

El clasificador Naïve Bayes (Minsky, 1961)

Seminaïve Bayes (Pazzani, 1997)

Naïve Bayes aumentado a árbol (Friedman y col., 1997)

Clasificador Bayesiano  $k$ -dependiente (Sahami, 1996)

Red Bayesiana (Jensen, 2001)

Resultados experimentales

# Clasificadores bayesianos

Abdelmalik Moujahid

Grupo de Inteligencia Computacional  
Universidad del País Vasco UPV/EHU  
Curso 2014-2015

El clasificador Naïve Bayes (Minsky, 1961)

Seminaïve Bayes (Pazzani, 1997)

Naïve Bayes aumentado a árbol (Friedman y col., 1997)

Clasificador Bayesiano  $k$ -dependiente (Sahami, 1996)

Red Bayesiana (Jensen, 2001)

Resultados experimentales

# Índice

- 1 El clasificador Naïve Bayes (Minsky, 1961)
- 2 Seminaïve Bayes (Pazzani, 1997)
- 3 Naïve Bayes aumentado a árbol (Friedman y col., 1997)
- 4 Clasificador Bayesiano  $k$ -dependiente (Sahami, 1996)
- 5 Red Bayesiana (Jensen, 2001)
- 6 Resultados experimentales

El clasificador Naïve Bayes (Minsky, 1961)

Seminaïve Bayes (Pazzani, 1997)

Naïve Bayes aumentado a árbol (Friedman y col., 1997)

Clasificador Bayesiano  $k$ -dependiente (Sahami, 1996)

Red Bayesiana (Jensen, 2001)

Resultados experimentales

# Índice

- 1 **El clasificador Naïve Bayes (Minsky, 1961)**
- 2 Seminaïve Bayes (Pazzani, 1997)
- 3 Naïve Bayes aumentado a árbol (Friedman y col., 1997)
- 4 Clasificador Bayesiano  $k$ -dependiente (Sahami, 1996)
- 5 Red Bayesiana (Jensen, 2001)
- 6 Resultados experimentales

## Introducción

### Clasificación Supervisada con Paradigmas Probabilistas

$$\gamma : (x_1, \dots, x_n) \rightarrow \{1, 2, \dots, r_0\}$$

- Suponemos la existencia de una distribución de probabilidad conjunta:

$$p(x_1, \dots, x_n, c) = p(c|x_1, \dots, x_n)p(x_1, \dots, x_n) = p(x_1, \dots, x_n|c)p(c)$$

- El clasificador Bayes consiste en asignar al ejemplo  $\mathbf{x} = (x_1, \dots, x_n)$  la *clase con mayor probabilidad a posteriori*. Es decir:

$$\gamma(\mathbf{x}) = \arg \max_c p(c|x_1, \dots, x_n)$$

## Introducción

### Teorema de Bayes (Bayes, 1764)

Sean  $C$  y  $X$  dos sucesos aleatorios cuyas probabilidades se denotan por  $p(C)$  y  $p(X)$  respectivamente, verificándose que  $p(X) > 0$ .

La probabilidad a posteriori del suceso  $C$  conocido que se verifica el suceso  $X$ , es decir  $p(C|X)$ , puede calcularse a partir de la siguiente fórmula:

$$p(C|X) = \frac{p(C, X)}{p(X)} = \frac{p(C)p(X|C)}{p(X)} = \frac{p(C)p(X|C)}{\sum_{C'} p(C')p(X|C')}$$

## Introducción

### Terminología

- **hallazgo**: valor de una variable predictora  $X_r$ ;
- **evidencia**, conjunto de todos los hallazgos para un determinado individuo;
- **diagnóstico**, valor que toman las  $m$  variables aleatorias  $Y_1, \dots, Y_m$ , cada una de las cuales se refiere a una enfermedad;
- **probabilidad a priori del diagnóstico**,  $p(\mathbf{y})$  o  $p(Y_1 = y_1, \dots, Y_m = y_m)$ ;
- **probabilidad a posteriori de un diagnóstico**,  $p(\mathbf{y}|\mathbf{x})$  o  $p(Y_1 = y_1, \dots, Y_m = y_m | X_1 = x_1, \dots, X_n = x_n)$ .

