MODELOS COMPUTACIONALES: CUARTO CURSO DEL GRADO

DE ING. INFORMÁTICA EN COMPUTACION

Clasificación: Redes Bayesianas

César Hervás-Martínez
Grupo de Investigación AYRNA

Departamento de Informática y Análisis Numérico
Universidad de Córdoba
Campus de Rabanales. Edificio Einstein.
Email: chervas@uco.es

2018-2019

Probabilidad condicionada. Espacio de probabilidad condicionado

La probabilidad condicionada es uno de los conceptos clave en Teoría de la Probabilidad.

Hay situaciones en las que se incorpora información suplementaria asociada a la ocurrencia de otro suceso, con lo que puede variar el espacio de resultados posibles y consecuentemente, sus probabilidades.

En este contexto aparece el concepto de probabilidad condicionada

Probabilidad condicionada. Espacio de probabilidad condicionado

Definición.- Sea $(\Omega, \mathcal{A}, P)$ un espacio probabilístico arbitrario y A un suceso ($A \in \mathcal{A}$) tal que $P(A) > 0$. Para cualquier otro suceso $B \in \mathcal{A}$, se define la probabilidad condicionada de B dado A o probabilidad de B condicionada a A como

$$P(B/A) = P(B \cap A) / P(A) .$$

Observemos que la condición $P(A) > 0$ es necesaria para que la definición tenga sentido. Por otra parte, la idea intuitiva de probabilidad condicionada hace lógica esta restricción ya que si $P(A)=0$, A es un suceso imposible y no tiene sentido condicionar a él.

Notemos que, sabiendo que $A \in \mathcal{A}$ ha ocurrido, tenemos una nueva evaluación de la probabilidad de cada suceso que pasa de ser $P(B)$ a $P(B/A)$, o sea, tenemos una nueva función de conjunto sobre $(\Omega, \mathcal{A})$, que, efectivamente, esta función es una medida de probabilidad sobre $(\Omega, \mathcal{A})$

INDEPENDENCIA ESTOCÁSTICA DE SUCESOS:

Sean A y B dos sucesos con probabilidades no nulas, luego existen P(A) y P(B) y no son cero, se dice que son independientes cuando se verifica que

$$P(A \cap B) = P(A)P(B),$$

$$\text{siendo } \mathrm{P(A)} > 0 \text{ y } \mathrm{P(B)} > 0 \text{ y } \mathrm{A} \cap \mathrm{B} \neq \varnothing$$

De esta forma dos sucesos A y B son estocásticamente independientes cuando la información sobre la ocurrencia de uno de ellos no modifica la probabilidad de que ocurra el otro. Esto es :

$$P(A \cap B) = P(A)P(B/A) = P(B)P(A/B)$$

$$\text{pero si son independientes } \mathrm{P(B/A)=P(B)}$$

$$\text{o equivalentemente } \mathrm{P(A/B)=P(A)}$$

Introducción

Hipótesis de independencia

Son necesarias para desarrollar una inferencia probabilística más practica

Método Naïve Bayes

Hace hipótesis de independencia que a menudo no son ciertas

También se le llama por esta razón Método Bayesiano Simple

Red Bayesiana

Modela explícitamente las relaciones de independencia de las variables asociadas a los datos

Utiliza estas relaciones de independencia para hacer inferencias probabilísticas

También se la conoce como: Red de Creencia, Red Bayes, Red Causal, etc

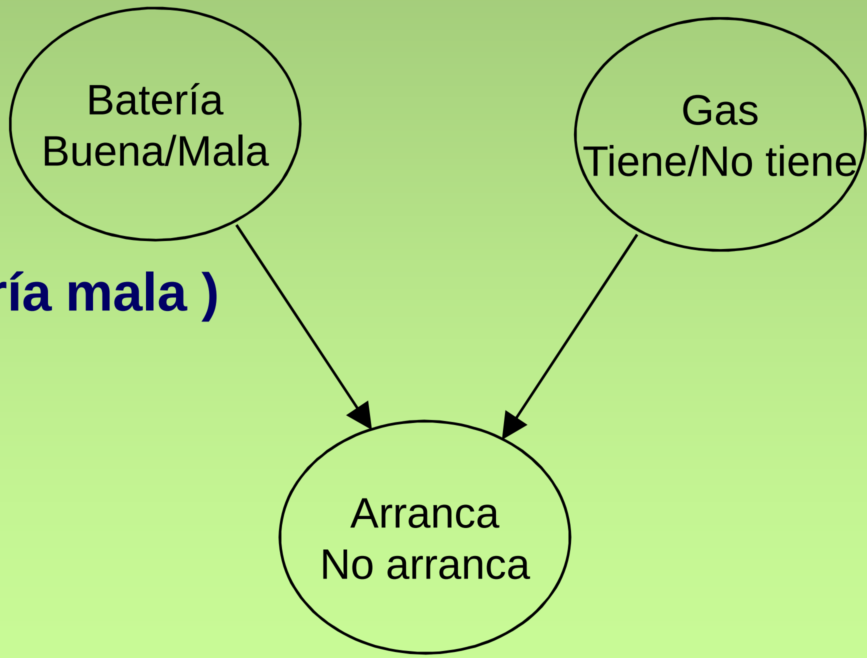
¿Porque Redes Bayesianas?

1) Lenguaje intuitivo

Puede utilizar conocimiento causal en la construcción de los modelos. Los expertos en el dominio de aplicación pueden construir de forma confortable la red.

2) Algoritmos de inferencia de propósito general

P(No arranca| Tiene gas, Batería mala)



3) Exacta: Las especificaciones modulares permiten grandes eficiencias computacionales

Redes Bayesianas: Aplicaciones

- **Búsqueda inteligente en Google**
- **Recuperación de información**
- **Filtrado colaborativo y tecnología de recomendación**
- **Sistemas expertos para el diagnóstico médico**
- **Minería de datos**
- **Evaluación de riesgos y predicción de la calidad en los sistemas**
- **Predicción de riesgo en el tráfico aéreo**
- **Visión por Computador**
- **Colisión de aeronaves y drones**
- **Análisis de defectos de software**
- **Fiabilidad de Sistemas y disponibilidad**
- **Riesgo operacional en instituciones financieras**
- **Cartera de proyectos de riesgo**

Independencia de variables

Definición de variable aleatoria

Un conjunto de variable aleatorias son independientes cuando la distribución de probabilidad conjunta es el producto de las distribuciones de probabilidad marginales:

Ejemplo

v. a. X: Presión sanguínea de un paciente {alta, media, baja} v. a. Y: Paciente con estornudos {si, no}

$$P(X=x_i, Y=y_j)= P(X=x_i). P(Y=y_j) \quad \text{para todo } i, j$$

Como un ejemplo

$$P(X=alta, Y=si)= P(X=alta). P(Y=si)$$

Se verifica una independencia condicional entre un conjunto de variables, si se verifica dicha independencia condicional entre todos los posibles sucesos de las variables.

Probabilidad Condicionada

A: **'Una persona tiene cáncer' P(*A*) = 0,1 ("a *priori*")**

B: **'Una persona es fumador' P(*B*) = 0,5**

¿Cual es la probabilidad P(*A* | *B*) ? ("a *posteriori*")

P(*B* | *A*) = 0,8 (verosimilitud "*likelihood*")

Probabilidad a posteriori Probabilidad a priori Verosimilitud

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B \mid A)}{P(B)}$$

$$P(A \mid B) = \frac{0{,}1 \times 0{,}8}{0{,}5} = 0{,}16$$

Introducción a las Redes Bayesianas

Las Redes Bayesianas son grafos acíclicos directos (DAGs).

Los nodos de las Redes Bayesianas representan a las variables aleatorias, las cuales “a priori” consideraremos que toman valores discretos.

Los enlaces de la red representan una influencia directa probabilística.

La estructura de la red representa la relación de dependencia/independencia probabilística entre las variables aleatorias representadas por los nodos

Independencia condicional de sucesos

Dos sucesos A y B son condicionalmente independientes si para cualquier suceso C con P(C)>0

$$P(A \cap B / C) = P(A / C)P(B / C) \qquad \textbf{(1)}$$

Si recordamos la definición de probabilidad condicional,

$$P(A / B) = \frac{P(A \cap B)}{P(B)}, \text{ si P(B)} \neq 0$$

Si condicionamos la anterior probabilidad sobre C, obtenemos

$$P(A / B \cap C) = \frac{P(A \cap B \cap C)}{P(B \cap C)} = \frac{P((A \cap B) / C)P(C)}{P(B / C)P(C)} = \frac{P((A \cap B) / C)}{P(B / C)},$$

$$\text{si P(C)} \neq 0 \text{ y si } P(B / C) \neq 0$$

Si A y B son condicionalmente independientes dado C, tenemos

$$P(A / B \cap C) = \frac{P((A \cap B) / C)}{P(B / C)} = \frac{P(A / C)P(B / C)}{P(B / C)} = P(A / C)$$

Entonces si A y B son condicionalmente independientes dado C, tenemos que

$$P(A / B \cap C) = P(A / C) \qquad \textbf{(2)}$$

De las ecuaciones (1) y (2) deducimos que son definiciones similares de independencia condicional

Independencia condicional de con juntos

Def.- Independencia Condicional

Dados Y, Z, W tres conjuntos disjuntos de variables, se dice que Y es condicionalmente independiente de Z dado W, si se verifica que para algún y, z, w:

$$p(y \mid z, w) = p(y \mid w)$$

Estructura de una Red Bayesiana

Para cada variable X_i, la estructura S para X representa la afirmación de que X_i y X_i y $\{X_1, \ldots, X_n\} / \mathbf{pa}_i^S$ **son independientes, dado que** $\mathbf{pa}_i^S$ **es el conjunto de padres de X_i en S)**

Factorización de la Distribución de Probabilidad Conjunta

$$p(\mathbf{x}) = p(x_1, \ldots, x_n) = \prod_{i=1}^{n} p(x_i \mid \mathbf{pa}_i^S)$$

Factorización de la distribución de probabilidad conjunta

Ejemplos $$p(\mathbf{x}) = p(x_1,\ldots,x_n) = \prod_{i=1}^{n} p(x_i \mid \mathbf{pa}_i^S)$$

$x_i \in \mathrm{B}(p_i)$, esto es, si x_i se distribuye según una distribución de Bernouilli de parámetro p_i

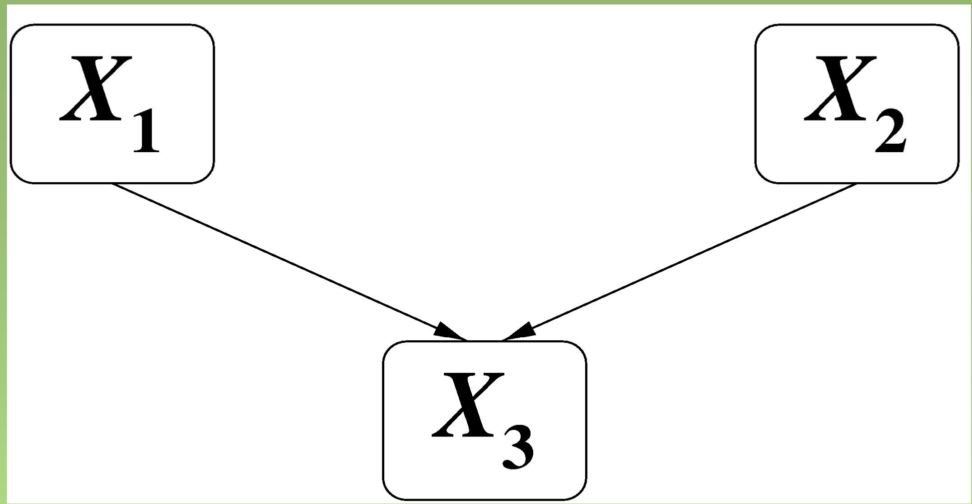
$$p(\mathbf{x}) = p(x_1,\ldots,x_n) = \prod_{i=1}^{n} p(x_i \mid \mathbf{pa}_i^S) = \prod_{i=1}^{n} (p_i)^{x_i}(1-p_i)^{(1-x_i)},$$

donde $p_i = p(x_i \mid \mathbf{pa}_i^S)$

$x_i \in N(\mu_i, \sigma_i^2)$, esto es, si x_i se distribuye según una distribución Normal de media μ_i y de varianza σ_i^2

$$p(\mathbf{x}) = p(x_1,\ldots,x_n) = \prod_{i=1}^{n} (2\pi\sigma_i)^{-1/2} \exp\left(\frac{(x_i-\mu_i)^2}{\sigma_i^2}\right)$$

Estructura y parámetros del modelo



Redes Bayesianas: Ventajas

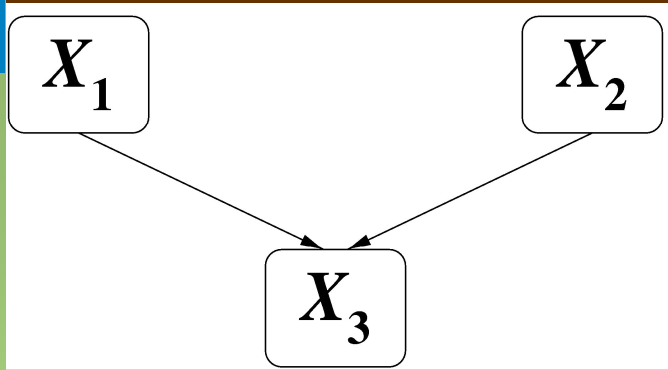
Representación

Basada en la teoría de la probabilidad, proporciona una semántica clara (basada en la independencia condicional entre tripletas de variables) y un sólido fundamento teórico

Representación del conocimiento: gráfico e intuitivo

Modularidad: reducción del numero de parámetros para especificar la distribución conjunta de probabilidad

Estructura y parámetros del modelo



$$\mathbf{x_1} = (x_1^1, x_1^2),\ \mathbf{x}_2 = (x_2^1, x_2^2),\ \mathbf{x}_3 = (x_3^1, x_3^2)$$

