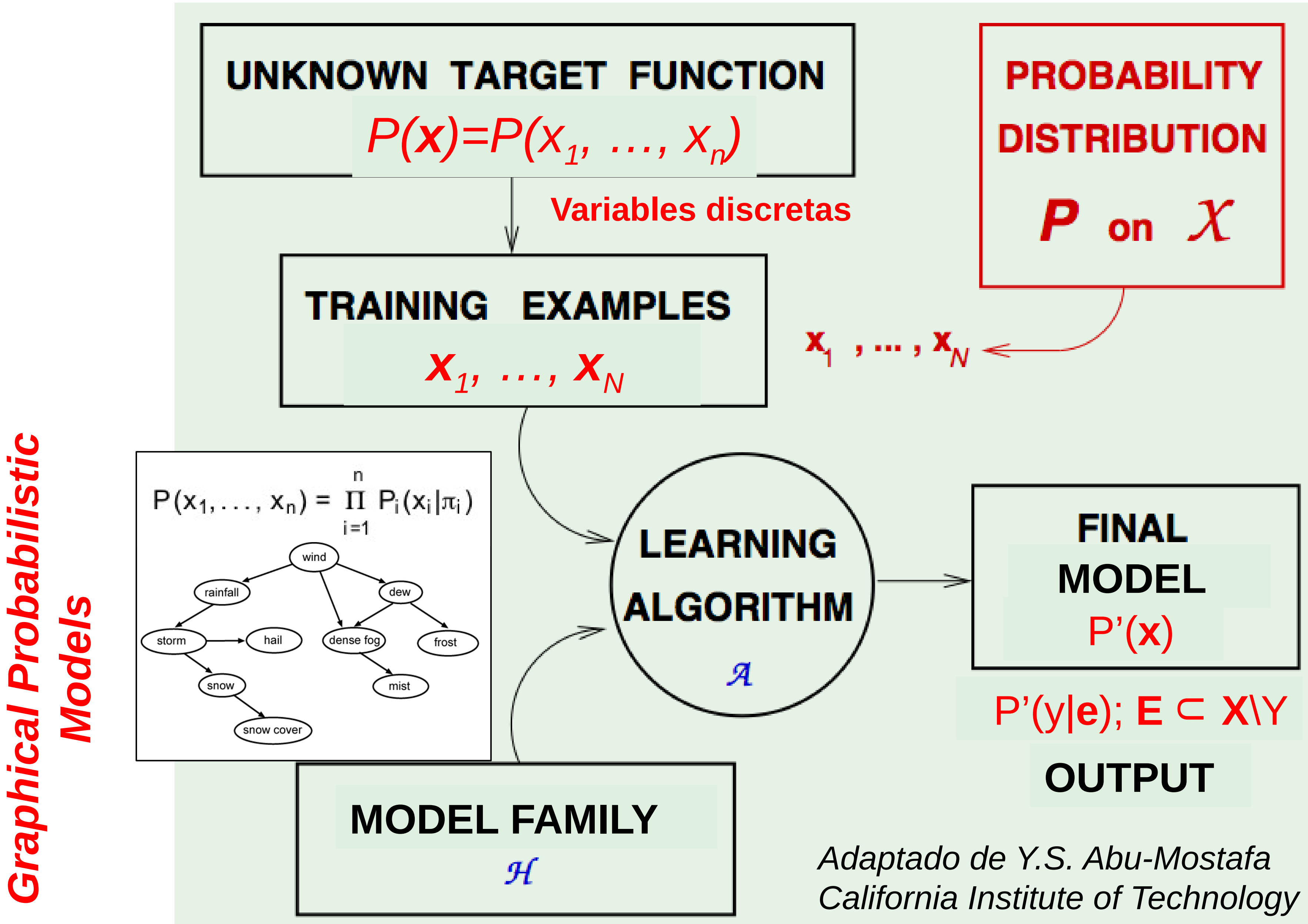
$$\underbrace{B \wedge X_i \, independent \Rightarrow P(X_i \vee B) = P(X_i) \wedge P(B \vee X_i) = P(B)}$$

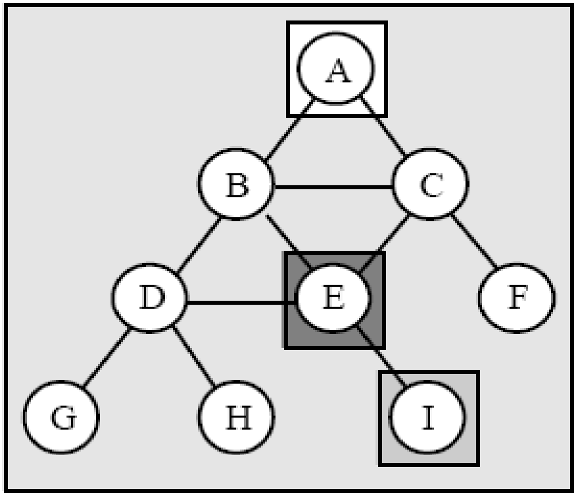
Can we build a model including some pre-defined independences? YES

$$\boxed{I(X_3, X_1 | X_2) \text{ and } I(X_4, \{X_1, X_3\} | X_2).}$$

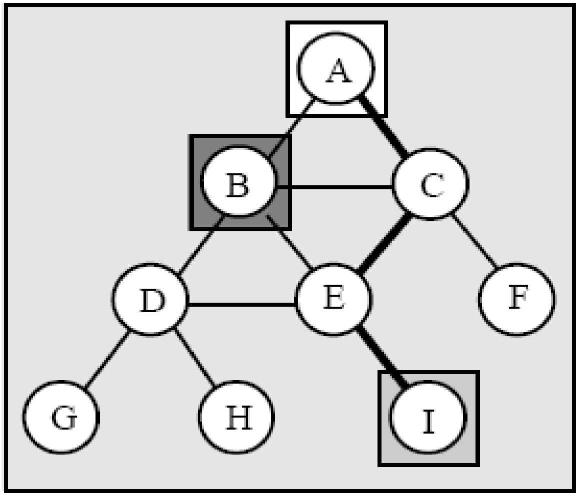
$$\boxed{p(x_1, \ldots, x_4) = p(x_1) p(x_2|x_1) p(x_3|x_2) p(x_4|x_2).}$$