El clasificador Naïve Bayes (Minsky, 1961)

Seminaïve Bayes (Pazzani, 1997)

Naïve Bayes aumentado a árbol (Friedman y col., 1997)

Clasificador Bayesiano  $k$ -dependiente (Sahami, 1996)

Red Bayesiana (Jensen, 2001)

Resultados experimentales

## Introducción

### Caso de $m$ diagnósticos no excluyentes.

	$X_1$	...	$X_i$	...	$X_n$	$Y_1 \dots Y_m$
$(x^1, y^1)$	$x_1^1$	...	$x_i^1$	...	$x_n^1$	$y_1^1 \dots y_m^1$
...	...	...	...	...	...	...
$(x^j, y^j)$	$x_1^j$	...	$x_i^j$	...	$x_n^j$	$y_1^j \dots y_m^j$
...	...	...	...	...	...	...
$(x^N, y^N)$	$x_1^N$	...	$x_i^N$	...	$x_n^N$	$y_1^N \dots y_m^N$

## Introducción

### Caso de $m$ diagnósticos no excluyentes.

$$(y_1^*, \dots, y_m^*) = \arg \max_{(y_1, \dots, y_m)} P(Y_1 = y_1, \dots, Y_m = y_m | X_1 = x_1, \dots, X_n = x_n)$$

$$P(Y_1 = y_1, \dots, Y_m = y_m | X_1 = x_1, \dots, X_n = x_n)$$

$$\propto p(Y_1 = y_1, \dots, Y_m = y_m) P(X_1 = x_1, \dots, X_n = x_n | Y_1 = y_1, \dots, Y_m = y_m)$$

número de parámetros a estimar:  $2^m - 1 + 2^m(2^n - 1)$

$m$	$n$	parámetros
3	10	$8 \cdot 10^3$
5	20	$33 \cdot 10^6$
10	50	$11 \cdot 10^{17}$



## Introducción

### Diagnósticos excluyentes.

$$c^* = \arg \max_c p(C = c | X_1 = x_1, \dots, X_n = x_n)$$

$$p(C = c | X_1 = x_1, \dots, X_n = x_n) \propto p(C = c) p(X_1 = x_1, \dots, X_n = x_n | C = c)$$

número de parámetros a estimar:  $(r_0 - 1) + r_0(2^n - 1)$

$m$	$n$	parámetros
3	10	$3 \cdot 10^3$
5	20	$5 \cdot 10^6$
10	50	$11 \cdot 10^{15}$

## Naïve Bayes

### Diagnósticos excluyentes y variables condicionalmente independientes dado el diagnóstico.

$$c^* = \arg \max_c p(C = c | X_1 = x_1, \dots, X_n = x_n)$$

$$= \arg \max_c p(C = c) \prod_{i=1}^n p(X_i = x_i | C = c)$$

número de parámetros a estimar:  $(r_0 - 1) + r_0 n$

$m$	$n$	parámetros
3	10	32
5	20	104
10	50	509

El clasificador Naïve Bayes (Minsky, 1961)

Seminaïve Bayes (Pazzani, 1997)

Naïve Bayes aumentado a árbol (Friedman y col., 1997)

Clasificador Bayesiano  $k$ -dependiente (Sahami, 1996)

Red Bayesiana (Jensen, 2001)

Resultados experimentales

## Naïve Bayes

### Variables predictoras discretas

$$c^* = \arg \max_c p(C = c) \prod_{i=1}^n p(X_i = x_i | C = c)$$

### Variables predictoras continuas y normales

$$c^* = \arg \max_c p(C = c) \prod_{i=1}^n \left[ \frac{1}{\sqrt{2\pi}\sigma_i^c} e^{-\frac{1}{2} \left( \frac{x_i - \mu_i^c}{\sigma_i^c} \right)^2} \right]$$

El clasificador Naïve Bayes (Minsky, 1961)

Seminaïve Bayes (Pazzani, 1997)