Parámetros del modelo y distribuciones de probabilidad locales

$$p(x_i^k \mid \mathbf{pa}_i^{j,S}, \theta_i) = \theta_{x_i^k|\mathbf{pa}_i^{j,S}} \equiv \theta_{ijk}$$

$$\boldsymbol{\theta}_\mathbf{1} = \left(\theta_{11}, \theta_{12}\right) \qquad \mathrm{p}(\mathrm{x}_1^1|\boldsymbol{\theta}_\mathbf{1}),\quad \mathrm{p}(\mathrm{x}_1^2|\boldsymbol{\theta}_\mathbf{1})$$

$$\boldsymbol{\theta}_\mathbf{2} = \left(\theta_{21}, \theta_{22}\right) \qquad \mathrm{p}(\mathrm{x}_2^1|\boldsymbol{\theta}_\mathbf{2}),\quad \mathrm{p}(\mathrm{x}_2^2|\boldsymbol{\theta}_\mathbf{2})$$

$$\boldsymbol{\theta}_\mathbf{3} = \left(\theta_{311}, \theta_{312}\right) \qquad \mathrm{p}(\mathrm{x}_3^1|\mathrm{x}_1^1,\mathrm{x}_2^1,\boldsymbol{\theta}_\mathbf{3}),\quad \mathrm{p}(\mathrm{x}_3^2|\mathrm{x}_1^1,\mathrm{x}_2^1,\boldsymbol{\theta}_\mathbf{3})$$

$$\left(\theta_{321}, \theta_{322}\right) \qquad \mathrm{p}(\mathrm{x}_3^1|\mathrm{x}_1^1,\mathrm{x}_2^2,\boldsymbol{\theta}_\mathbf{3}),\quad \mathrm{p}(\mathrm{x}_3^2|\mathrm{x}_1^1,\mathrm{x}_2^2,\boldsymbol{\theta}_\mathbf{3})$$

$$\left(\theta_{331}, \theta_{332}\right) \qquad \mathrm{p}(\mathrm{x}_3^1|\mathrm{x}_1^2,\mathrm{x}_2^1,\boldsymbol{\theta}_\mathbf{3}),\quad \mathrm{p}(\mathrm{x}_3^2|\mathrm{x}_1^2,\mathrm{x}_2^1,\boldsymbol{\theta}_\mathbf{3})$$

$$\left(\theta_{341}, \theta_{342}\right) \qquad \mathrm{p}(\mathrm{x}_3^1|\mathrm{x}_1^2,\mathrm{x}_2^2,\boldsymbol{\theta}_\mathbf{3}),\quad \mathrm{p}(\mathrm{x}_3^2|\mathrm{x}_1^2,\mathrm{x}_2^2,\boldsymbol{\theta}_\mathbf{3})$$

Factorización de la distribución de probabilidad conjunta

$$p(\mathbf{x} \mid \boldsymbol{\theta}_S) = p(\mathbf{x_1} \mid \boldsymbol{\theta}_\mathbf{1})\, p(\mathbf{x_2} \mid \boldsymbol{\theta}_\mathbf{2})\, p(\mathbf{x_3} \mid \mathbf{x_1}, \mathbf{x_2}, \boldsymbol{\theta}_\mathbf{3})$$

Redes Bayesianas: Ventajas

Inferencia.- Razonamiento (propagación de la evidencia) mediante algoritmos exactos y aproximados

- Probabilidad de un suceso dada una evidencia
- Explicación más probable (inferencia abductiva o generación de hipótesis) como el suceso que mejor explica la evidencia actual. Es un tipo de razonamiento que a partir de la descripción de un hecho o fenómeno ofrece o llega a una hipótesis.

Aprendizaje a partir de los datos

Tres maneras principales de construir redes Bayesianas:

- **Manual** (mediante la ayuda de un experto en el dominio a ser modelado)
- **Induciendo a partir de una base de datos de casos** (Aprendizaje a partir de datos)
- **Aproximaciones mixtas**

Estimador de máxima probabilidad, MPE, para C dado X

$$C^* = \arg\max_c P(C = c \mid X_1 = x_1, \ldots, X_n = x_n) =$$

$$= \arg\max_c \frac{P((C = c) \cap (X_1 = x_1, \ldots, X_n = x_n))}{P(X_1 = x_1, \ldots, X_n = x_n)} =$$

$$= \arg\max_c P(C = c) P(X_1 = x_1, \ldots, X_n = x_n \mid C = c)$$

Dependiendo de la factorización de:

$$P(X_1 = x_1, \ldots, X_n = x_n \mid C = c)$$

Obtenemos diferentes clasificadores Bayesianos

Clasificadores Bayesianos

Dependiendo de la factorización de:

$$P(X_1 = x_1, \ldots, X_n = x_n \mid C = c)$$

Obtenemos diferentes clasificadores Bayesianos

Modelos Jerárquicos de Clasificadores Bayesianos

- **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)**

Clasificación Supervisada con Paradigmas Probabilistas

$$C^* = \arg\max_c P(C=c \mid X_1 = x_1, \ldots, X_n = x_n) =$$

$$= \arg\max_c \frac{P((C=c) \cap (X_1 = x_1, \ldots, X_n = x_n))}{P(X_1 = x_1, \ldots, X_n = x_n)} =$$

$$= \arg\max_c P(C=c) P(X_1 = x_1, \ldots, X_n = x_n \mid C = c)$$

$$\gamma : (x_1, \ldots, x_n) \rightarrow \{1, 2, \ldots, \mathrm{J}\}$$

Matriz de costes: C(r,s)

Minimización del coste total de errores

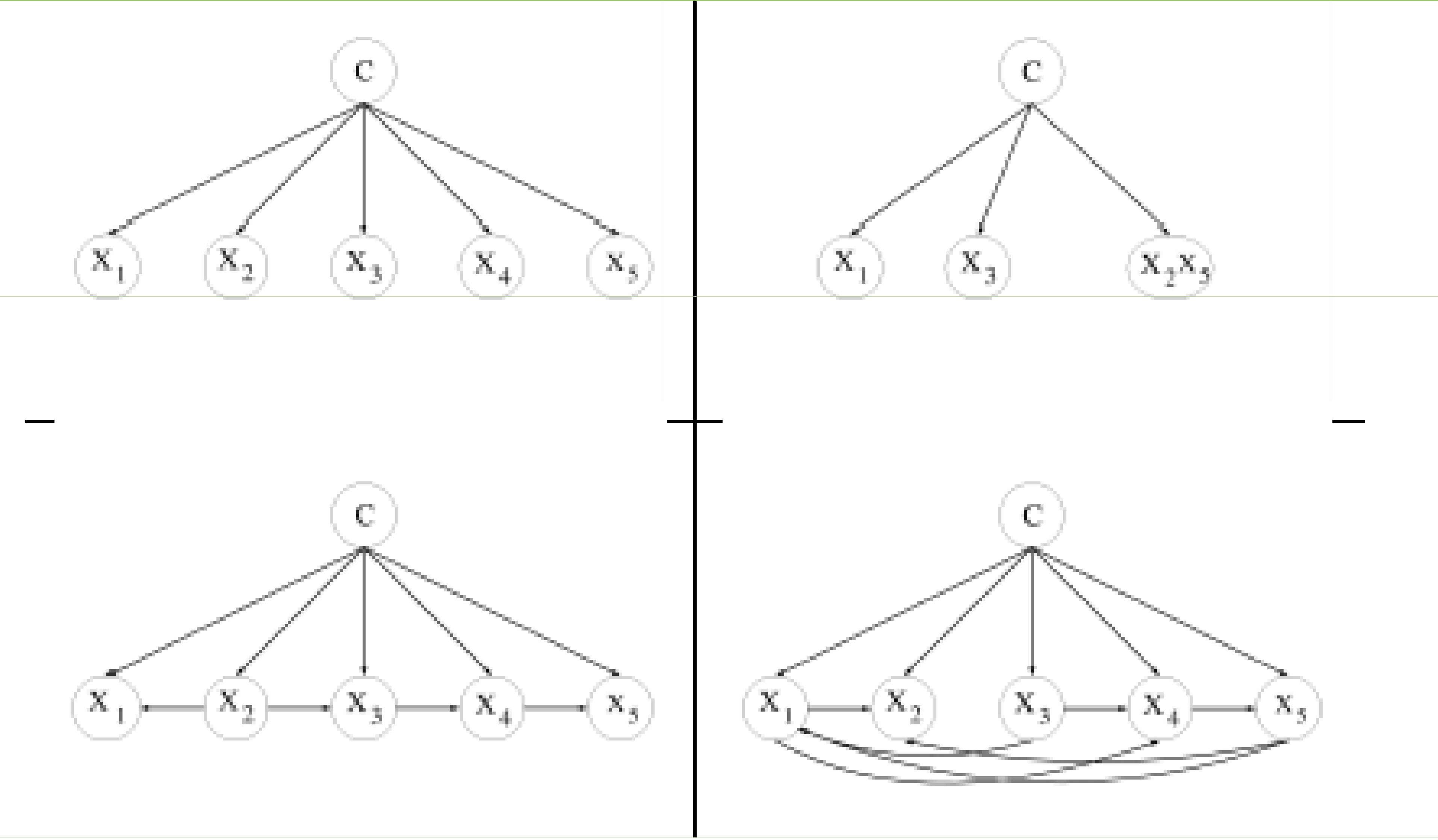
$$\gamma(\mathbf{x}) = \arg \min_k \sum_{c=1}^{\mathrm{J}} \mathrm{C}(k, c)\, \mathrm{P}(c \mid x_1, \ldots, x_n)$$

Función de pérdida 0/1

$$\gamma(\mathbf{x}) = \arg \max_c \mathrm{P}(c \mid x_1, \ldots, x_n)$$

Estructura de diferentes clasificadores Bayesianos dependiendo de la factorización de

$$P(X_1 = x_1, \ldots, X_n = x_n \mid C = c)$$



Naïve Bayes

Formulación clásica de un problema de diagnóstico

• m diagnósticos posibles no excluyentes

	X_1,	… ,	X_n	Y_1,	… ,	Y_m
$(x^{(1)}, y^{(1)})$	$x_1^{(1)}$	…	$x_n^{(1)}$	$y_1^{(1)}$	…	$y_m^{(1)}$
$(x^{(2)}, y^{(2)})$	$x_1^{(2)}$	…	$x_n^{(2)}$	$y_1^{(2)}$	…	$y_m^{(2)}$
…		…			…	
$(x^{(N)}, y^{(N)})$	$x_1^{(N)}$	…	$x_n^{(N)}$	$y_1^{(N)}$	…	$y_m^{(N)}$

N es el número de patrones, n el número de características y m el número de clases no excluyentes, las características son variables binarias y las variables de clase también

Formulación clásica de un problema de diagnóstico: Clases no excluyentes

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

$$\mathrm{P}(Y_1 = y_1,\ldots,Y_m = y_m \mid X_1 = x_1,\ldots,X_n = x_n) =$$

$$= \frac{\mathrm{P}((Y_1 = y_1,\ldots,Y_m = y_m) \cap (X_1 = x_1,\ldots,X_n = x_n))}{P(X_1 = x_1,\ldots,X_n = x_n)} \propto$$

$$\propto P(Y_1 = y_1,\ldots,Y_m = y_m)\,\mathrm{P}(X_1 = x_1,\ldots,X_n = x_n \mid Y_1 = y_1,\ldots,Y_m = y_m)$$

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

m	n		$parámetros$
3	10	$= 8191 \simeq$	8×10^3
5	20	$\simeq$	33×10^6
10	50	$\simeq$	11×10^{17}

m es el número de clases no excluyentes y n el número de características

Diagnósticos excluyentes, una sola clase de salida

$$c^* = \arg\ \text{máx}\ \mathrm{P}(C = c \mid X_1 = x_1, \dots, X_n = x_n)$$

$$P(C = c \mid X_1 = x_1, \dots, X_n = x_n)$$
$$\propto P(C = c)\,\mathrm{P}(X_1 = x_1, \dots, X_n = x_n \mid C = c)$$

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

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

m es el número de clases excluyentes
y n el número de características

Diagnósticos excluyentes y variables condicionalmente independientes dado el diagnóstico (Naïve Bayes)

$$c^* = \arg \max_c \mathrm{P}(C = c \mid X_1 = x_1, \ldots, X_n = x_n) =$$

$$= \arg \max_c P(C = c) \prod_{i=1}^{n} \mathrm{P}(X_i = x_i \mid C = c)$$

número de parámetros a estimar: $(m-1) + mn$

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

m es el número de clases excluyentes
y n el número de características

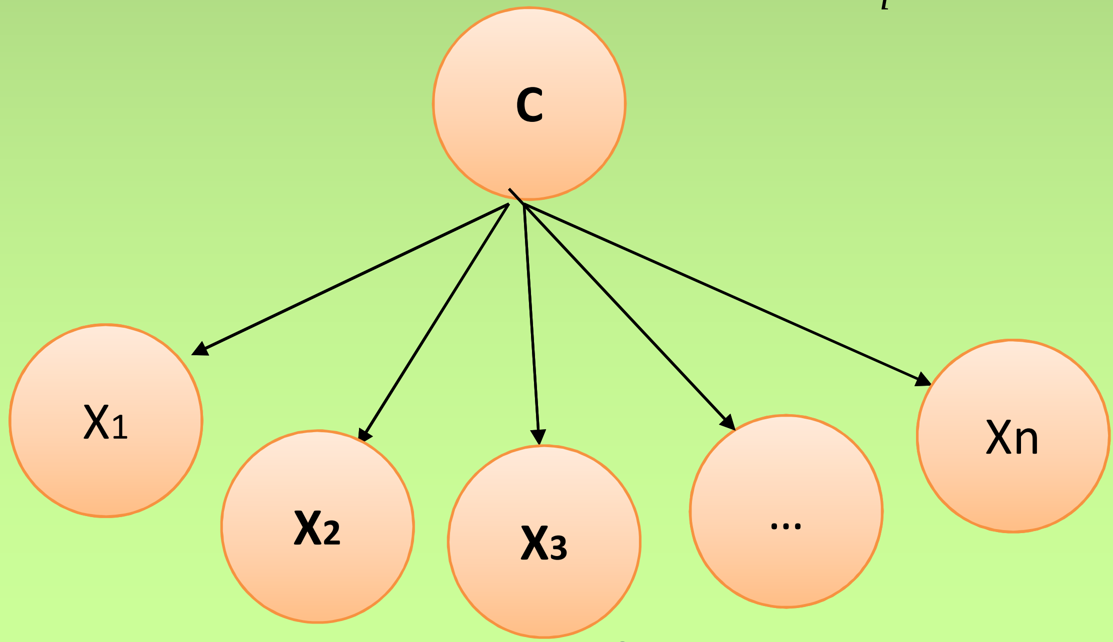
Variables predictoras condicionalmente independientes dado que C=c

Predictoras discretas

$$c^* = \arg\max_c P(C=c).\prod_{i=1}^{n} P(X_i = x_i \mid C=c)$$

Predictoras continuas y normales

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



Estructura de un clasificador Naïve Bayes

Seminaïve Bayes: Pseudocódigo del algoritmo *FSSJ* (Pazzani, 1997)

Paso 1. Inicializar el conjunto de variables a utilizar a conjunto vacío.