Can we obtain efficiently these (in)dependences? → Graphs


UNKNOWN TARGET FUNCTION
$P(\mathbf{x})=P(x_1, \ldots, x_n)$
Variables discretas
PROBABILITY DISTRIBUTION
$\boldsymbol{P}$ on $\mathcal{X}$
TRAINING EXAMPLES
$\mathbf{x}_1, \ldots, \mathbf{x}_N$
$\mathbf{x}_1, \ldots, \mathbf{x}_N$
$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P_i(x_i|\pi_i)$
wind
rainfall
dew
storm
hail
dense fog
frost
snow
mist
snow cover
Graphical Probabilistic Models
LEARNING ALGORITHM
$\mathcal{A}$
FINAL MODEL
$P'(\mathbf{x})$
$P'(y|\mathbf{e})$; $\mathbf{E} \subset \mathbf{X} \backslash Y$
OUTPUT
MODEL FAMILY
$\mathcal{H}$
Adaptado de Y.S. Abu-Mostafa
California Institute of Technology



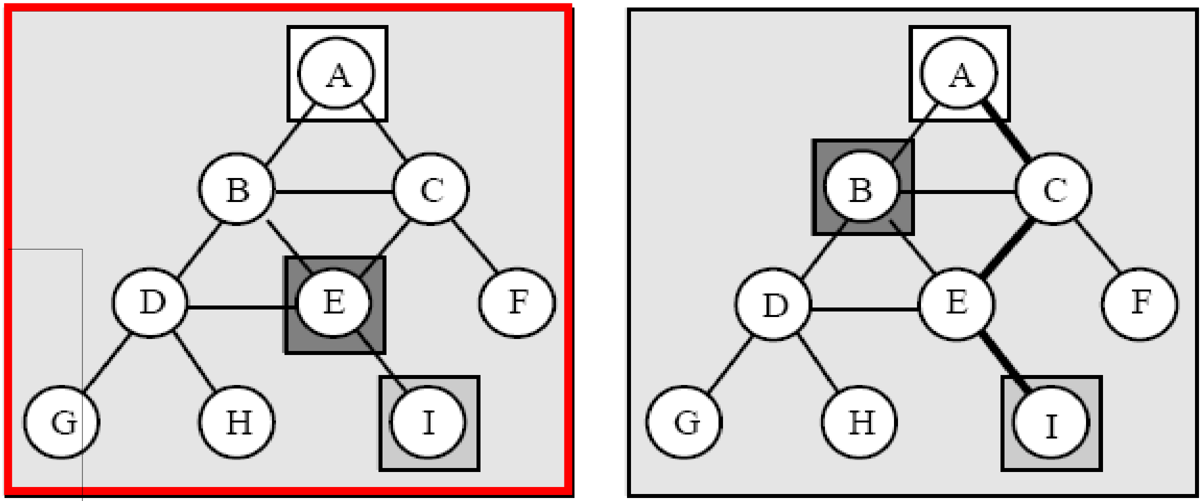
(a) $I(A, I \mid E)$



(b) $D(A, I \mid B)$

Links of the graph reflect **dependences** between the linked variables.

Non directed graphs define the conditional dependence through the **d-separation** concept.

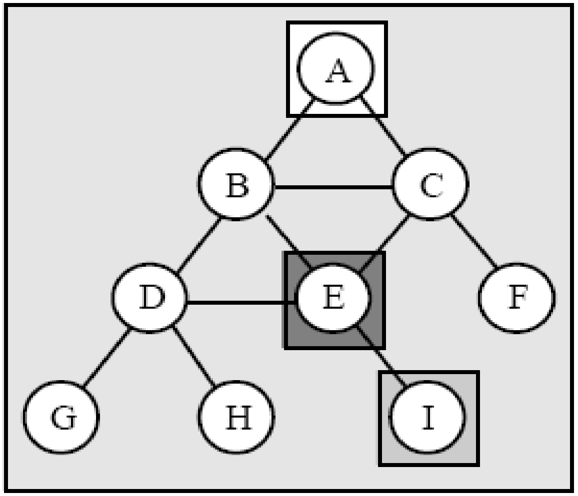


(a) $I(A, I \mid E)$ (b) $D(A, I \mid B)$

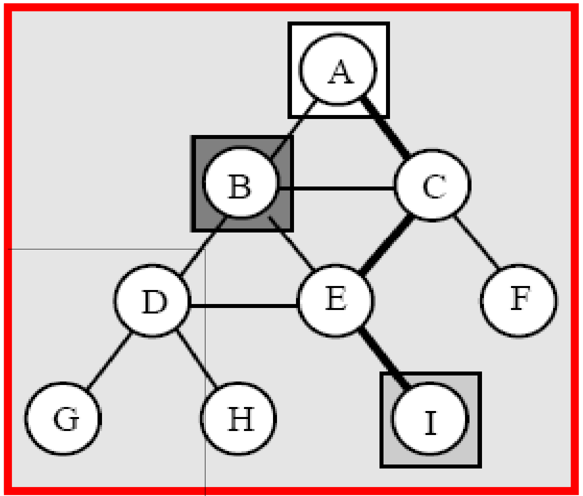
Links of the graph reflect **dependences** between the linked variables.

Non directed graphs define the conditional dependence through the **d-separation** concept.

There is not a path linking A and I not passing for E.
Thus A and I are dependent but conditional independent given E.