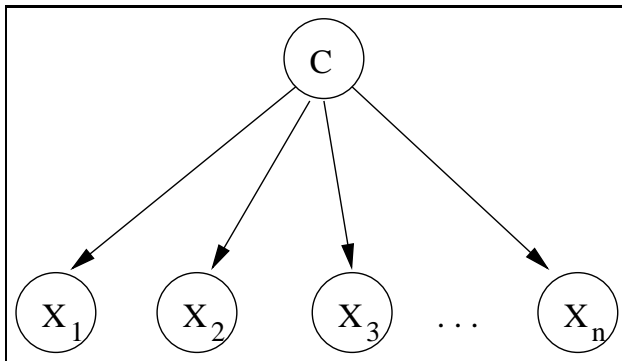
Naïve Bayes aumentado a árbol (Friedman y col., 1997)

Clasificador Bayesiano  $k$ -dependiente (Sahami, 1996)

Red Bayesiana (Jensen, 2001)

Resultados experimentales

## Naïve Bayes



Estructura gráfica de un modelo naïve Bayes

**Simple Bayesian classifier:**

<http://www.ics.uci.edu/~pazzani/Publications/mlc96-pedro.pdf>

El clasificador Naïve Bayes (Minsky, 1961)

**Seminaïve Bayes (Pazzani, 1997)**

Naïve Bayes aumentado a árbol (Friedman y col., 1997)

Clasificador Bayesiano  $k$ -dependiente (Sahami, 1996)

Red Bayesiana (Jensen, 2001)

Resultados experimentales

# Índice

- 1 El clasificador Naïve Bayes (Minsky, 1961)
- 2 Seminaïve Bayes (Pazzani, 1997)**
- 3 Naïve Bayes aumentado a árbol (Friedman y col., 1997)
- 4 Clasificador Bayesiano  $k$ -dependiente (Sahami, 1996)
- 5 Red Bayesiana (Jensen, 2001)
- 6 Resultados experimentales

## El algoritmo FSSJ (Pazzani, 1997)

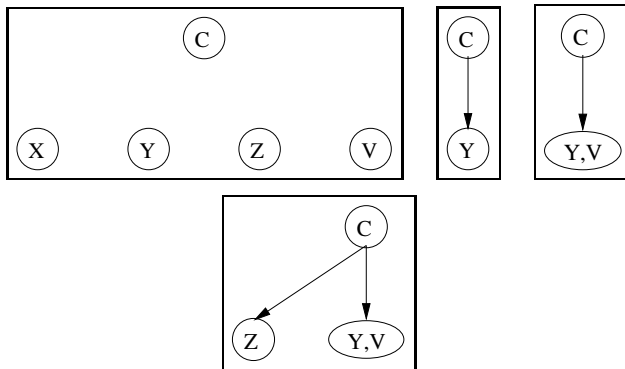
**Paso1** Inicializar el conjunto de variables a utilizar a vacío. Clasificar todos los ejemplos como pertenecientes a la clase más frecuente

**Paso2 Repetir** en cada paso la mejor opción entre:

- (a) Considerar cada variable que no está en el modelo como una variable a incluir en el modelo. Dicha variable debe incluirse condicionalmente independiente de las variables presentes en el modelo, dada la variable clase
- (b) Juntar cada variable no presente en el modelo con una variable que ya forme parte del mismo Evaluar cada posible opción por medio de la estimación del porcentaje de bien clasificados

**Hasta que** ninguna opción produzca mejoras.

## Seminaïve Bayes



Proceso de construcción de un modelo seminaïve Bayes.  $p(c|x, y, z, v) \propto p(c)p(z|c)p((y, v)|c)$

El clasificador Naïve Bayes (Minsky, 1961)

Seminaïve Bayes (Pazzani, 1997)

**Naïve Bayes aumentado a árbol (Friedman y col., 1997)**

Clasificador Bayesiano  $k$ -dependiente (Sahami, 1996)

Red Bayesiana (Jensen, 2001)

Resultados experimentales

# Índice

- 1 El clasificador Naïve Bayes (Minsky, 1961)
- 2 Seminaïve Bayes (Pazzani, 1997)
- 3 Naïve Bayes aumentado a árbol (Friedman y col., 1997)**
- 4 Clasificador Bayesiano  $k$ -dependiente (Sahami, 1996)
- 5 Red Bayesiana (Jensen, 2001)
- 6 Resultados experimentales