Clasificar todos los ejemplos como pertenecientes a la clase más frecuente.

Paso 2. 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.

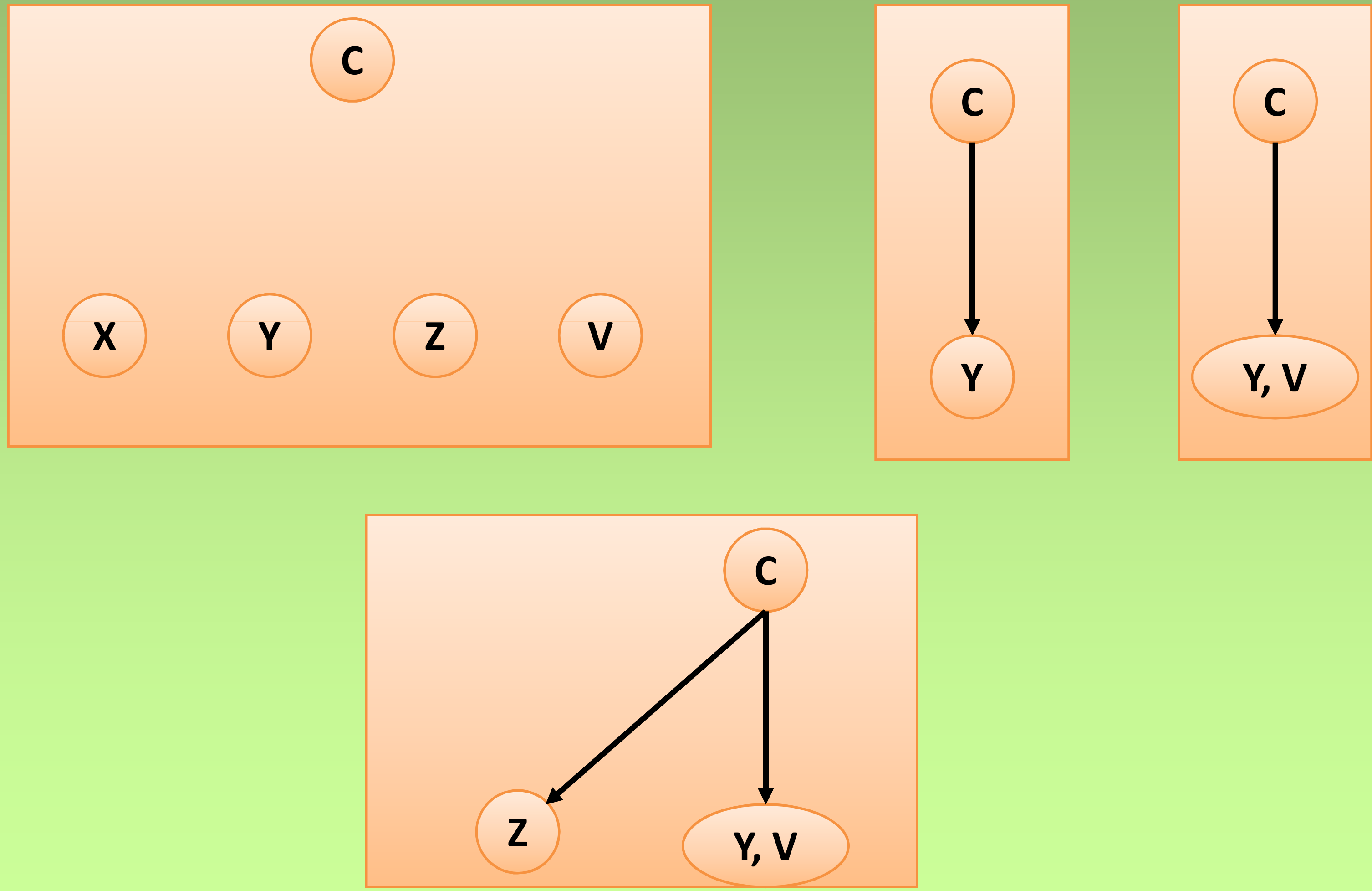
Evaluar cada posible opción por medio de la estimación del porcentaje de patrones bien clasificados.

Hasta que ninguna opción produzca mejoras

Seminaïve Bayes

Proceso de construcción de un modelo seminaïve Bayes.

$$P(c|x, y, z, v) \propto P(c).P(z|c).P((y, v)|c)$$



Naïve Bayes aumentando el árbol

La cantidad de información mutua entre X e Y

$$I(X,Y)=\sum_{i=1}^{r_X}\sum_{j=1}^{r_Y}\mathrm{P}(x_i,y_j)\log\frac{\mathrm{P}(x_i,y_j)}{\mathrm{P}(x_i).\mathrm{P}(y_j)}$$

mide la reducción de la incertidumbre de una de las variables cuando se conoce la otra

Cantidad de información mutua entre X e Y condicionada a C

$$I(X,Y\,|\,C)=\sum_{c}\mathrm{P}(C=c)I(X,Y\,|\,C=c)$$

$$=\sum_{i=1}^{r_X}\sum_{j=1}^{r_Y}\sum_{k=1}^{r_0}\mathrm{P}(x_i,y_j,c_k)\log\frac{\mathrm{P}(x_i,y_j\,|\,c_k)}{\mathrm{P}(x_i\,|\,c_k)\mathrm{P}(y_j\,|\,c_k)}$$

Relación entre la cantidad de información mutua y las verosimilitudes o probabilidades condicionadas a c_k

Naïve Bayes aumentado en árbol

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

Paso 1. Calcular $I(X_i, X_j \mid C)$ con $i < j$, $i, j = 1, \ldots, n$

Paso 2. Construir un grafo no dirigido completo cuyos nodos corresponden a las variables predictoras : $X_1, \ldots, X_n$. Asignar a cada arista conectando las variables X_i y X_j un peso dado por $I(X_i, X_j \mid C)$

Paso 3. Asignar las dos aristas de mayor peso al árbol a construir

Paso 4. 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

Paso 5. Repetir el paso 4 hasta seleccionar $n - 1$ aristas

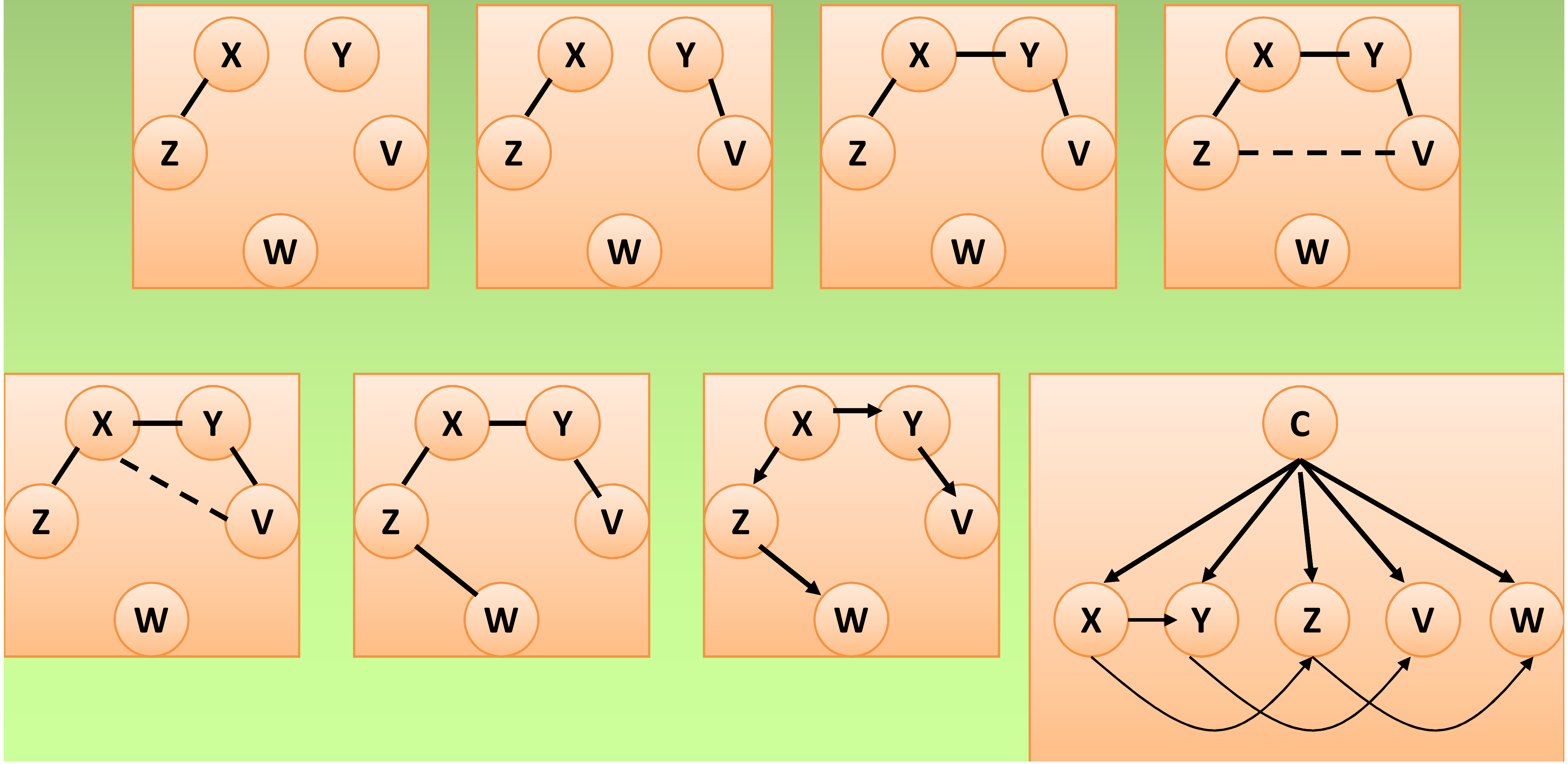
Paso 6. Transformar el árbol no dirigido resultante en uno dirigido, escogiendo una variable como raíz, para a continuación direccionar el resto de aristas

Paso 7. Construir un modelo *TAN* añadiendo un nodo etiquetado como C y posteriormente un arco desde C a cada variable predictora X_i

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)

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



Clasificador Bayesiano k-dependiente (Sahami, 1996)

- **Precalcula I(Xi,C) y I(Xi,Xj |C) para todo par de variables**

- **Añade en cada iteración, de entre las variables que no están en el modelo, aquella Xmax que tenga mayor I(Xi,C)**

- **Asigna a la variable añadida como padres la variable C y aquellas k variables con mayor I(Xj,Xmax|C)**

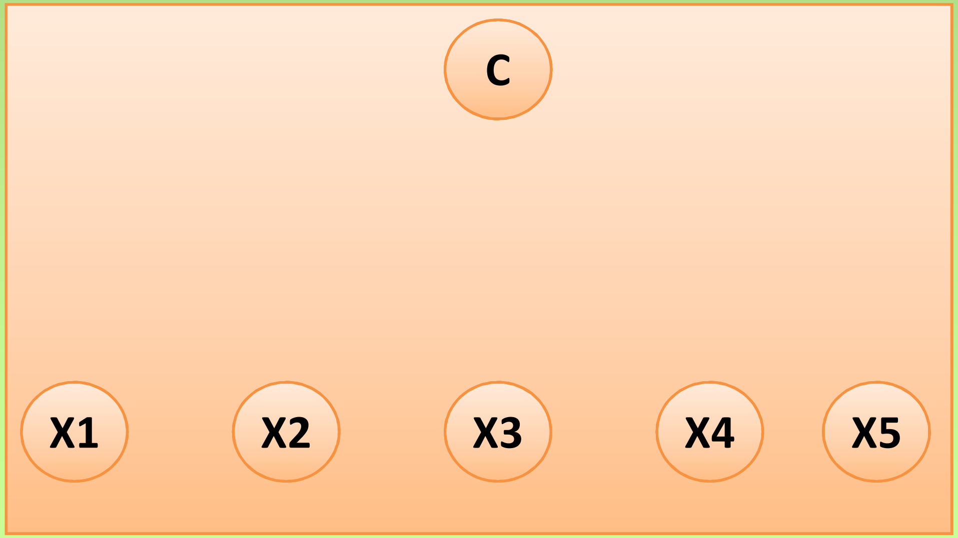
Clasificador Bayesiano k-dependiente

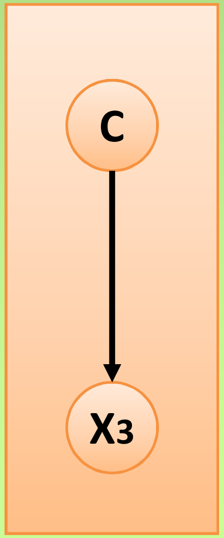
Proceso de construcción de kDB con k = 2.