(a) $I(A, I \mid E)$

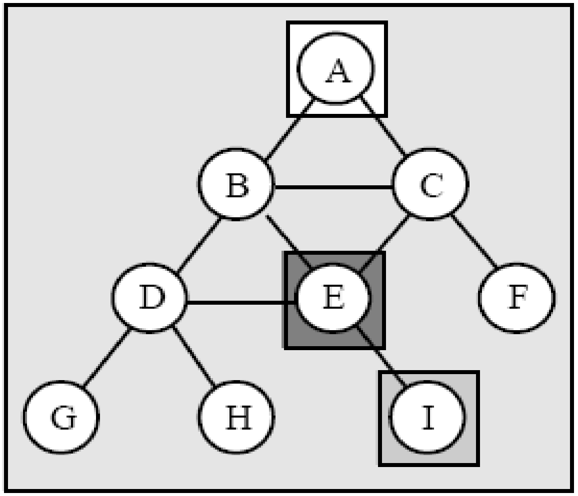


(b) $D(A, I \mid B)$

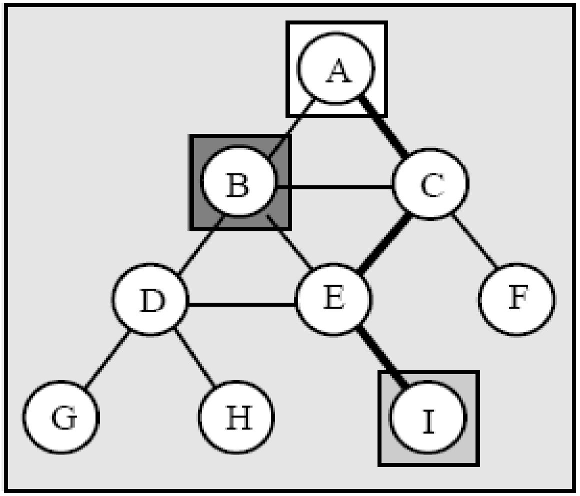
Links of the graph reflect **dependences** between the linked variables.

Non directed graphs define the conditional dependence through the **d-separation** concept.

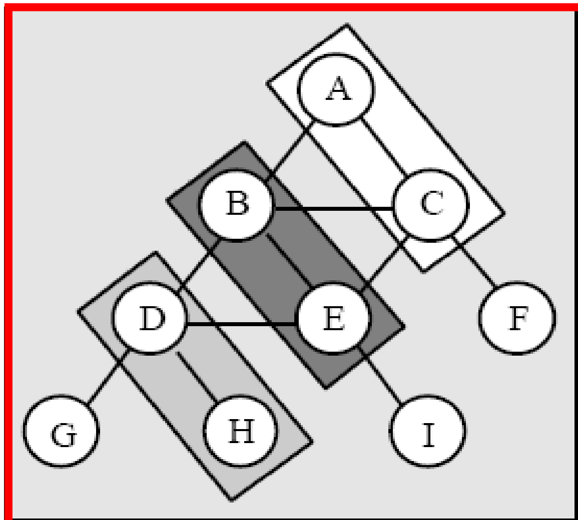
There is a path linking A and I not passing for B (A->C->E->I).
Thus A and I are dependent given B and B doesn't d-separate A and I.



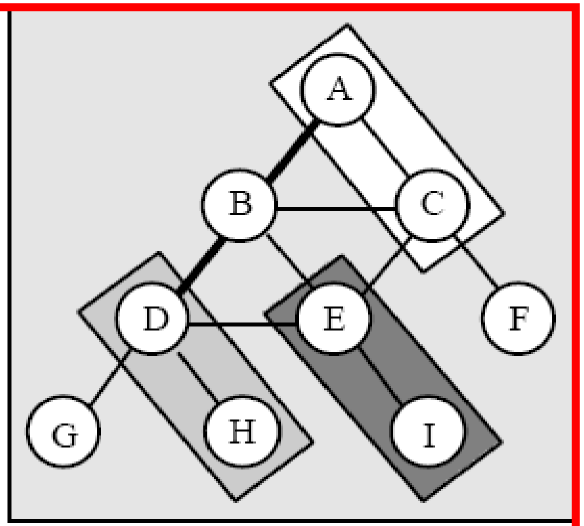
(a) $I(A, I \mid E)$



(b) $D(A, I \mid B)$



(c) $I(\{A, C\}, \{D, H\} \mid \{B, E\})$



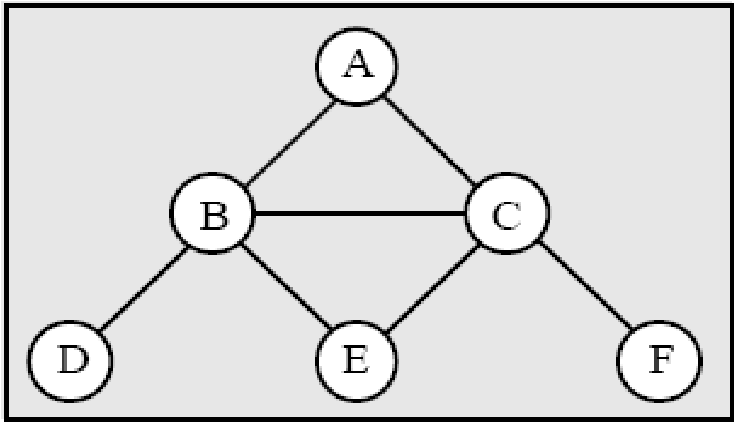
(d) $D(\{A, C\}, \{D, H\} \mid \{E, I\})$

Links of the graph reflect **dependences** between the linked variables.

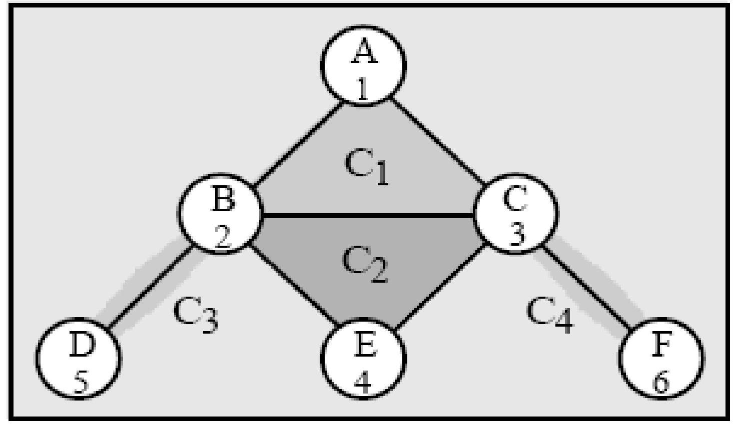
Non directed graphs define the conditional dependence through the **d-separation** concept.

D-separation is extended to set of variables.

Non-directed graphs define a graphical probabilistic model family based on the **cliques** of the graph and the factorization of the joint probability function given by them.