## Naïve Bayes aumentado a árbol.

- Cantidad de información mutua entre  $X$  e  $Y$

$$I(X, Y) = H(X) - H(X|Y) = \sum_{i=1}^{r_X} \sum_{j=1}^{r_Y} p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)}$$

- Cantidad de información mutua entre  $X$  e  $Y$  condicionada a  $C$

$$\begin{aligned} I(X, Y|C) &= \sum_c p(c) I(X, Y|C=c) \\ &= \sum_{i=1}^{r_X} \sum_{j=1}^{r_Y} \sum_{k=1}^{r_0} p(x_i, y_j, c_k) \log \frac{p(x_i, y_j|c_k)}{p(x_i|c_k)p(y_j|c_k)} \end{aligned}$$

## Naïve Bayes aumentado a árbol.

**Paso1** Calcular  $I(X_i, X_j|C)$  con  $i < j, i, j = 1, \dots, n$

**Paso2** Construir un grafo no dirigido completo cuyos nodos corresponden a las variables predictoras:  $X_1, \dots, X_n$ . Asignar a cada arista conectando las variables  $X_i$  y  $X_j$  un peso dado por  $I(X_i, X_j|C)$

**Paso3** Asignar las dos aristas de mayor peso al árbol a construir

**Paso4** Examinar la siguiente arista de mayor peso, y añadirla al árbol a no ser que forme un ciclo, en cuyo caso se descarta y se examina la siguiente arista de mayor peso

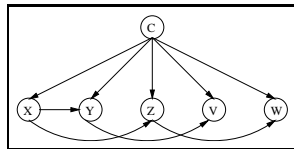
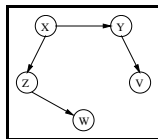
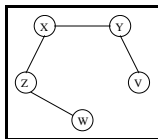
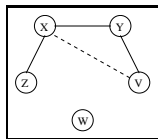
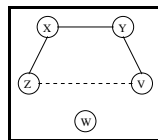
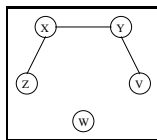
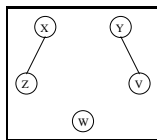
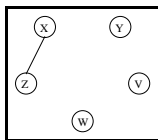
**Paso5** Repetir el paso 4 hasta seleccionar  $n - 1$  aristas

**Paso6** Transformar el árbol no dirigido resultante en uno dirigido, escogiendo una variable como raíz, para a continuación direccionar el resto de aristas

**Paso7** Construir un modelo TAN añadiendo un nodo etiquetado como  $C$  y posteriormente un arco desde  $C$  a cada variable predictora  $X_i$

Pseudocódigo del algoritmo TAN (Friedman y col., 1997)

## Naïve Bayes aumentado a árbol.



Proceso de construcción de TAN.  $I(X, Z|C) > I(Y, V|C) > I(X, Y|C) > I(Z, V|C) > I(X, V|C) >$

$I(Z, W|C) > I(X, W|C) > I(Y, Z|C) > I(Y, W|C) > I(V, W|C)$

$p(c|x, y, z, v, w) \propto p(c)p(x|c)p(y|x, c)p(z|x, c)p(v|y, c)p(w|z, c)$

El clasificador Naïve Bayes (Minsky, 1961)

Seminaïve Bayes (Pazzani, 1997)

Naïve Bayes aumentado a árbol (Friedman y col., 1997)

**Clasificador Bayesiano  $k$ -dependiente (Sahami, 1996)**

Red Bayesiana (Jensen, 2001)

Resultados experimentales

# Índice

- 1 El clasificador Naïve Bayes (Minsky, 1961)
- 2 Seminaïve Bayes (Pazzani, 1997)
- 3 Naïve Bayes aumentado a árbol (Friedman y col., 1997)
- 4 Clasificador Bayesiano  $k$ -dependiente (Sahami, 1996)**
- 5 Red Bayesiana (Jensen, 2001)
- 6 Resultados experimentales

## Clasificador Bayesiano $k$ -dependiente.

- Precalcula  $I(X_i, C)$  y  $I(X_i, X_j|C)$  para todo par de variables
- Añade en cada iteración, de entre las variables que no están en el modelo, aquella  $X_{max}$  que tenga mayor  $I(X_i, C)$
- Asigna a la variable añadida como padres la variable  $C$  y aquellas  $k$  variables con mayor  $I(X_j, X_{max}|C)$