I(X3,C) > I(X1,C) > I(X4,C) > I(X5,C) > I(X2,C)

I(X3,X4|C) > I(X2,X5|C) > I(X1,X3|C) > I(X1,X2|C) > I(X2,X4|C) >

I(X2,X3|C) > I(X1,X4|C) > I(X4,X5|C) > I(X1,X5|C) > I(X3,X5|C)





Clasificador Bayesiano k-dependiente

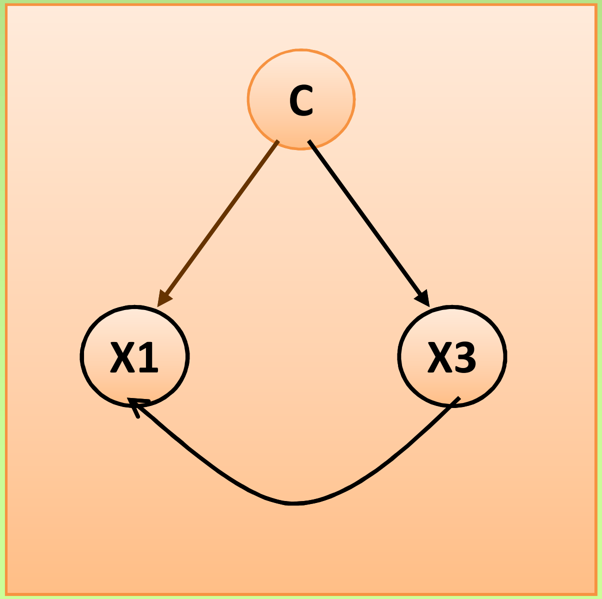
Proceso de construcción de kDB con k = 2.

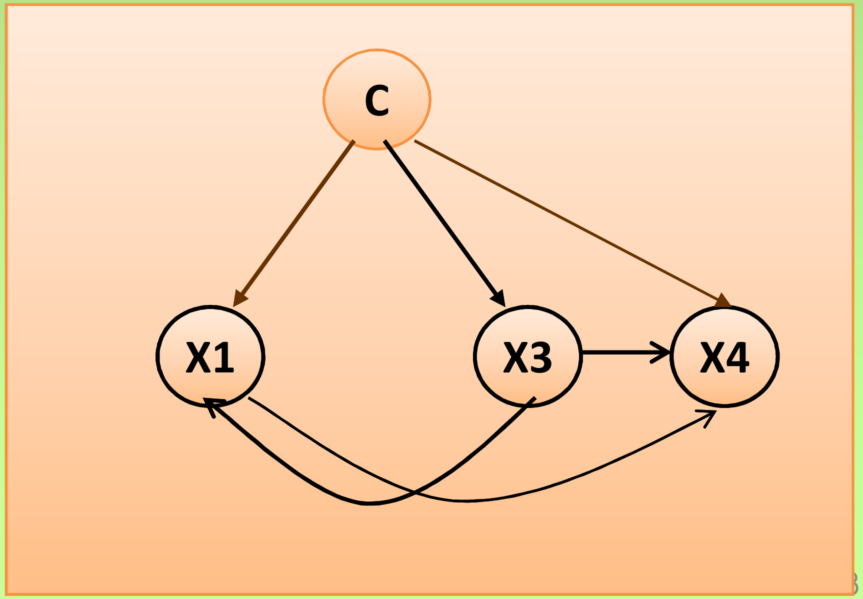
I(X3,C) > I(X1,C) > I(X4,C) > I(X5,C) > I(X2,C)

I(X3,X4|C) > I(X2,X5|C) > I(X1,X3|C) > I(X1,X2|C) > I(X2,X4|C) >

I(X2,X3|C) > I(X1,X4|C) > I(X4,X5|C) > I(X1,X5|C) > I(X3,X5|C)

P(c|x1, x2, x3, x4, x5)





Clasificador Bayesiano k-dependiente

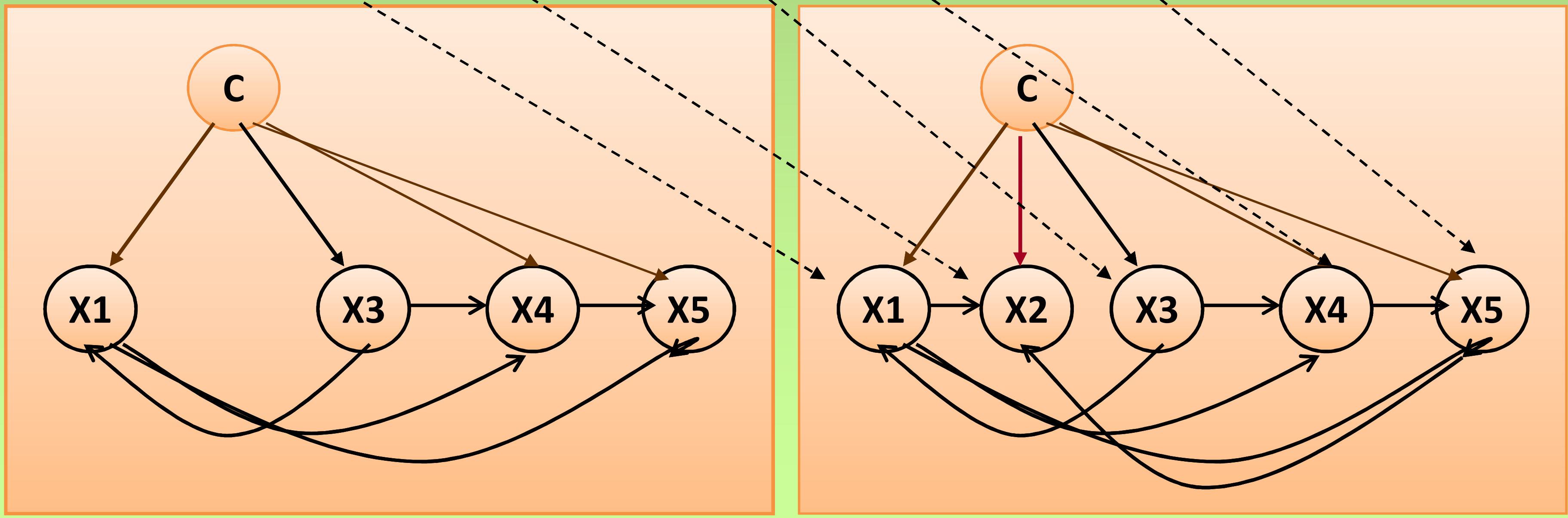
Proceso de construcción de kDB con k = 2.

I(X3,C) > I(X1,C) > I(X4,C) > I(X5,C) > I(X2,C)

I(X3,X4|C) > I(X2,X5|C) > I(X1,X3|C) > I(X1,X2|C) > I(X2,X4|C) > I(X2,X3|C) > I(X1,X4|C) > I(X4,X5|C) > I(X1,X5|C) > I(X3,X5|C)

$$p(c|x1, x2, x3, x4, x5) \propto$$

$$p(c)p(x1|x3, c)p(x2|x1, x5, c)p(x3|c)p(x4|x1, x3, c)p(x5|x1, x4, c)$$



Tutorial de Weka
Clasificador Naïve Bayes

Naive Bayes (NB) (Maron and Kuhns, 1960)

Clasificador OneR

¿Cuál es el mejor predictor?

Outlook	Temp.	Humidity	Windy	Play
sunny	hot	high	false	no
sunny	hot	high	true	no
overcast	hot	high	false	yes
rainy	mild	high	false	yes
rainy	cool	normal	false	yes
rainy	cool	normal	true	no
overcast	cool	normal	true	yes
sunny	mild	high	false	no
sunny	cool	normal	false	yes
rainy				yes
sunny				yes
overcast				yes
overcast				yes
rainy				no

```
=== Classifier model (full training set) ===

Outlook:
IF   sunny    -> no
IF   overcast        -> yes
IF   rainy    -> yes
(10/14 instances correct)
```

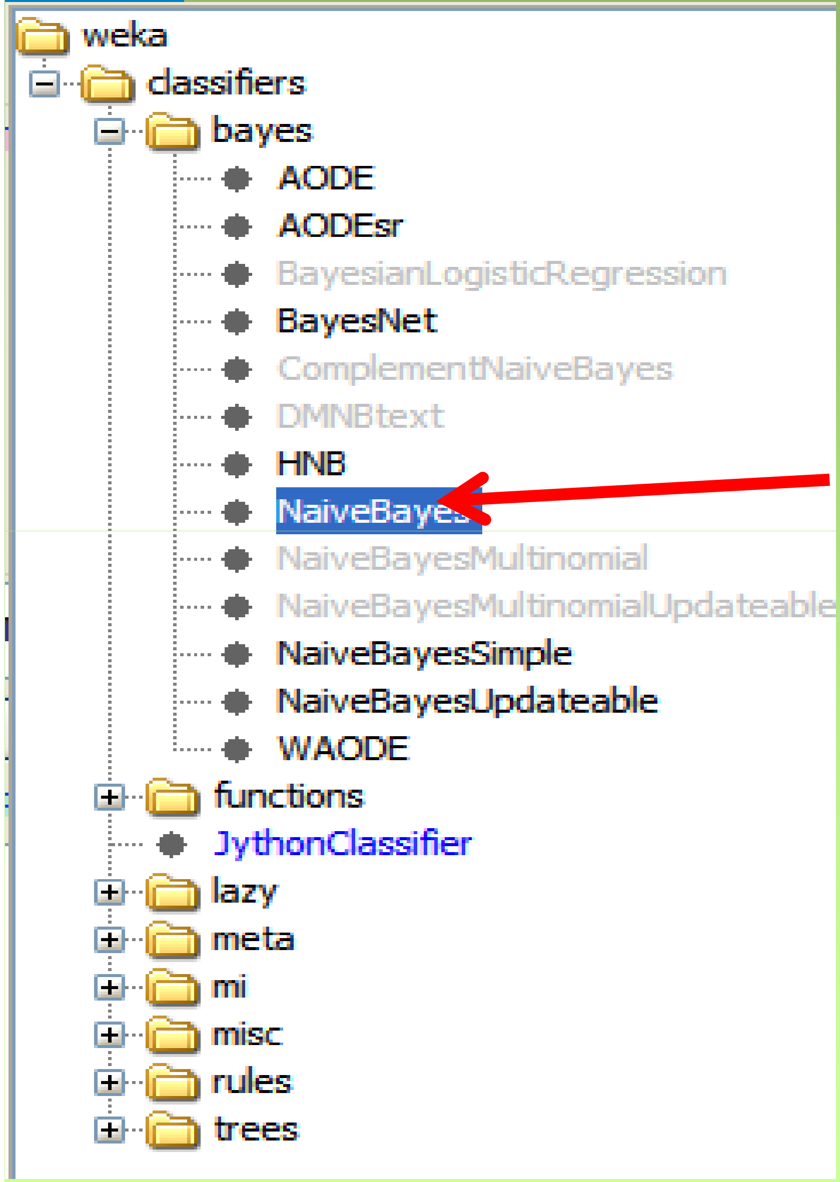
Regla OneR

3 aciertos

3 aciertos

4 aciertos

Total 10 de 14



Es suficiente utilizar la regla OneR o es mejor utilizar el algoritmo naivë Bayes

Clasificador Bayesiano

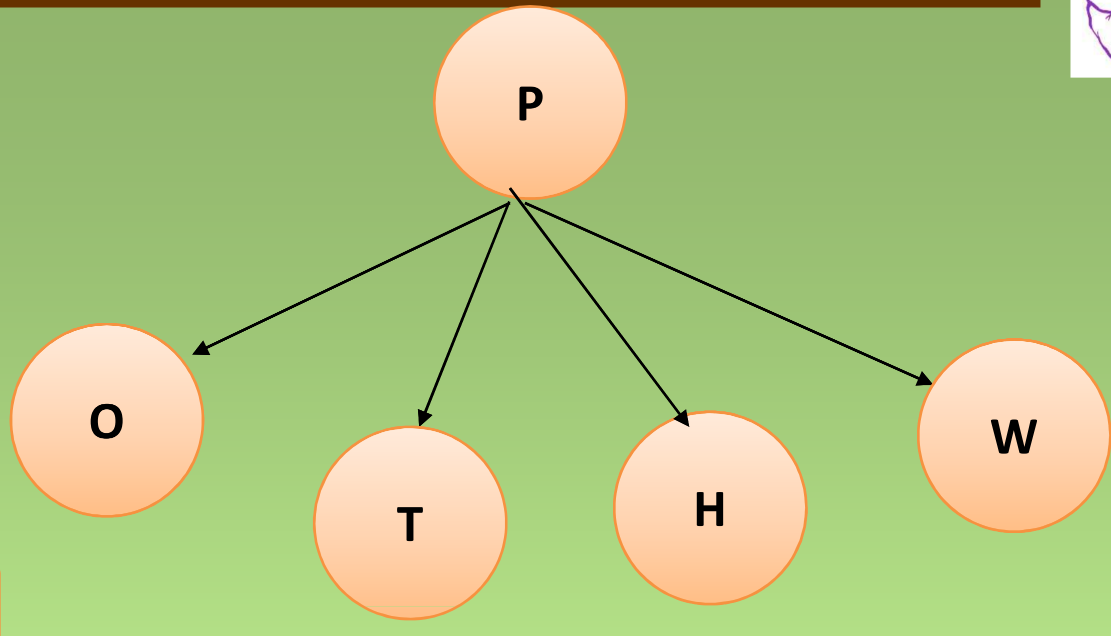
Como de bien el atributo aspecto del cielo “Outlook” Nublado, predice si jugar o no a tenis?