(a)



(b)

$$C_1 = \{A, B, C\}, \quad C_2 = \{B, C, E\},$$
$$C_3 = \{B, D\}, \quad C_4 = \{C, F\}.$$

i	Clique C_i	Separator S_i	Residual R_i
1	A, B, C	ϕ	A, B, C
2	B, C, E	B, C	E
3	B, D	B	D
4	C, F	C	F

$$\begin{aligned} p(a,b,c,d,e,f) &= \psi_1(c_1)\psi_2(c_2)\psi_3(c_3)\psi_4(c_4) \\ &= \psi_1(a,b,c)\psi_2(b,c,e)\psi_3(b,d)\psi_4(c,f). \end{aligned}$$

$$p(a,b,c,d,e,f) = \prod_{i=1}^{4} p(r_i|s_i) = p(a,b,c)p(e|b,c)p(d|b)p(f|c).$$

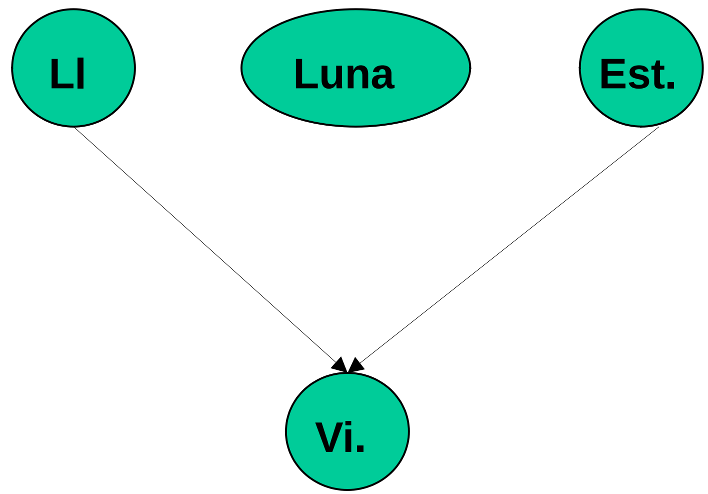
$$P_Z(Y \vee X) = P(Y \vee X, Z) = P(Y \vee Z) = P_Z(Y) \Rightarrow I(X, Y \vee Z)$$

	Anual		Invierno		Primavera		Verano		Otoño	
	S	Ll	S	Ll	S	Ll	S	Ll	S	Ll
NE	1014	516	190	99	287	166	360	162	177	89
SE	64	57	24	18	6	4	1	9	33	26
SW	225	661	98	223	18	119	15	71	94	248
NW	288	825	49	150	95	277	108	251	36	147
Total	1591	2059	361	490	406	566	484	493	340	510

P (Ll / Primavera) = 0.576
P (Ll / Invierno) = 0.582

Direct independence variables → Involve only two variables

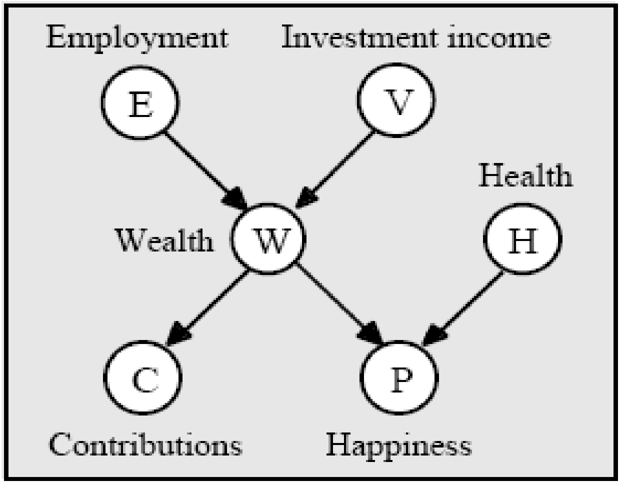
P (Ll) = 0.564

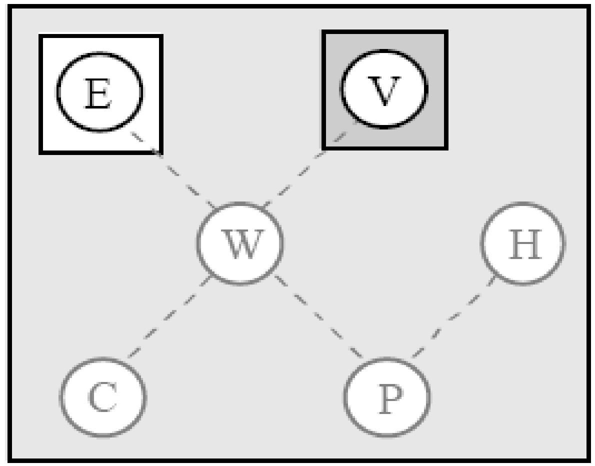


Non-directed graphs are not able to represent this kind of dependence!!!

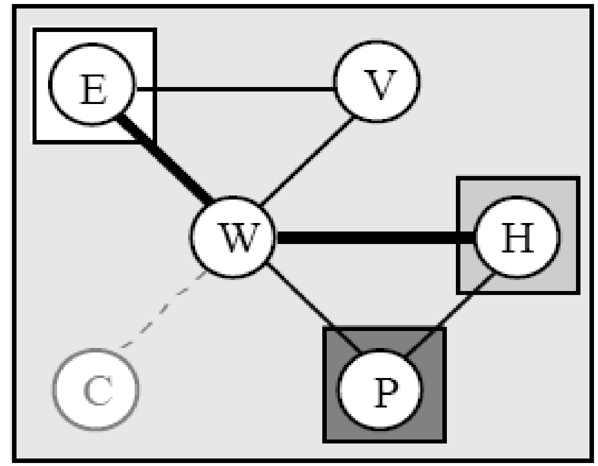
Conditional dependence between rainfall and season, given the wind

➔ **D-separation** concept for directed graphs enrich the representativity of the model → **Moral graph.**