El clasificador Naïve Bayes (Minsky, 1961)

Seminaïve Bayes (Pazzani, 1997)

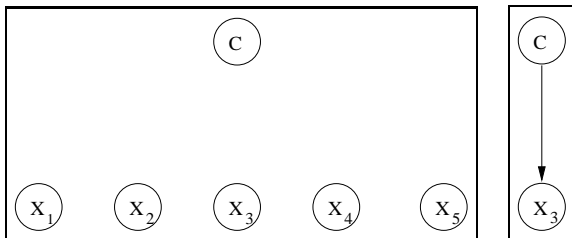
Naïve Bayes aumentado a árbol (Friedman y col., 1997)

**Clasificador Bayesiano  $k$ -dependiente (Sahami, 1996)**

Red Bayesiana (Jensen, 2001)

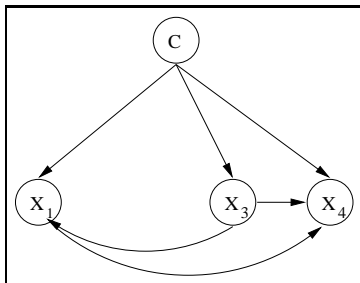
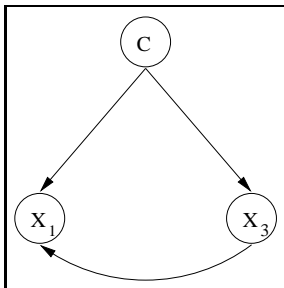
Resultados experimentales

## Clasificador Bayesiano $k$ -dependiente.



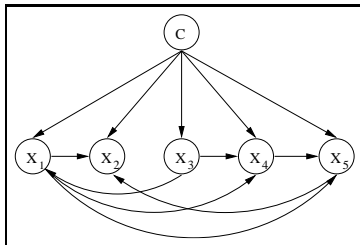
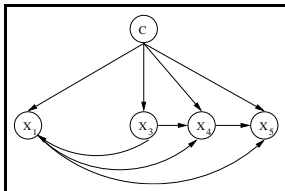
Proceso de construcción de  $k$ DB con  $k = 2$ .  $I(X_3, C) > I(X_1, C) > I(X_4, C) > I(X_5, C) > I(X_2, C)$   
 $I(X_3, X_4|C) > I(X_2, X_5|C) > I(X_1, X_3|C) > I(X_1, X_2|C) > I(X_2, X_4|C) > I(X_2, X_3|C) > I(X_1, X_4|C) >$   
 $I(X_4, X_5|C) > I(X_1, X_5|C) > I(X_3, X_5|C)$

## Clasificador Bayesiano $k$ -dependiente.



Proceso de construcción de  $k$ DB con  $k = 2$ .  $I(X_3, C) > I(X_1, C) > I(X_4, C) > I(X_5, C) > I(X_2, C)$   
 $I(X_3, X_4|C) > I(X_2, X_5|C) > I(X_1, X_3|C) > I(X_1, X_2|C) > I(X_2, X_4|C) > I(X_2, X_3|C) > I(X_1, X_4|C) >$   
 $I(X_4, X_5|C) > I(X_1, X_5|C) > I(X_3, X_5|C) p(c|x_1, x_2, x_3, x_4, x_5)$

## Clasificador Bayesiano $k$ -dependiente.



Proceso de construcción de  $k$ DB con  $k = 2$ .  $I(X_3, C) > I(X_1, C) > I(X_4, C) > I(X_5, C) > I(X_2, C)$   
 $I(X_3, X_4|C) > I(X_2, X_5|C) > I(X_1, X_3|C) > I(X_1, X_2|C) > I(X_2, X_4|C) > I(X_2, X_3|C) > I(X_1, X_4|C) >$   
 $I(X_4, X_5|C) > I(X_1, X_5|C) > I(X_3, X_5|C)$   
 $p(c|x_1, x_2, x_3, x_4, x_5) \propto p(c)p(x_1|x_3, c)p(x_2|x_1, x_5, c)p(x_3|c)p(x_4|x_1, x_3, c)p(x_5|x_1, x_4, c)$