Outlook	Temp.	Humidity	Windy	Play
sunny	hot	high	false	no
sunny	h			no
overcast	h			yes
rainy	m			yes
rainy	c			yes
rainy	c			no
overcast	c			yes
sunny	m			no
sunny	c			yes
rainy	m			yes
sunny	m			yes
overcast	mild	high	true	yes
overcast	hot	normal	false	yes
rainy	mild	high	true	no

	Play	
Outlook	**yes**	**no**
sunny	2	3
overcast	4	0
rainy	3	2
TOTAL	9	5

Clasificador Bayesiano

	Play	
Outlook	**yes**	**no**
sunny	2	3
overcast	4	0
rainy	3	2
TOTAL	9	5



Repetirlo con todos los atributos…

Estructura del clasificador Naïve Bayes

	Play			Play			Play			Play			Play
Outlook	**yes**	**no**	**Temp.**	**yes**	**no**	**Humid.**	**yes**	**no**	**Windy**	**yes**	**no**		
sunny	2	3	hot	2	2	high	3	4	false	6	2	yes	9
overcast	4	0	mild	4	2	normal	6	1	true	3	3	no	5
rainy	3	2	cool	3	1								
TOTAL	9	5	TOTAL	9	5	TOTAL	9	5	TOTAL	9	5	TOTAL	14

Clasificador Bayesiano

	Play	
Outlook	yes	no
sunny	2	3
overcast	4	0
rainy	3	2
TOTAL	9	5

	Play	
Temp.	yes	no
hot	2	2
mild	4	2
cool	3	1
TOTAL	9	5

	Play	
Humid.	yes	no
high	3	4
normal	6	1
TOTAL	9	5

	Play	
Windy	yes	no
false	6	2
true	3	3
TOTAL	9	5

	Play
yes	9
no	5
TOTAL	14

P(Outlook = Sunny | PlayGolf =Yes)
=2/9 = 0.22

Convertir los valores a proporciones o probabilidades

	Play	
Outlook	yes	no
sunny	0.22	0.60
overcast	0.44	0.00
rainy	0.33	0.40

	Play	
Temp.	yes	no
hot	0.22	0.40
mild	0.44	0.40
cool	0.33	0.20

	Play	
Humid.	yes	no
high	0.33	0.80
normal	0.67	0.20

	Play	
Windy	yes	no
false	0.67	0.40
true	0.33	0.60

	Play
yes	0.64
no	0.36

2/5=0.40

2 casos de Play = no, dado que Outlook = rainy (casos favorables)

5 casos de Play = no (casos posibles)

Clasificador Bayesiano

Ejemplo: Calcular la verosimilitud de JUGAR bajo las siguientes condiciones atmosféricas:

Calcular la verosimilitud de que

Outlook = sunny (0.22)

Temperature = cool (0.33)

Humidity = high (0.33)

Windy = true (0.33)

Play = yes (0.64)

0.22 x 0.33 x 0.33 x 0.33 x 0.64 = **0.0053**

Verosimilitud de jugar bajo las condiciones atmosféricas dadas, bajo la hipótesis de que son sucesos condicionalmente independientes.

P(O|T,H,W,Yes)=P(O|Yes)

	Play			Play			Play			Play			Play
Outlook	**yes**	**no**	**Temp.**	**yes**	**no**	**Humid.**	**yes**	**no**	**Windy**	**yes**	**no**		
sunny	0.22	0.60	hot	0.22	0.40	high	0.33	0.80	false	0.67	0.40	yes	0.64
overcast	0.44	0.00	mild	0.44	0.40	normal	0.67	0.20	true	0.33	0.60	no	0.36
rainy	0.33	0.40	cool	0.33	0.20								

Ejemplo: Calcular la verosimilitud de NO JUGAR bajo estas condiciones atmosféricas:

Outlook = sunny (0.60)
Temperature = cool (0.20)
Humidity = high (0.80)
Windy = true (0.60)
Play = no (0.36)

0.60 x 0.20 x 0.80 x 0.60 x 0.36 = **0.0206**

Verosimilitud de NO JUGAR bajo las condiciones atmosféricas dadas, bajo la hipótesis de que son sucesos condicionalmente independientes.

	Play			Play			Play			Play			Play
Outlook	**yes**	**no**	**Temp.**	**yes**	**no**	**Humid.**	**yes**	**no**	**Windy**	**yes**	**no**		
sunny	0.22	0.60	hot	0.22	0.40	high	0.33	0.80	false	0.67	0.40	yes	0.64
overcast	0.44	0.00	mild	0.44	0.40	normal	0.67	0.20	true	0.33	0.60	no	0.36
rainy	0.33	0.40	cool	0.33	0.20								

Clasificador Bayesiano

Probabilidad de jugar y de no jugar bajo las siguientes condiciones atmosféricas:

Dadas las siguientes condiciones atmosféricas:

- Outlook = sunny
- Temperature = cool
- Humidity = high
- Windy = true

Probabilidad de Play = yes: $\frac{0.0053}{0.0053 + 0.0206} =$ **20.5%**

Cálculo de las verosimilitudes

Probabilidad de Play = no: $\frac{0.0206}{0.0053 + 0.0206} =$ **79.5%**

Ejemplo

Verosimilitud de <u>NO jugar</u> bajo estas condiciones atmosféricas:

Calcular la verosimilitud de que

- Outlook = overcast (0.00)
- Temperature = cool (0.20)
- Humidity = high (0.80)
- Windy = true (0.60)
- Play = no (0.36)

El juego nunca fué cancelado en un día nublado

0.00 x 0.20 x 0.80 x 0.60 x 0.36 = 0.0000

	Play			Play			Play			Play			Play
Outlook	yes	no	Temp.	yes	no	Humid.	yes	no	Windy	yes	no		
sunny	0.22	0.60	hot	0.22	0.40	high	0.33	0.80	false	0.67	0.40	yes	0.64
overcast	0.44	0.00	mild	0.44	0.40	normal	0.67	0.20	true	0.33	0.60	no	0.36
rainy	0.33	0.40	cool	0.33	0.20								

Counts - from dataset

Outlook	Play yes	Play no	Temp.	Play yes	Play no	Humid.	Play yes	Play no	Windy	Play yes	Play no		Play
sunny	2	3	hot	2	2	high	3	4	false	6	2	yes	9
overcast	4	0	mild	4	2	normal	6	1	true	3	3	no	5
rainy	3	2	cool	3	1								
TOTAL	9	5	TOTAL	9	5	TOTAL	9	5	TOTAL	9	5	TOTAL	14

estimador de Laplace:
Añadir una unidad a cada valor

Add 1 to count (Laplace estimator)

Outlook	Play yes	Play no	Temp.	Play yes	Play no	Humid.	Play yes	Play no	Windy	Play yes	Play no		Play
sunny	3	4	hot	3	3	high	4	5	false	7	3	yes	12
overcast	5	1	mild	5	3	normal	7	2	true	4	4	no	8
rainy	4	3	cool	4	2								
TOTAL	12	8	TOTAL	12	8	TOTAL	11	7	TOTAL	11	7	TOTAL	20

	Play			Play			Play			Play			Play
Outlook	**yes**	**no**	**Temp.**	**yes**	**no**	**Humid.**	**yes**	**no**	**Windy**	**yes**	**no**		
sunny	3	4	hot	3	3	high	4	5	false	7	3	yes	9
overcast	5	1	mild	5	3	normal	7	2	true	4	4	no	5
rainy	4	3	cool	4	2								
TOTAL	12	8	TOTAL	12	8	TOTAL	11	7	TOTAL	11	7	TOTAL	14

Convertir los valores incrementados a proporciones o probabilidades

	Play			Play			Play			Play			Play
Outlook	**yes**	**no**	**Temp.**	**yes**	**no**	**Humid.**	**yes**	**no**	**Windy**	**yes**	**no**		
sunny	0.25	0.50	hot	0.25	0.38	high	0.36	0.71	false	0.64	0.43	yes	0.64
overcast	0.42	0.13	mild	0.42	0.38	normal	0.64	0.29	true	0.36	0.57	no	0.36
rainy	0.33	0.38	cool	0.33	0.25								

Se mantienen las probabilidades de pertenencia a cada clase

Con estimador de Laplace

	Play			Play			Play			Play			Play
Outlook	**yes**	**no**	**Temp.**	**yes**	**no**	**Humid.**	**yes**	**no**	**Windy**	**yes**	**no**		
sunny	0.25	0.50	hot	0.25	0.38	high	0.36	0.71	false	0.64	0.43	yes	0.64
overcast	0.42	0.13	mild	0.42	0.38	normal	0.64	0.29	true	0.36	0.57	no	0.36
rainy	0.33	0.38	cool	0.33	0.25								

Outlook = overcast, Temperature = cool, Humidity = high, Windy = true

Verosimilitud

Play = no: 0.13 x 0.25 x 0.71 x 0.57 x 0.36 = 0.0046

Play = yes: 0.42 x 0.33 x 0.36 x 0.36 x 0.64 = 0.0118

Probabilidad de Play = no: $\frac{0.0046}{0.0046 + 0.0118} = 28\%$

Probabilidad de Play = yes: $\frac{0.0118}{0.0046 + 0.0118} = 72\%$

Previamente 100%, porque la probabilidad de No jugar era 0

Con estimador de Laplace

	Play			Play			Play			Play			Play
Outlook	**yes**	**no**	**Temp.**	**yes**	**no**	**Humid.**	**yes**	**no**	**Windy**	**yes**	**no**		
sunny	0.25	0.50	hot	0.25	0.38	high	0.36	0.71	false	0.64	0.43	yes	0.64
overcast	0.42	0.13	mild	0.42	0.38	normal	0.64	0.29	true	0.36	0.57	no	0.36
rainy	0.33	0.38	cool	0.33	0.25								

Outlook = sunny, Temperature = cool, Humidity = high, Windy = true

Verosimilitud

Play = no: 0.50 x 0.25 x 0.71 x 0.57 x 0.36 =0,0182115

Play = yes: 0.25 x 0.33 x 0.36 x 0.36 x 0.64 =0,00684288

Probabilidad de Play = no: $\frac{0.0182115}{0.0182115 + 0.00684288}$ **=72,68%**

Probabilidad de Play = yes: $\frac{0.00684288}{0.00684288 + 0.0182115}$ **= 27,31%**

Probabilidad de jugar y de no jugar bajo las siguientes condiciones atmosféricas:

Dado:

- **Outlook = sunny**
- **Temperature = cool**
- **Humidity = high**
- **Windy = true**

<u>Sin utilizar el estimador de Laplace</u>

Play = no: 79.5%
Play = yes: 20.5%

<u>Utilizando el estimador de Laplace</u>

Play = no: 72.68%
Play = yes: 27,31%

En ambos casos la regla equivalente es la misma:

IF Outlook = sunny
AND Temperature = cool
AND Humidity = high
AND Windy = true
THEN Play = no.

Verdadero utilizando ambos cálculos

Dado:

Outlook = sunny

Temperature = cool

Humidity = high

Windy = true

Pregunta: Jugaran a tenis?

Dos posibles reglas:

- Regla 1: Si Outlook = sunny Y Temperature = cool Y Humidity = high Y Windy = true, Entonces Play = no.
- Regla 2: Si Outlook = sunny Y Temperature = cool Y Humidity = high Y Windy = true, Entonces Play = yes.

Probabilidades de que las reglas sean correctas :

- Regla 1 = 72%
- Regla 2 = 28%

Por tanto: Utilizar la Regla 1 para estas condiciones meteorológicas.

Ahora se repite para todas las otras combinaciones de condiciones

Outlook	Temp.	Humid.	Windy	Play
overcast	cool	high	false	no
overcast	cool	high	false	yes
overcast	cool	high	true	no
overcast	cool	high	true	yes
overcast	cool	normal	false	no
overcast	cool	normal	false	yes
overcast	cool	normal	true	no
overcast	cool	normal	true	yes
overcast	hot	high	false	no
overcast	hot	high	false	yes
overcast	hot	high	true	no
overcast	hot	high	true	yes
overcast	hot	normal	false	no
overcast	hot	normal	false	yes
overcast	hot	normal	true	no
overcast	hot	normal	true	yes

Repetir todos los cálculos anteriores para todas las posibles combinaciones de condiciones atmosféricas

	Play			Play			Play			Play			Play
Outlook	**yes**	**no**	**Temp.**	**yes**	**no**	**Humid.**	**yes**	**no**	**Windy**	**yes**	**no**		
sunny	0.25	0.50	hot	0.25	0.38	high	0.36	0.71	false	0.64	0.43	yes	0.64
overcast	0.42	0.13	mild	0.42	0.38	normal	0.64	0.29	true	0.36	0.57	no	0.36
rainy	0.33	0.38	cool	0.33	0.25								

Calculo de las probabilidades de las 36 posibles combinaciones=3x3x2x2

Inst	Outlook	Temp.	Humid.	Windy	Play	Outlook	Temp.	Humid.	Windy	Play	Like.	Prob.
	overcast	cool	high	false	no	0.13	0.25	0.71	0.43	0.36	0.0034	14.2%
	overcast	cool	high	false	yes	0.42	0.33	0.36	0.64	0.64	0.0207	85.8%
	overcast	cool	high	true	no	0.13	0.25	0.71	0.57	0.36	0.0046	27.8%
	overcast	cool	high	true	yes	0.42	0.33	0.36	0.36	0.64	0.0118	72.2%
	overcast	cool	normal	false	no	0.13	0.25	0.29	0.43	0.36	0.0014	3.6%
	overcast	cool	normal	false	yes	0.42	0.33	0.64	0.64	0.64	0.0362	96.4%
	overcast	cool	normal	true	no	0.13	0.25	0.29	0.57	0.36	0.0018	8.1%
7	overcast	cool	normal	true	yes	0.42	0.33	0.64	0.36	0.64	0.0207	91.9%
	overcast	hot	high	false	no	0.13	0.38	0.71	0.43	0.36	0.0051	24.9%
3	overcast	hot	high	false	yes	0.42	0.25	0.36	0.64	0.64	0.0155	75.1%

Inst	Outlook	Temp.	Humid.	Windy	Play	Prob.
	overcast	cool	normal	false	yes	96.4%
	overcast	mild	normal	false	yes	95.7%
13	overcast	hot	normal	false	yes	93.0%
7	overcast	cool	normal	true	yes	91.9%
	overcast	mild	normal	true	yes	90.4%
5	rainy	cool	normal	false	yes	87.6%
	overcast	cool	high	false	yes	85.8%
10	rainy	mild	normal	false	yes	85.5%
	overcast	hot	normal	true	yes	85.0%
2	sunny	hot	high	true	no	83.7%
	overcast	mild	high	false	yes	83.4%
9	sunny	cool	normal	false	yes	79.9%
	rainy	hot	normal	false	yes	77.9%
	sunny	mild	normal	false	yes	76.8%
	sunny	mild	high	true	no	75.5%
3	overcast	hot	high	false	yes	75.1%
	rainy	cool	normal	true	yes	75.1%
	rainy	hot	high	true	no	74.3%