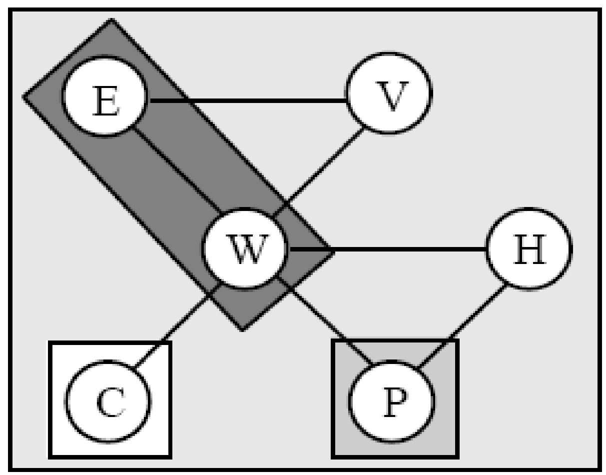
(a) $I(E, V \mid \emptyset)$



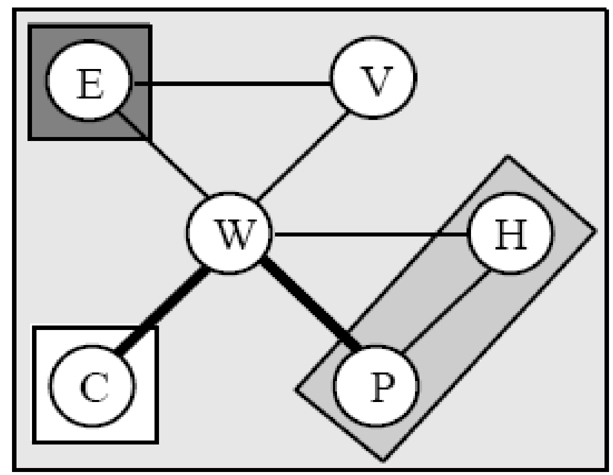
(b) $D(E, H \mid P)$

Links between variables imply probabilistic dependence **NOT CAUSALITY !!!!!!**

Causal Networks (not seen)

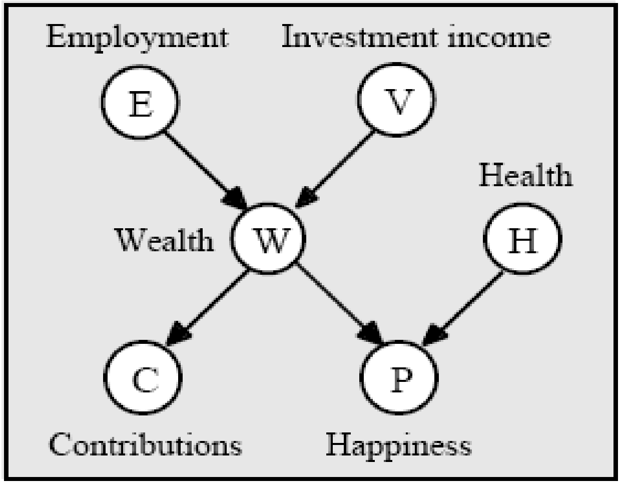


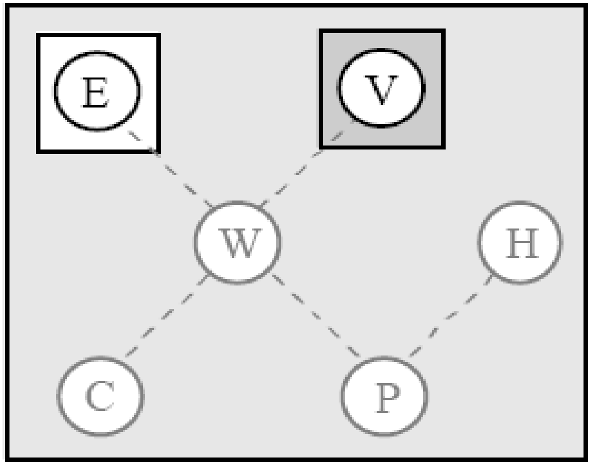
(c) $I(C, P \mid \{E, W\})$



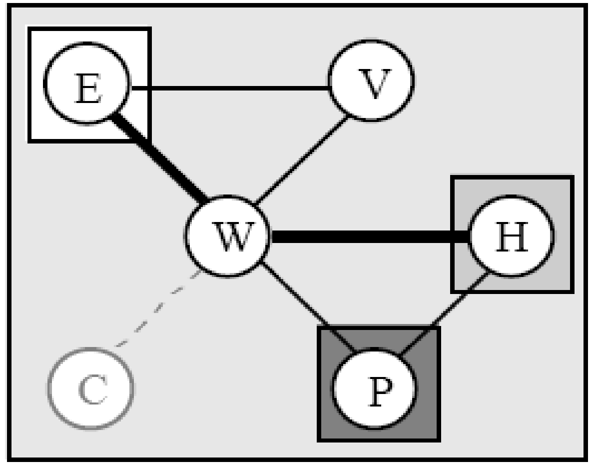
(d) $D(C, \{H, P\} \mid E)$

➔ **D-separation** concept for directed graphs enrich the representativity of the model → **Moral graph.**





(a) $I(E, V \mid \emptyset)$

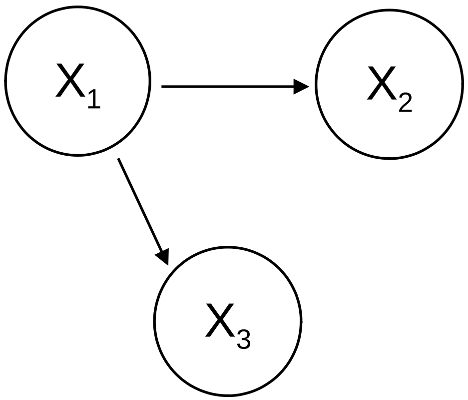


(b) $D(E, H \mid P)$

```
## Load bnlearn:
library(bnlearn)
## Defining an empty graph:
dag<-empty.graph(nodes=c("E","V","W","H","C","P"))
class(dag)
print(dag)
plot(dag)
## Adding link between nodes:
dag<-set.arc(dag,from="E",to="W")
dag<-set.arc(dag,from="V",to="W")
## Complete and plot the graph:
## Evaluate the separation included in the previous slide (See ? dsep and ?path):
```

➔ **D-separation** concept for directed graphs enrich the representativity of the model → **Moral graph.**