El clasificador Naïve Bayes (Minsky, 1961)

Seminaïve Bayes (Pazzani, 1997)

Naïve Bayes aumentado a árbol (Friedman y col., 1997)

Clasificador Bayesiano  $k$ -dependiente (Sahami, 1996)

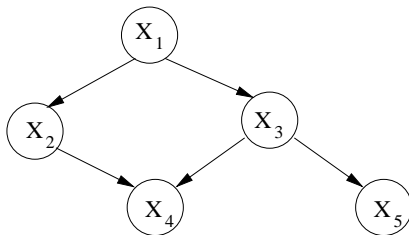
**Red Bayesiana (Jensen, 2001)**

Resultados experimentales

# Índice

- 1 El clasificador Naïve Bayes (Minsky, 1961)
- 2 Seminaïve Bayes (Pazzani, 1997)
- 3 Naïve Bayes aumentado a árbol (Friedman y col., 1997)
- 4 Clasificador Bayesiano  $k$ -dependiente (Sahami, 1996)
- 5 Red Bayesiana (Jensen, 2001)**
- 6 Resultados experimentales

## Red Bayesiana múltiplemente conectada.



$$p(X_1 = 0) = 0,20$$

$$p(X_2 = 0 | X_1 = 0) = 0,80$$

$$p(X_2 = 0 | X_1 = 1) = 0,80$$

$$p(X_3 = 0 | X_1 = 0) = 0,20$$

$$p(X_3 = 0 | X_1 = 1) = 0,05$$

$$p(X_4 = 0 | X_2 = 0, X_3 = 0) = 0,80$$

$$p(X_4 = 0 | X_2 = 1, X_3 = 0) = 0,80$$

$$p(X_4 = 0 | X_2 = 0, X_3 = 1) = 0,80$$

$$p(X_4 = 0 | X_2 = 1, X_3 = 1) = 0,05$$

$$p(X_5 = 0 | X_3 = 0) = 0,80$$

$$p(X_5 = 0 | X_3 = 1) = 0,60$$

Factorización de la distribución de probabilidad conjunta obtenida con la red Bayesiana adjunta

El clasificador Naïve Bayes (Minsky, 1961)

Seminaïve Bayes (Pazzani, 1997)

Naïve Bayes aumentado a árbol (Friedman y col., 1997)

Clasificador Bayesiano  $k$ -dependiente (Sahami, 1996)

Red Bayesiana (Jensen, 2001)

**Resultados experimentales**

# Índice

- 1 El clasificador Naïve Bayes (Minsky, 1961)
- 2 Seminaïve Bayes (Pazzani, 1997)
- 3 Naïve Bayes aumentado a árbol (Friedman y col., 1997)
- 4 Clasificador Bayesiano  $k$ -dependiente (Sahami, 1996)
- 5 Red Bayesiana (Jensen, 2001)
- 6 Resultados experimentales**

## Cuadro: Table Caption

Dataset	(1)	(2)	(3)	(4)
breast-cancer	66.91±6.11	74.28±6.05 ○	72.85±6.93 ○	72.70±7.74 ○
german-credit	66.19±3.30	71.25±3.17 ○	71.88±3.68 ○	75.16±3.48 ○
DRUG1n	60.30±6.99	92.15±5.42 ○	86.00±7.21 ○	87.15±7.33 ○
pima-diabetes	72.00±4.72	74.49±5.27	70.62±4.67	75.75±5.32 ○
hepatitis	82.48±8.27	79.22±9.57	81.40±8.55	83.81±9.70
iris	93.53±5.56	94.73±5.30	95.40±4.80	95.53±5.02
market-basket	74.70±2.11	77.82±1.31 ○	67.51±4.39 ●	57.98±5.70 ●
soybean	39.75±2.71	91.78±3.19 ○	91.20±3.00 ○	92.94±2.92 ○
vehicle	52.21±4.71	72.28±4.32 ○	69.59±3.77 ○	44.68±4.59 ●
Average	67.56	80.89	78.49	76.19

○, ● statistically significant improvement or degradation