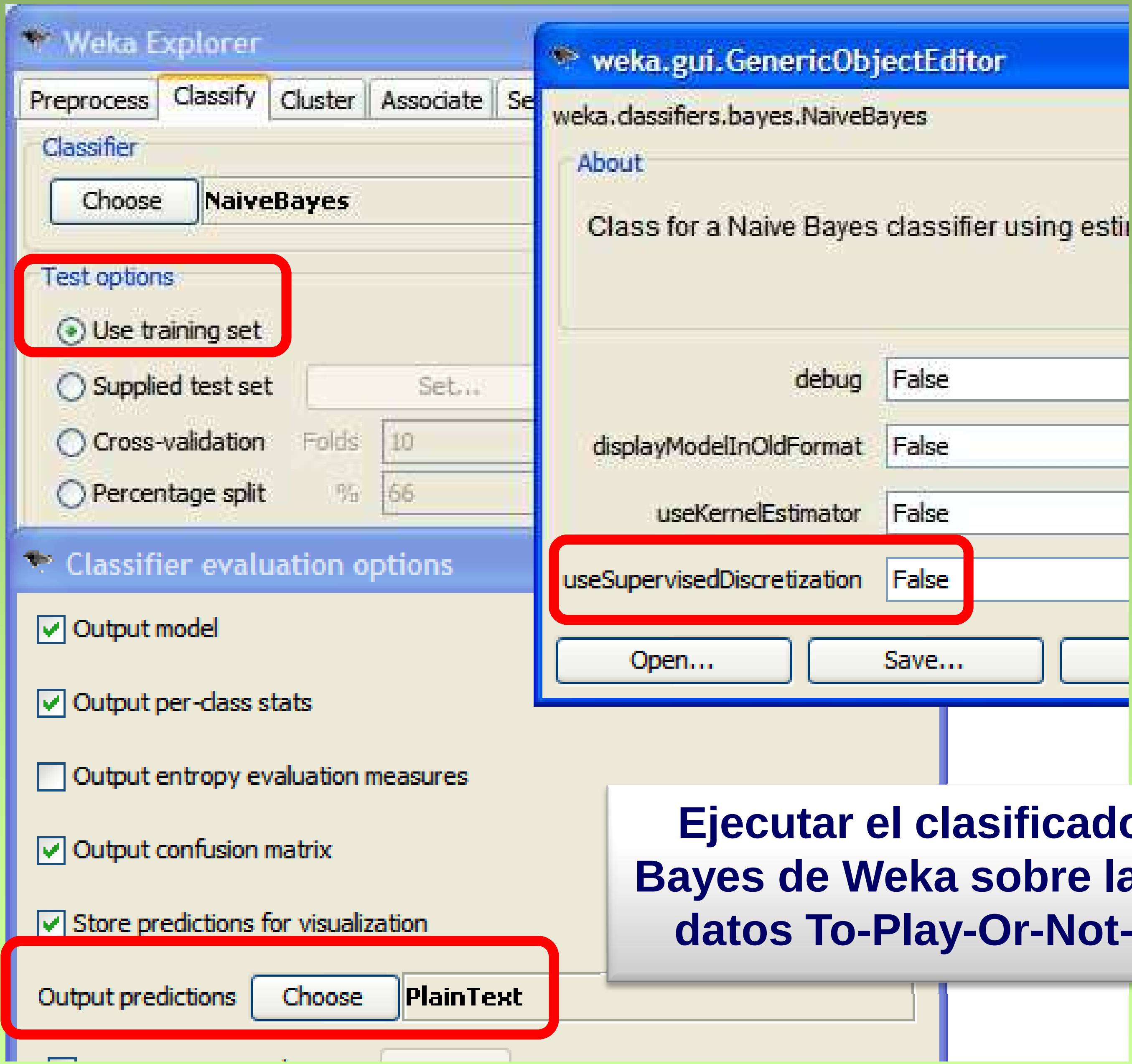
Inst	Outlook	Temp.	Humid.	Windy	Play	Prob.
	overcast	cool	high	true	yes	72.2%
	sunny	cool	high	true	no	72.0%
	rainy	mild	normal	true	yes	71.6%
1	sunny	hot	high	false	no	68.8%
12	overcast	mild	high	true	yes	68.4%
	sunny	hot	normal	false	yes	66.5%
14	rainy	mild	high	true	no	63.5%
	sunny	cool	normal	true	yes	63.0%
	rainy	cool	high	false	yes	61.7%
	rainy	hot	normal	true	yes	60.2%
	rainy	cool	high	true	no	59.1%
11	sunny	mild	normal	true	yes	58.6%
4	rainy	mild	high	false	yes	57.3%
8	sunny	mild	high	false	no	57.0%
	overcast	hot	high	true	yes	56.4%
	rainy	hot	high	false	no	55.4%
	sunny	hot	normal	true	no	54.0%
	sunny	cool	high	false	no	52.4%

Reglas de predicción de clase para todas las combinaciones de atributos.

¿Donde esta la instancia 6?

Comparación de decisiones predichas y actuales (o reales)

Inst	Outlook	Temp.	Humid.	Windy	Play	Prob.	Actual
1	sunny	hot	high	false	no	72.6%	no
2	sunny	hot	high	true	no	86.1%	no
3	overcast	hot	high	false	yes	71.6%	yes
4	rainy	mild	high	false	yes	52.8%	yes
5	rainy	cool	normal	false	yes	85.5%	yes
6	rainy	cool	normal	true	yes	75.1%	no
7	overcast	cool	normal	true	yes	90.4%	yes
8	sunny	mild	high	false	no	61.4%	no
9	sunny	cool	normal	false	yes	76.8%	yes
10	rainy	mild	normal	false	yes	83.0%	yes
11	sunny	mild	normal	true	yes	54.2%	yes
12	overcast	mild	high	true	yes	64.3%	yes
13	overcast	hot	normal	false	yes	91.7%	yes
14	rainy	mild	high	true	no	67.6%	no



Ejecutar el clasificador naïve Bayes de Weka sobre la base de datos To-Play-Or-Not-To-Play

Add 1 to count (Laplace estimator)

	Play			Play			Play			Play			Play
Outlook	yes	no	Temp.	yes	no	Humid.	yes	no	Windy	yes	no		
sunny	3	4	hot	3	3	high	4	5	false	7	3	yes	12
overcast	5	1	mild	5	3	normal	7	2					
rainy	4	3	cool	4	2								
TOTAL	12	8	TOTAL	12	8	TOTAL	11	7					

Classifier output

```
Attribute              no      yes
                    (0.38)  (0.63)
===================================
Outlook
  sunny               4.0      3.0
  overcast            1.0      5.0
  rainy               3.0      4.0
  [total]             8.0     12.0

Temp.
  hot                 3.0      3.0
  mild                3.0      5.0
  cool                2.0      4.0
  [total]             8.0     12.0

Humidity
  high                5.0      4.0
  normal              2.0      7.0
  [total]             7.0     11.0
```

El clasificador ha utilizado el estimador de Laplace

El clasificador naïve Bayes clasifica mal la instancia o patrón 6

Classifier output

```
=== Predictions on training set ===

 inst#     actual  predicted error prediction
     1       1:no       1:no       0.704
     2       1:no       1:no       0.847
     3      2:yes      2:yes       0.737
     4      2:yes      2:yes       0.554
     5      2:yes      2:yes       0.867
     6       1:no      2:yes   +   0.737
     7      2:yes      2:yes       0.913
     8       1:no       1:no       0.588
     9      2:yes      2:yes       0.786
    10      2:yes      2:yes       0.845
    11      2:yes      2:yes       0.568
    12      2:yes      2:yes       0.667
    13      2:yes      2:yes       0.925
    14       1:no       1:no       0.652
```

Errores de predicción

Salida del clasificador o “prediction”

Classifier output

```
=== Predictions on training set ===

 inst#     actual  predicted error prediction
     1       1:no       1:no       0.704
     2       1:no       1:no       0.847
                       2:yes       0.737
                       2:yes       0.554
                       2:yes       0.867
                       2:yes   +   0.737
                       2:yes       0.913
                        1:no       0.588
                       2:yes       0.786
                       2:yes       0.845
                       2:yes       0.568
                       2:yes       0.667
                       2:yes       0.925
                        1:no       0.652
```

Inst	Outlook	Temp.	Humid.	Windy	Play	Prob.	Actual
1	sunny	hot	high	false	no	72.6%	no
2	sunny	hot	high	true	no	86.1%	no
3	overcast	hot	high	false	yes	71.6%	yes
4	rainy	mild	high	false	yes	52.8%	yes
5	rainy	cool	normal	false	yes	85.5%	yes
6	rainy	cool	normal	true	yes	75.1%	no
7	overcast	cool	normal	true	yes	90.4%	yes
8	sunny	mild	high	false	no	61.4%	no
9	sunny	cool	normal	false	yes	76.8%	yes
10	rainy	mild	normal	false	yes	83.0%	yes
11	sunny	mild	normal	true	yes	54.2%	yes
12	overcast	mild	high	true	yes	64.3%	yes
13	overcast	hot	normal	false	yes	91.7%	yes
14	rainy	mild	high	true	no	67.6%	no

Cuestiones críticas de minería de datos:

- **Cómo aplicar los clasificadores.**
- **Mantener el modelo lo más sencillo posible**
- **Discretización.**
- **Impacto de los valores perdidos.**

Como aborda estas cuestiones el clasificador Naive Bayes

¿Donde está el modelo equivalente al clasificador Naïve Bayes de Weka?

```
=== Classifier model (full training set) ===

Outlook:
	sunny	-> no
	overcast		-> yes
	rainy	-> yes
(10/14 instances correct)
```

Conjunto de reglas generado por OneR de Weka

Inst	Outlook	Temp.	Humid.	Windy	Play	Prob.
	overcast	cool	normal	false	yes	96.4%
	overcast	mild	normal	false	yes	95.7%
13	overcast	hot	normal	false	yes	93.0%
7	overcast	cool	normal	true	yes	91.9%
	overcast	mild	normal	true	yes	90.4%
5	rainy	cool	normal	[illegible]	[illegible]	[illegible]
	overcast	cool	high	[illegible]	[illegible]	[illegible]
10	rainy	mild	normal	[illegible]	[illegible]	[illegible]
	overcast	hot	normal	[illegible]	[illegible]	[illegible]
2	sunny	hot	high	[illegible]	[illegible]	[illegible]
	overcast	mild	high	false	yes	83.4%
9	sunny	cool	normal	false	yes	79.9%
	rainy	hot	normal	false	yes	77.9%
	sunny	mild	normal	false	yes	76.8%
	sunny	mild	high	true	no	75.5%
3	overcast	hot	high	false	yes	75.1%
	rainy	cool	normal	true	yes	75.1%
	rainy	hot	high	true	no	74.3%

Inst	Outlook	Temp.	Humid.	Windy	Play	Prob.
	overcast	cool	high	true	yes	72.2%
	sunny	cool	high	true	no	72.0%
	rainy	mild	normal	true	yes	71.6%
1	sunny	hot	high	false	no	68.8%
12	overcast	mild	high	true	yes	68.4%
[illegible]	[illegible]	[illegible]	normal	false	yes	66.5%
[illegible]	[illegible]	[illegible]	high	true	no	63.5%
[illegible]	[illegible]	[illegible]	normal	true	yes	63.0%
[illegible]	[illegible]	[illegible]	high	false	yes	61.7%
[illegible]	[illegible]	[illegible]	normal	true	yes	60.2%
	rainy	cool	high	true	no	59.1%
11	sunny	mild	normal	true	yes	58.6%
4	rainy	mild	high	false	yes	57.3%
8	sunny	mild	high	false	no	57.0%
	overcast	hot	high	true	yes	56.4%
	rainy	hot	high	false	no	55.4%
	sunny	hot	normal	true	no	54.0%
	sunny	cool	high	false	no	52.4%

Reglas que predicen la clase para todas la combinaciones de atributos

Una aplicación con 5 atributos cada uno de ellos con 5 valores tendría 3125 reglas!

	Play			Play			Play			Play			Play
Outlook	yes	no	Temp	yes	no	Humid	yes	no	Windy	yes	no		
sunny	3	4	hot	3	3	high	4	5	false	7	3	yes	9
overcast	5	1	mild	5	3	normal	7	2	true	4	4	no	5
rainy	4	3	cool	4	2								
TOTAL	12	8	TOTAL	12	8	TOTAL	11	7	TOTAL	11	7	TOTAL	14

Conjunto de datos historicos de decisiones tomadas

Retroalimentar los resultados

Predecir en función de las condiciones atmosféricas de hoy: sunny, hot, high, false

Calcular probabilidades

No play!

Nueva instancia de condiciones atmosféricas

Calculo de probabilidades genérico Naïve Bayes

Cuestiones críticas de minería de datos:

- **Cómo aplicar los clasificadores.**
- **Mantener el modelo lo más sencillo posible**
- **Discretización.**
- **Impacto de los valores perdidos.**

¿Que es “ingenuo” del clasificador ingenuo Bayes “Naivë Bayes”?

Respuesta sencilla: atributos o variables aleatorias condicionalmente *Independientes*

Respuesta no tan sencilla : teorema de Bayes

Que significa esto?