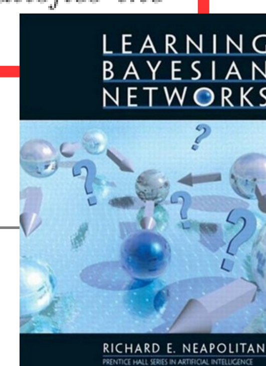
$$P(X_1, X_2, X_3) = P(X_1) P(X_2 \vee X_1) P(X_3 \vee X_1) = P(X_1, X_2) P(X_3 \vee X_1)$$



Definition 1.9 *Suppose we have a joint probability distribution P of the random variables in some set V and a DAG $\mathbb{G} = (\mathsf{V}, \mathsf{E})$. We say that $(\mathbb{G}, P)$ satisfies the **Markov condition** if for each variable $X \in \mathsf{V}$, $\{X\}$ is conditionally independent of the set of all its nondescendents given the set of all its parents.*

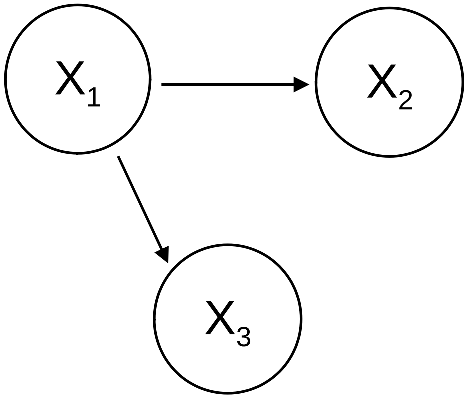
Theorem 1.5 *Let a DAG $\mathbb{G}$ be given in which each node is a random variable, and let a discrete conditional probability distribution of each node given values of its parents in $\mathbb{G}$ be specified. Then the product of these conditional distributions yields a joint probability distribution P of the variables, and $(\mathbb{G}, P)$ satisfies the Markov condition.*

Theorem 1.4 *If $(\mathbb{G}, P)$ satisfies the Markov condition, then P is equal to the product of its conditional distributions of all nodes given values of their parents, whenever these conditional distributions exist.*



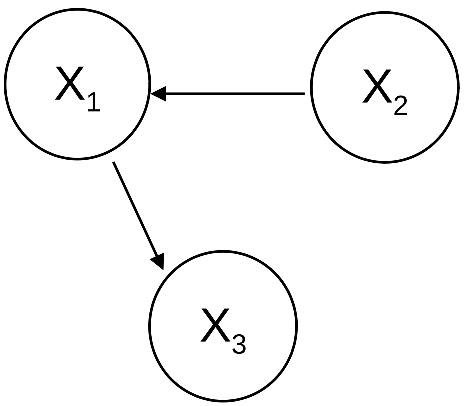
- ➔ **D-separation** concept for directed graphs enrich the representativity of the model → **Moral graph.**
- ➔ Two directed graph are **equivalents** when they lead to the same probabilistic model:

$$P\left(X_1, X_2, X_3\right) = P\left(X_1\right) P\left(X_2 \vee X_1\right) P\left(X_3 \vee X_1\right) = P\left(X_1, X_2\right) P\left(X_3 \vee X_1\right)$$



Equivalents

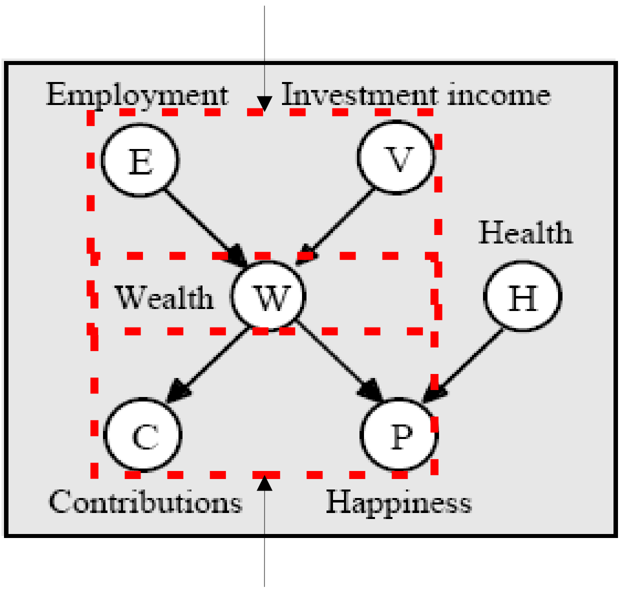
$$P\left(X_1, X_2, X_3\right) = P\left(X_2\right) P\left(X_1 \vee X_2\right) P\left(X_3 \vee X_1\right) = P\left(X_1, X_2\right) P\left(X_3 \vee X_1\right)$$

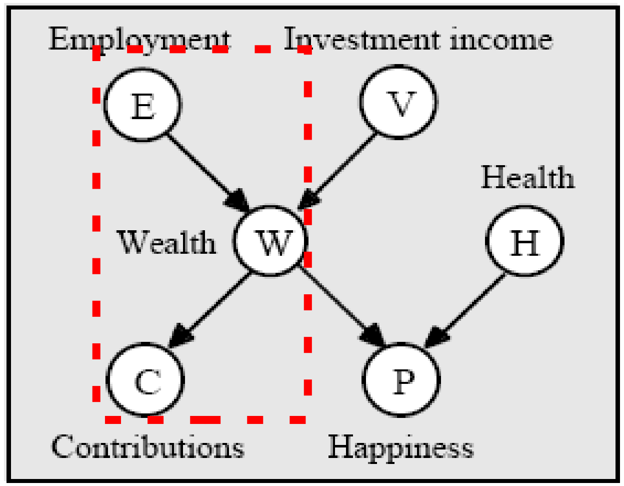


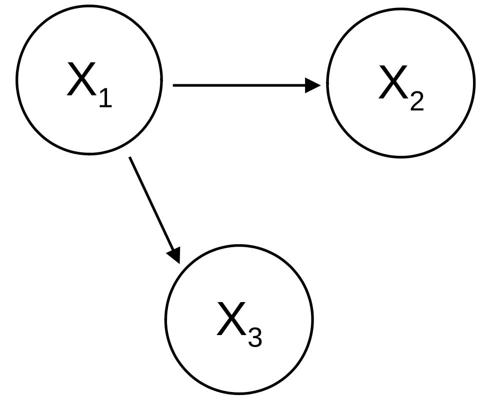
- **D-separation** concept for directed graphs enrich the representativity of the model → **Moral graph.**
- Two directed graph are **equivalents** when they lead to the same probabilistic model.
- This occurs when the **subyacent non-directed graph** is the same and include the same **V-structures**.

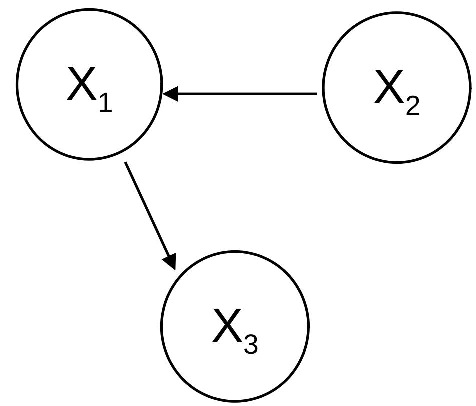
$$P(X_1, X_2, X_3) = P(X_1, X_2)\,P(X_3 \vee X_1)$$

Common effect



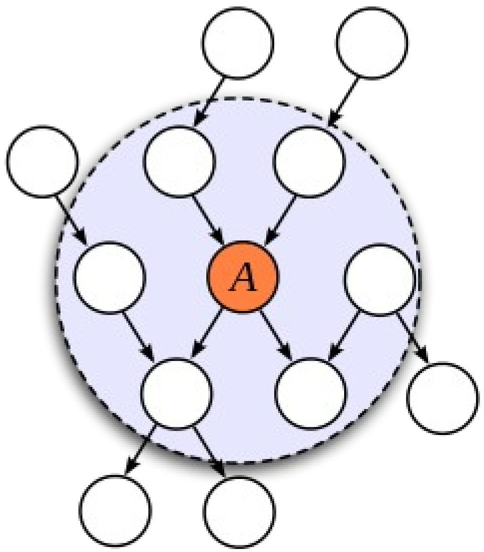






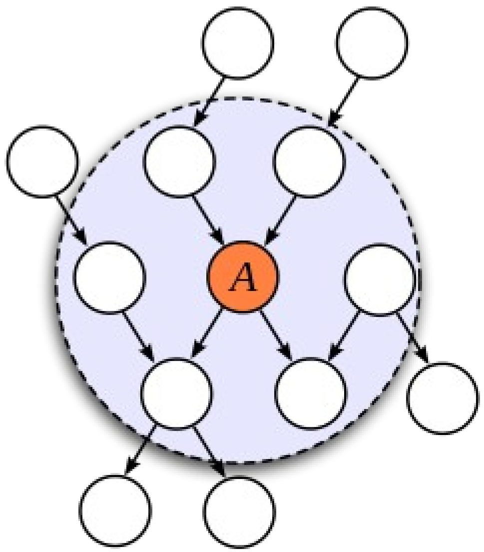
Common cause **Indirect evidential/causal effect**

- **D-separation** concept for directed graphs enrich the representativity of the model → **Moral graph.**
- Two directed graph are **equivalents** when they lead to the same probabilistic model.
- This occurs when the **subyacent non-directed graph** is the same and include the same **V-structures**.
- The **Skeleton** of the graph is the undirected graph underlying.
- The **Markov Blanket** of a node ***A*** is the set of nodes that completely separates ***A*** from the rest of the graph. In particular, it includes the parents and childrens of the node ***A***, and those children's other parents.



Source: Image from https://en.wikipedia.org/wiki/Markov_blanket

- **D-separation** concept for directed graphs enrich the representativity of the model → **Moral graph.**
- Two directed graph are **equivalents** when they lead to the same probabilistic model.
- This occurs when the **subyacent non-directed graph** is the same and include the same **V-structures**.
- The **Skeleton** of the graph is the undirected graph underlying.
- The **Markov Blanket** of a node ***A*** is the set of nodes that completely separates ***A*** from the rest of the graph. In particular, it includes the parents and childrens of the node ***A***, and those children's other parents.

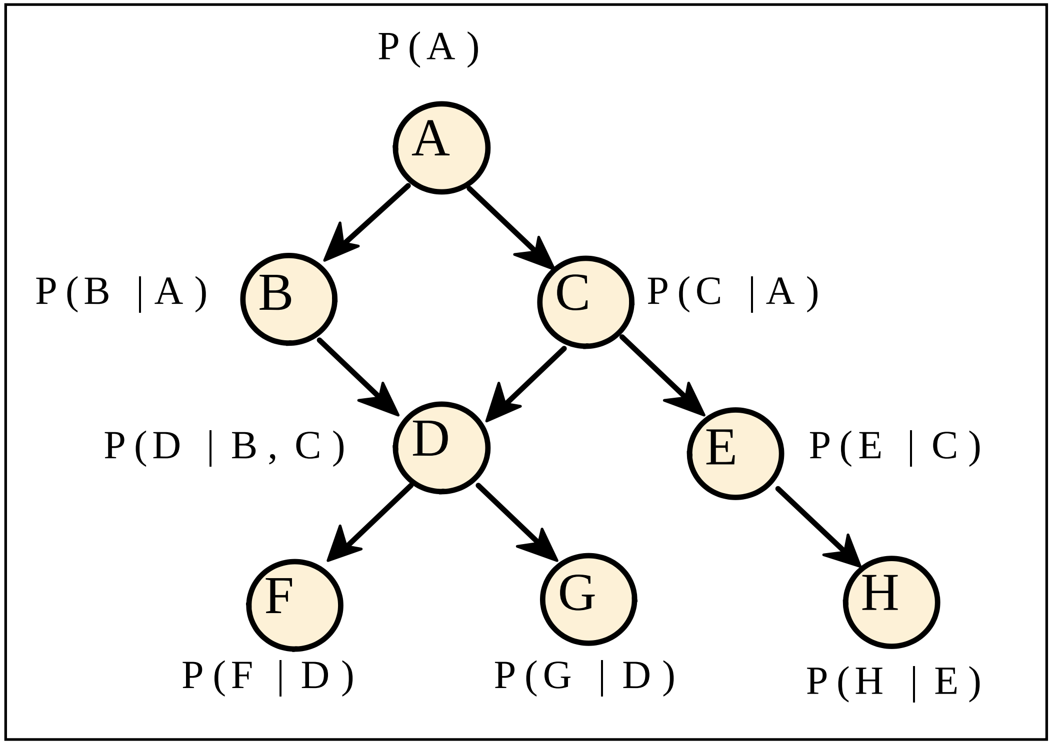


- The **Markov Blanket** is the set of nodes that includes all the knowledge needed to do inference on the node ***A***, from estimation to hypothesis testing to prediction.