Energía solar para sistema de sonido

Outlook	Power
sunny	high
overcast	medium
rainy	low

Outlook	Power
sunny	high
overcast	medium
rainy	low

Añadir un nuevo atributo

Outlook	Temp.	Humidity	Windy	Power	Play
sunny	hot	high	false	high	no
sunny	hot	high	true	high	no
overcast	hot	high	false	medium	yes
rainy	mild	high	false	low	yes
rainy	cool	normal	false	low	yes
rainy	cool	normal	true	low	no
overcast	cool	normal	true	medium	yes
sunny	mild	high	false	high	no
sunny	cool	normal	false	high	yes
rainy	mild	normal	false	low	yes
sunny	mild	normal	true	high	yes
overcast	mild	high	true	medium	yes
overcast	hot	normal	false	medium	yes
rainy	mild	high	true	low	no

El atributo Potencia es una variable dependiente del estado del cielo o "Outlook"

Añadir el atributo potencia

Add 1 to count (Laplace estimator)

	Play			Play			Play			Play			Play			Play
Outlook	**yes**	**no**	**Temp.**	**yes**	**no**	**Humid.**	**yes**	**no**	**Windy**	**yes**	**no**	**Power**	**yes**	**no**		
sunny	3	4	hot	3	3	high	4	5	false	7	3	high	3	4	yes	9
overcast	5	1	mild	5	3	normal	7	2	true	4	4	medium	5	1	no	5
rainy	4	3	cool	4	2							low	4	3		
TOTAL	12	8	TOTAL	12	8	TOTAL	11	7	TOTAL	11	7	TOTAL	12	8	TOTAL	14

Ratios

	Play			Play			Play			Play			Play			Play
Outlook	**yes**	**no**	**Temp.**	**yes**	**no**	**Humid.**	**yes**	**no**	**Windy**	**yes**	**no**	**Power**	**yes**	**no**		
sunny	0.25	0.50	hot	0.25	0.38	high	0.36	0.71	false	0.64	0.43	high	0.25	0.50	yes	0.64
overcast	0.42	0.13	mild	0.42	0.38	normal	0.64	0.29	true	0.36	0.57	medium	0.42	0.13	no	0.36
rainy	0.33	0.38	cool	0.33	0.25							low	0.33	0.38		

Salida de los porcentajes de potencia

SIN Atributo Potencia

Para Play = yes (.64)

Outlook = overcast (.42)

Temperature = cool (.33)

Humidity = high (.36)

Windy = true (.36)

Verosimilitud = **0.42 * 0.33 * 0.36 * 0.36 * 0.64 =** 0.**0118**

CON Atributo Potencia

Para Play = yes (.64)

Outlook = overcast (.42)

Temperature = cool (.33)

Humidity = high (.36)

Windy = true (.36)

Power = medium (.42)

Verosimilitud = **0.42 * 0.33 * 0.36 * 0.36 * 0.42 * 0.64** = 0.**00496**

El efecto de la variable Outlook se aplica dos veces

Nuevas condiciones atmosféricas

Outlook = overcast
Temperature = cool
Humidity = high
Windy = true
Power = medium

SIN Atributo Potencia

Probabilidad de Play = no: $\frac{0.046}{0.0046 + 0.0118} = 28\%$

Probabilidad de Play = yes: $\frac{0.0118}{0.0046 + 0.0118} = 72\%$

CON Atributo Potencia

Probabilidad de Play = no: $\frac{0.00060}{0.00060 + 0.00496} = 10\%$

Probabilidad de Play = yes: $\frac{0.00496}{0.00060 + 0.00496} = 90\%$

Las primeras 18 reglas son las mismas cuando utilizamos atributos dependientes

					NO Power			WITH Power		
Inst	Outlook	Temp.	Humid.	Windy	Play	Prob.	Err	Play	Prob.	Err
	overcast	cool	high	false	yes	85.8%		yes	95.3%	
	overcast	cool	high	true	yes	72.2%		yes	89.6%	
	overcast	cool	normal	false	yes	96.4%		yes	98.9%	
7	overcast	cool	normal	true	yes	91.9%		yes	97.4%	
3	overcast	hot	high	false	yes	75.1%		yes	91.0%	
	overcast	hot	high	true	yes	56.4%		yes	81.2%	
13	overcast	hot	normal	false	yes	93.0%		yes	97.8%	
	overcast	hot	normal	true	yes	85.0%		yes	95.0%	
	overcast	mild	high	false	yes	83.4%		yes	94.4%	
12	overcast	mild	high	true	yes	68.4%		yes	87.8%	
	overcast	mild	normal	false	yes	95.7%		yes	98.7%	
	overcast	mild	normal	true	yes	90.4%		yes	96.9%	
	rainy	cool	high	false	yes	61.7%		yes	58.9%	
	rainy	cool	high	true	no	59.1%		no	61.9%	
5	rainy	cool	normal	false	yes	87.6%		yes	86.2%	
6	rainy	cool	normal	true	yes	75.1%	+	yes	72.9%	+
	rainy	hot	high	false	no	55.4%		no	58.3%	
	rainy	hot	high	true	no	74.3%		no	76.5%	

Las últimas 18 tienen algunas diferencias sobre las reglas de menor probabilidad

					NO Power			WITH Power		
Inst	Outlook	Temp.	Humid.	Windy	Play	Prob.	Err	Play	Prob.	Err
	rainy	hot	normal	false	yes	77.9%		yes	75.8%	
	rainy	hot	normal	true	yes	60.2%		yes	57.3%	
4	rainy	mild	high	false	yes	57.3%		yes	54.4%	
14	rainy	mild	high	true	no	63.5%		no	66.1%	
10	rainy	mild	normal	false	yes	85.5%		yes	83.9%	
	rainy	mild	normal	true	yes	71.6%		yes	69.1%	
	sunny	cool	high	false	no	52.4%		no	68.8%	
	sunny	cool	high	true	no	72.0%		no	83.7%	
9	sunny	cool	normal	false	yes	79.9%		yes	66.5%	
	sunny	cool	normal	true	yes	63.0%		no	54.0%	
1	sunny	hot	high	false	no	68.8%		no	81.5%	
2	sunny	hot	high	true	no	83.7%		no	91.1%	
	sunny	hot	normal	false	yes	66.5%		no	50.2%	
	sunny	hot	normal	true	no	54.0%		no	70.2%	
8	sunny	mild	high	false	no	57.0%		no	72.6%	
	sunny	mild	high	true	no	75.5%		no	86.1%	
	sunny	mild	normal	false	yes	76.8%		yes	62.3%	
11	sunny	mild	normal	true	yes	58.6%		no	58.5%	+

¿Porqué OneR trabaja tan bien?

Entonces ¿por qué Naivë Bayes trabaja tan bien?

Por qué buscar un clasificador que sea sencillo es a menudo la mejor política

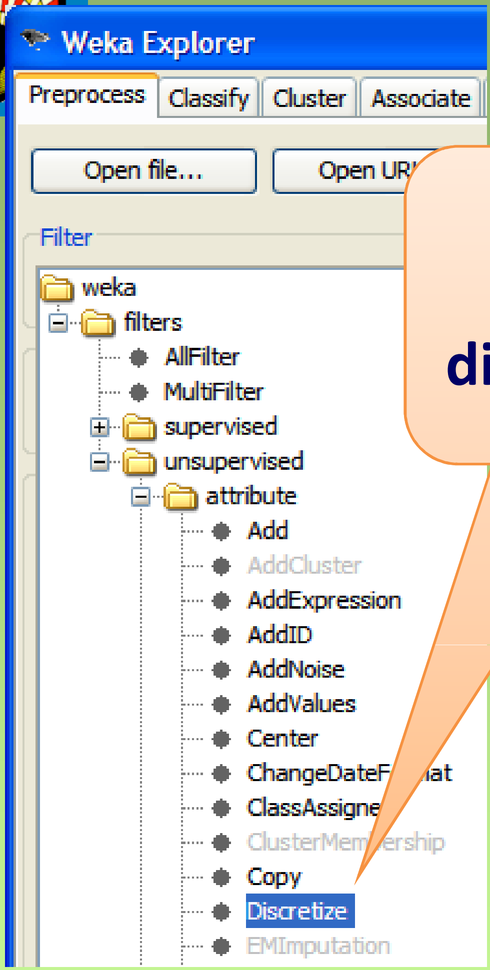
Porque a menudo la clase tiempo está dominada por un solo atributo

Debido a que no se preocupa por las probabilidades, sólo queremos saber si debemos ir a jugar!

Cuestiones críticas de minería de datos:

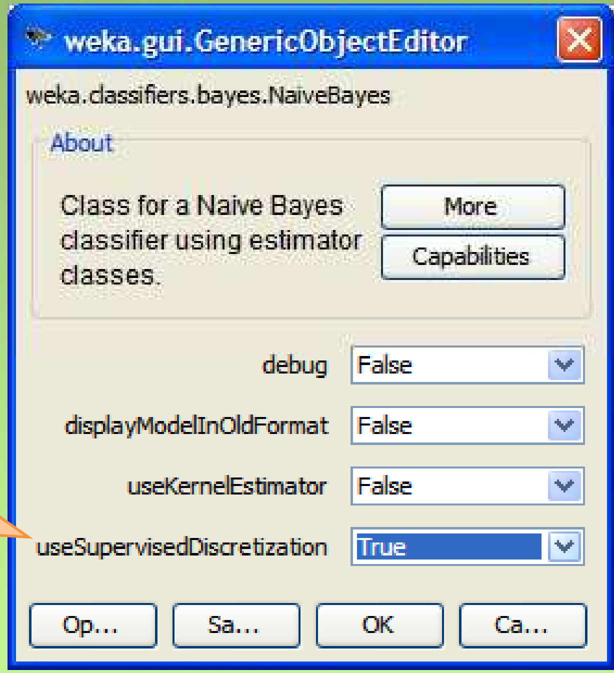
- **Cómo aplicar los clasificadores.**
- **Mantener el modelo lo más sencillo posible**
- **Discretización.**
- **Impacto de los valores perdidos.**

Como se pueden utilizar atributos numéricos en el clasificador naïve Bayes?



Opción 1: Utilizar discretización no supervisada

Opción 2: Utilizar discretización supervisada



Opción 3: Utilizar la capacidad de construir atributos numéricos de NaiveBayes de Weka

	Play			Play			Play			Play			Play
Outlook	**yes**	**no**	**Temp.**	**yes**	**no**	**Humid.**	**yes**	**no**	**Windy**	**yes**	**no**		
sunny	2	3	hot	2	2	high	3	4	false	6	2	yes	9
overcast	4	0	mild	4	2	normal	6	1	true	3	3	no	5
rainy	3	2	cool	3	1								
TOTAL	9	5	TOTAL	9	5	TOTAL	9	5	TOTAL	9	5	TOTAL	14

Empezar contando las ocurrencias de los sucesos (frecuencias absolutas)

Pasar las frecuencias absolutas a probabilidades

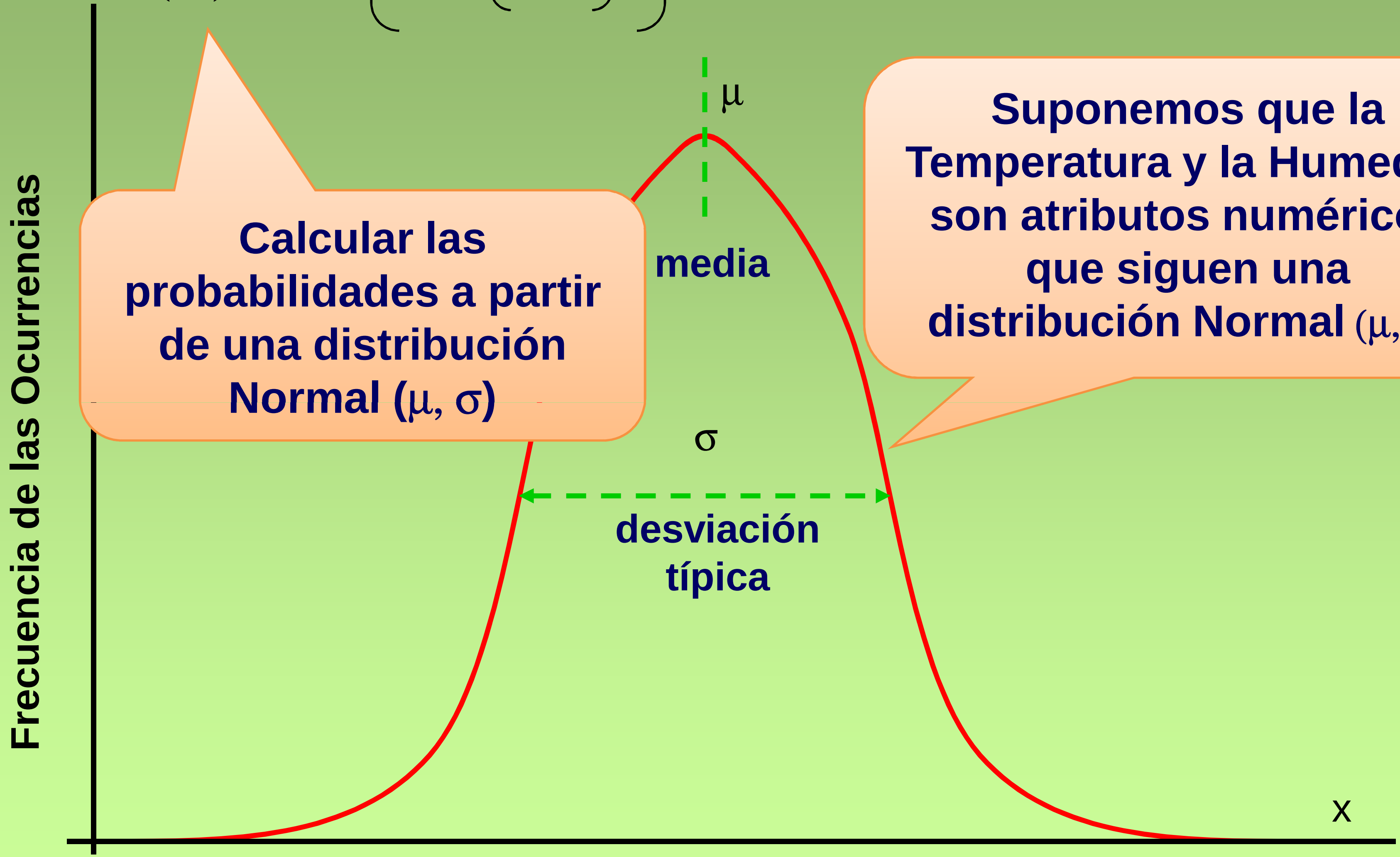
	Play			Play			Play			Play			Play
Outlook	**yes**	**no**	**Temp.**	**yes**	**no**	**Humid.**	**yes**	**no**	**Windy**	**yes**	**no**		
sunny	0.22	0.60	hot	0.22	0.40	high	0.33	0.80	false	0.67	0.40	yes	0.64
overcast	0.44	0.00	mild	0.44	0.40	normal	0.67	0.20	true	0.33	0.60	no	0.36
rainy	0.33	0.40	cool	0.33	0.20								

Calculo de las probabilidades a partir de atributos numéricos

$$\frac{1}{\sigma(2\pi)^{1/2}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^{2}\right)$$

Calcular las probabilidades a partir de una distribución Normal (μ, σ)

Suponemos que la Temperatura y la Humedad son atributos numéricos que siguen una distribución Normal (μ, σ)



	Temperature	
Play =	Yes	No
	83	85
	70	80
	68	65
	64	72
	69	71
	75	
	75	
	72	
	81	
Mean =	73.0	74.6
S.D. =	6.2	7.9

Calcular la probabilidad de que Temp. = 66 cuando Play = yes

$$\frac{1}{\sigma(2\pi)^{1/2}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

$$\frac{1}{6.2(2\pi)^{1/2}} \exp\left(-\frac{1}{2}\left(\frac{66-73.0}{6.2}\right)^2\right) = 0.0340$$