Source: Image from https://en.wikipedia.org/wiki/Markov_blanket

Directed graphs lead to a probabilistic model directly obtained from the graph, defining the factorization of the joint probability function as product of conditional probabilities of each node x_i given his parents π_i.

$$P(X) = \prod_{i=1}^{n} P(X_i \vee \pi_i)$$



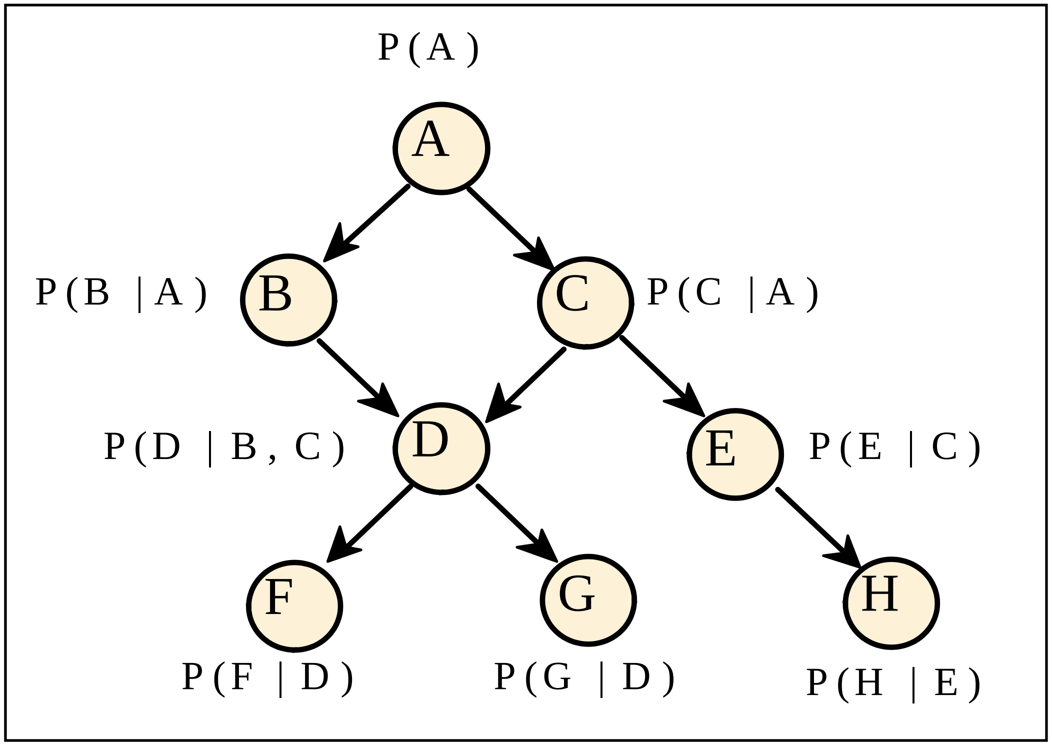
$$P(A,B,C,D,E,F,G,H) = P(A)\,P(B \vee A)\,P(C \vee A)\,P(D \vee B,C)\,x\ldots$$

$$xP(E \vee C)\,P(F \vee D)\,P(G \vee D)\,P(H \vee E)$$

```
## Define the graph using both the graph and the factorization expression (See ?modelstring):

## Plot both graphs, is there any difference between them?
```

$$P(X)=\prod_{i=1}^{n} P(X_i \vee \pi_i)$$



$$P(A,B,C,D,E,F,G,H)=P(A)P(B \vee A)P(C \vee A)P(D \vee B,C)x\ldots$$

$$xP(E \vee C)P(F \vee D)P(G \vee D)P(H \vee E)$$

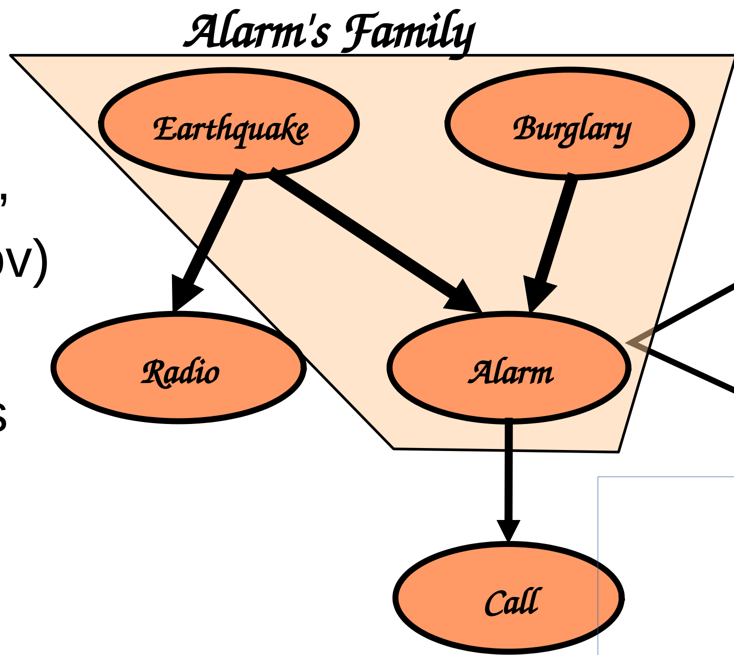
Bayesian Networks obtain a compact representation of the joint probability function through the conditional independences.

Structure:

Acyclic Directed Graph (DAG),

or non-directed graphs (Markov)

- Nodes – variables
- Links – direct dependences



E	B	$P(A \mid E,B)$	
e	b	0.9	0.1
e	$\bar{b}$	0.2	0.8
$\bar{e}$	b	0.9	0.1
$\bar{e}$	$\bar{b}$	0.01	0.99

Factorization of the joint probability function.

$$P(B,E,A,R,C)=P(E)P(B)P(R\vee E)P(A\vee E,B)P(C\vee A)$$

Parameters: Probabilities and tables.