	Temperature	
Play =	Yes	No
	83	85
	70	80
	68	65
	64	72
	69	71
	75	
	75	
	72	
	81	
Mean =	73.0	74.6
S.D. =	6.2	7.9
@ Temp. = 66	0.0340	0.0279

	Humidity	
Play =	Yes	No
	86	85
	96	90
	80	70
	65	95
	70	91
	80	
	70	
	90	
	75	
Mean =	79.1	86.2
S.D. =	10.2	9.7
@ Humid. = 90	0.0221	0.0380

Verosimilitud de Play = yes para atributos nominales y numéricos

Con todos los atributos nominales:

- Outlook = sunny (.22)
- Temperature = cool (.33)
- Humidity = high (.33)
- Windy = true (.33)
- Play = yes (.64) → .22 x .33 x .33 x .33 x .64 = 0.0053

Con atributos numéricos para Temperature = 66 y Humidity = 90:

- Outlook = sunny (.22)
- Temperature = 66 (.0340)
- Humidity = 90 (.0221)
- Windy = true (.33)
- Play = yes (.64) → .22 x .0340 x .0221 x .33 x .64 = 0.000036

Verosimilitud de Play = no para atributos nominales y numéricos

Con todos los atributos nominales:

Outlook = sunny (.60)
Temperature = cool (.20)
Humidity = high (.80)
Windy = true (.60)
Play = no (.36) → .60 x .20 x .80 x .60 x .36 = 0.0206

Con atributos numéricos para Temperature = 66 y Humidity = 90:

Outlook = sunny (.60)
Temperature = 66 (.0279)
Humidity = 90 (.0380)
Windy = true (.60)
Play = no (.36) → .60 x .0279 x .0380 x .60 x .36 = 0.000137

Probabilidad de jugar y de no jugar bajo ciertas condiciones atmosféricas

Dadas esta condiciones atmosféricas y suponiendo distribuciones de probabilidad normales para la temperatura y humedad tenemos:

- **Outlook = sunny**
- **Temperature = 66**
- **Humidity = 90**
- **Windy = true**

Probabilidad de Play = yes:

$$\frac{0.000036}{0.000036 + 0.000137} = 20.8\%$$

Probabilidad de Play = no:

$$\frac{0.000137}{0.000036 + 0.000137} = 79.2\%$$

Cuestiones críticas de minería de datos:

- **Cómo aplicar los clasificadores.**
- **Mantener el modelo lo más sencillo posible**
- **Discretización.**
- **Impacto de los valores perdidos.**

Outlook	Temp.	Humidity	Windy	Play
sunny	hot	high	false	no
sunny	hot	high	true	no
overcast	hot	high	false	yes
rainy	mild		false	yes
rainy	cool	normal	false	yes
rainy	cool	normal	true	no
overcast	cool	normal	true	yes
sunny	mild	high	false	no
	cool	normal	false	yes
rainy	mild	normal	false	yes
sunny	mild	normal	true	yes
overcast	mild	high	true	yes
overcast	hot	normal	false	yes
rainy	mild	high	true	no

Con fines ilustrativos, eliminamos un valor de Outlook = soleado y un valor de Humedad = alta

Un recuento de menos debido a la falta de un valor

	Play	
Outlook	yes	no
sunny	1	3
overcast	4	0
rainy	3	2
TOTAL	8	5

	Play	
Temp.	yes	no
hot	2	2
mild	4	2
cool	3	1
TOTAL	9	5

	Play	
Humid.	yes	no
high	2	4
normal	6	1
TOTAL	8	5

	ay	
W	yes	no
se	6	2
true	3	3
TOTAL	9	5

	Play
yes	9
no	5
TOTAL	14

Counts - no missing data

	Play			Play			Play			Play			Play
Outlook	**yes**	**no**	**Temp.**	**yes**	**no**	**Humid.**	**yes**	**no**	**Windy**	**yes**	**no**		
sunny	2	3	hot	2	2	high	3	4	false	6	2	yes	9
overcast	4	0	mild	4	2	normal	6	1	true	3	3	no	5
rainy	3	2	cool	3	1								
TOTAL	9	5	TOTAL	9	5	TOTAL	9	5	TOTAL	9	5	TOTAL	14

Counts - with missing data

	Play			Play			Play			Play			Play
Outlook	**yes**	**no**	**Temp.**	**yes**	**no**	**Humid.**	**yes**	**no**	**Windy**	**yes**	**no**		
sunny	1	3	hot	2	2	high	2	4	false	6	2	yes	9
overcast	4	0	mild	4	2	normal	6	1	true	3	3	no	5
rainy	3	2	cool	3	1								
TOTAL	8	5	TOTAL	9	5	TOTAL	8	5	TOTAL	9	5	TOTAL	14

APRENDIZAJE AUTOMÁTICO: TERCER CURSO DEL GRADO
DE ING. INFORMÁTICA EN COMPUTACION

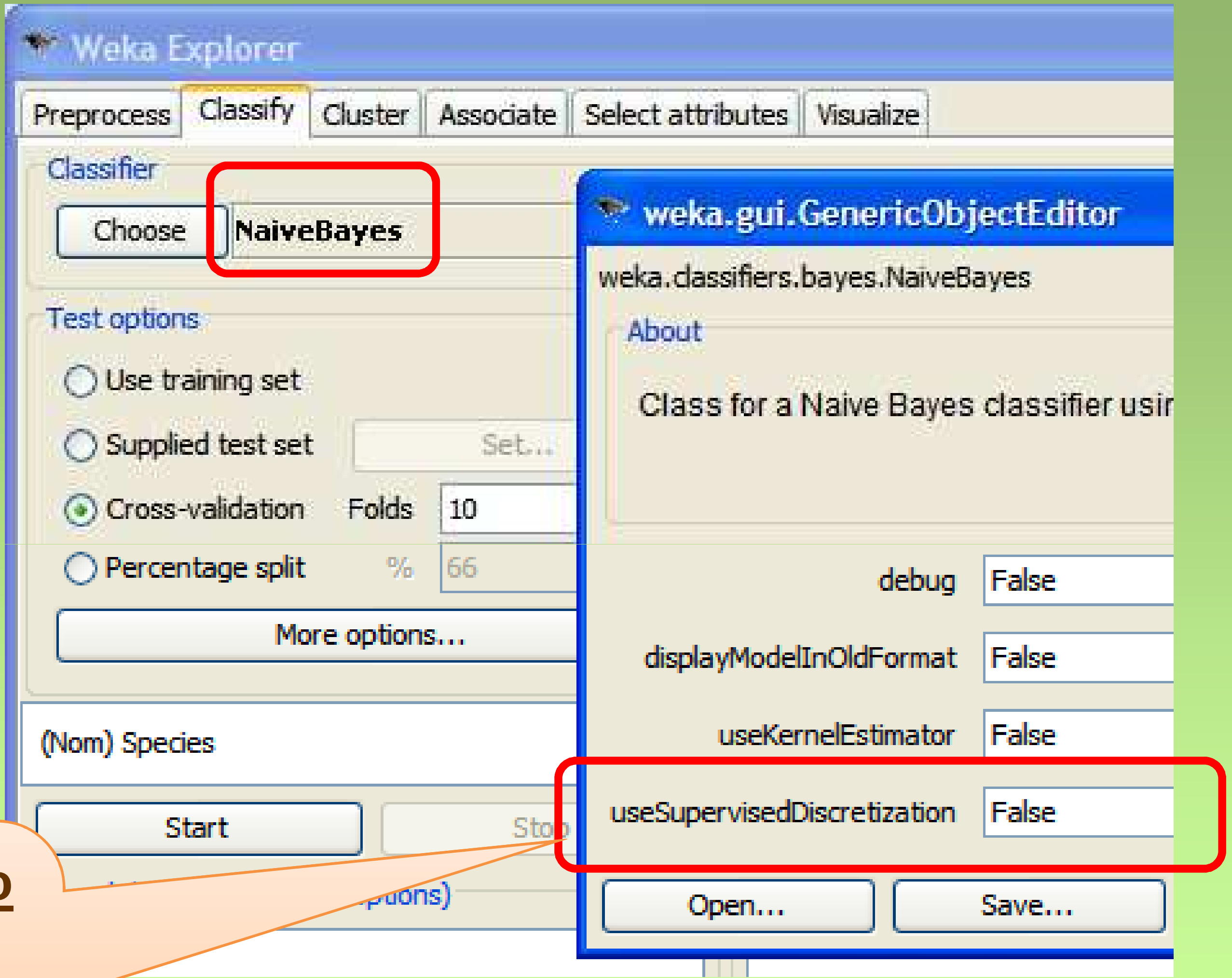
Clasificación: Redes Bayesianas

TRABAJO

César Hervás-Martínez
Grupo de Investigación AYRNA

Departamento de Informática y Análisis Numérico
Universidad de Córdoba
Campus de Rabanales. Edificio Einstein.
Email: chervas@uco.es

2017-2018



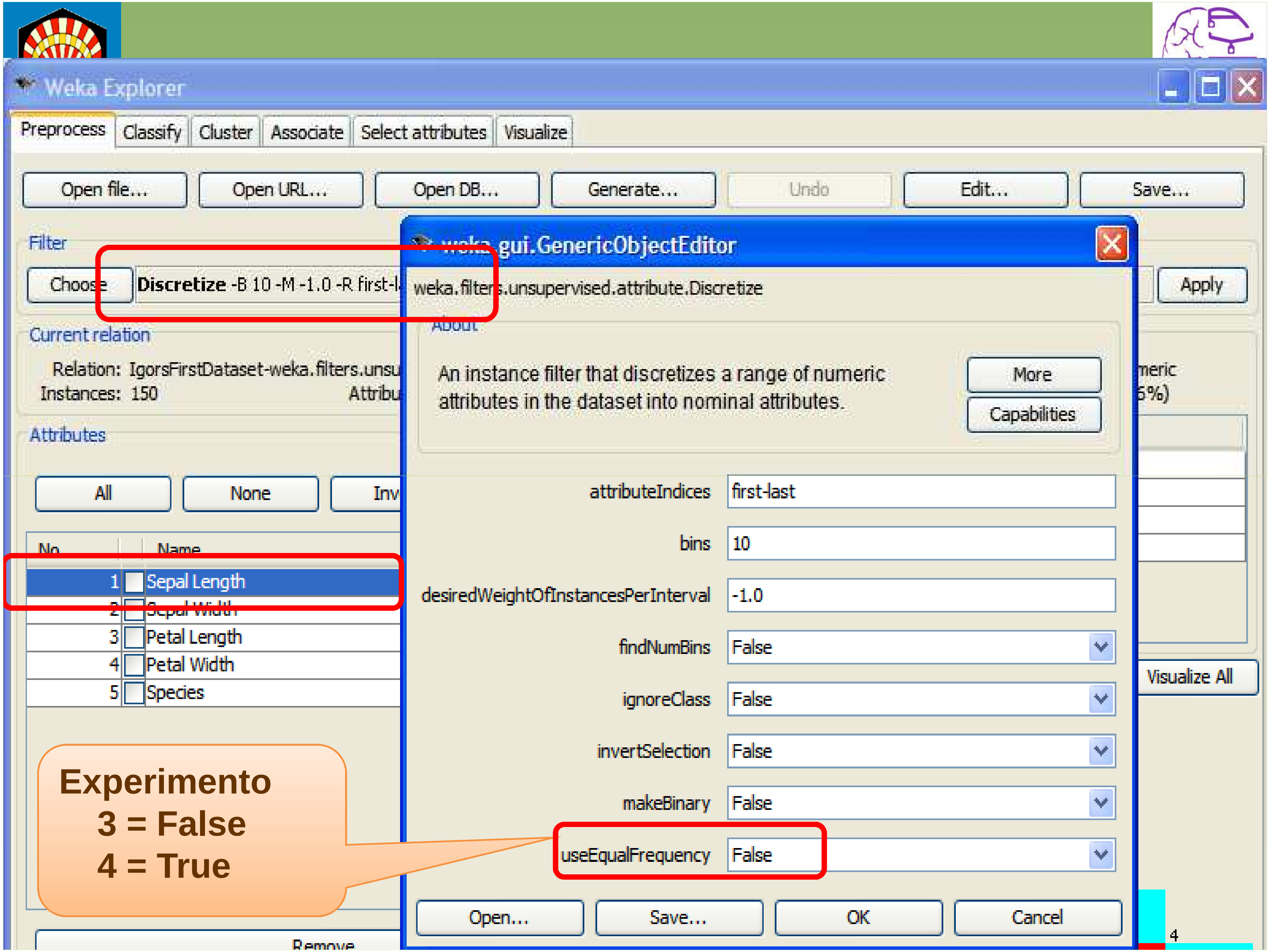
Experimento

1 = False

2 = True

3 = False

4 = False


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
Discretize -B 10 -M -1.0 -R first-l
Apply
Current relation
Relation: IgorsFirstDataset-weka.filters.unsu
Instances: 150
Attributes
All
None
No.
Name
1 Sepal Length
2 Sepal Width
3 Petal Length
4 Petal Width
5 Species
Remove
Visualize All
weka.gui.GenericObjectEditor
weka.filters.unsupervised.attribute.Discretize
About
An instance filter that discretizes a range of numeric attributes in the dataset into nominal attributes.
More
Capabilities
attributeIndices first-last
bins 10
desiredWeightOfInstancesPerInterval -1.0
findNumBins False
ignoreClass False
invertSelection False
makeBinary False
useEqualFrequency False
Open...
Save...
OK
Cancel
Experimento
3 = False
4 = True

Debe de ser un 8

Número total de casos mal clasificados para cada experimento

Experiment	1	2	3	4
Total	9	6	6	6
Instance	45	42	45	45
	72	45	72	72
	74	72	107	107
	100	107	116	116
	107	116	140	140
	116	140	148	148
	140			
	148			

Número de la instancia de cada clase mal clasificada

MODELOS COMPUTACIONALES: CUARTO CURSO DEL GRADO DE ING. INFORMÁTICA EN COMPUTACION

Clasificación: Redes Bayesianas

César Hervás-Martínez
Grupo de Investigación AYRNA

Departamento de Informática y Análisis Numérico
Universidad de Córdoba
Campus de Rabanales. Edificio Einstein.
Email: chervas@uco.es

2018-2019