



Machine Learning

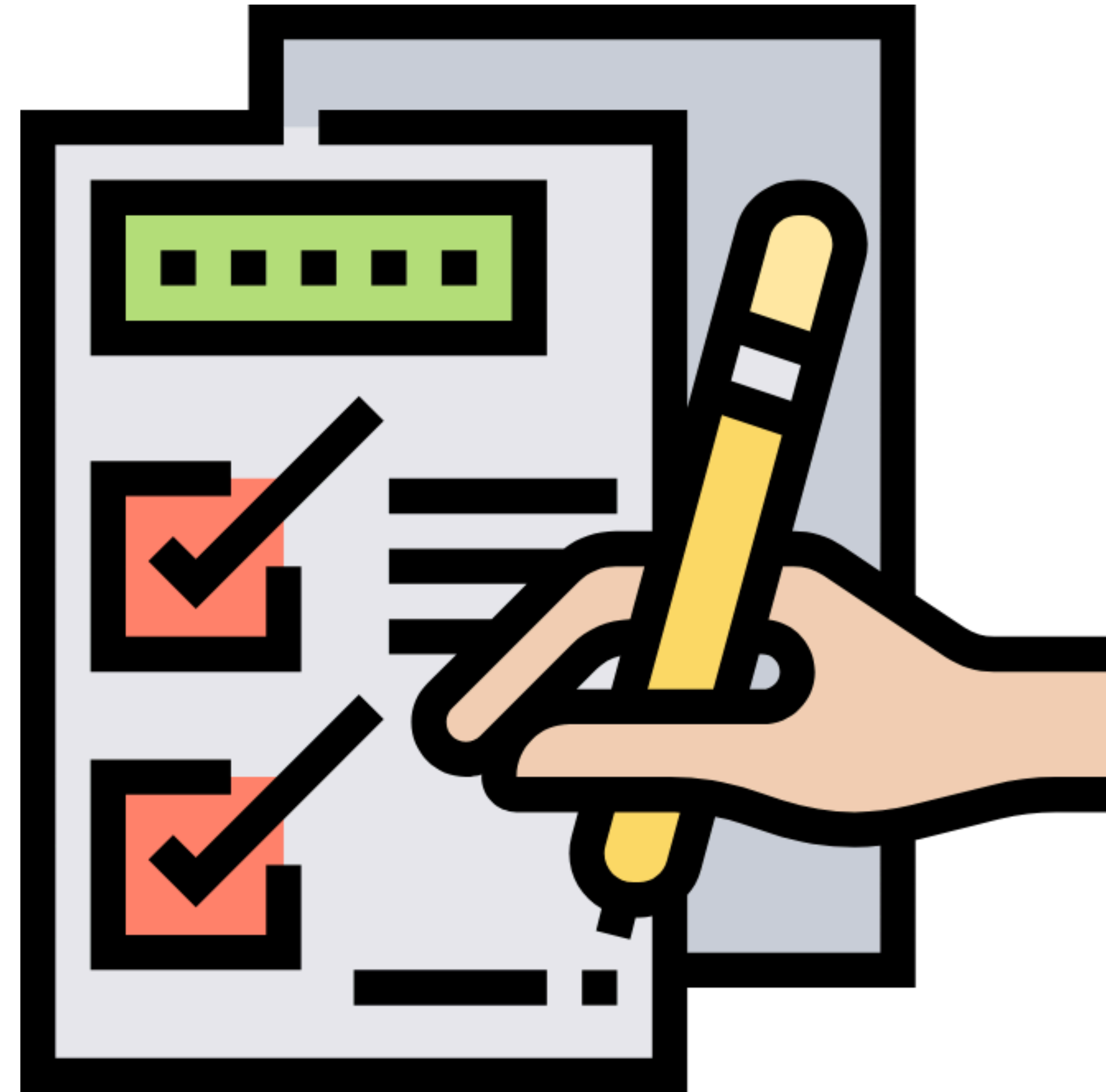
Unidad # 2 - Aprendizaje Supervisado
CC57 – 2019-1

Profesor
Andrés Melgar



Competencias a adquirir en la sesión

- Al finalizar la sesión el alumno comprenderá el funcionamiento del **aprendizaje inductivo**.
- Al finalizar la sesión el alumno implementará **modelos algoritmos de clasificación** usando conjuntos de datos.
- Al finalizar la sesión el alumno **entenderá** el funcionamiento del algoritmo **Naïve Bayes**.
- Al finalizar la sesión el alumno **aplicará** el algoritmo **Naïve Bayes** para obtener modelos algorítmicos.





Revisión de la sesión anterior

- ¿Por qué se dice que el aprendizaje de máquina es **aprendizaje inductivo**?
- ¿En qué se **fundamenta** el algoritmo **OneR**?
- ¿Para qué situaciones se suele **utilizar** el algoritmo **OneR**?



Naïve Bayes

Texto guía

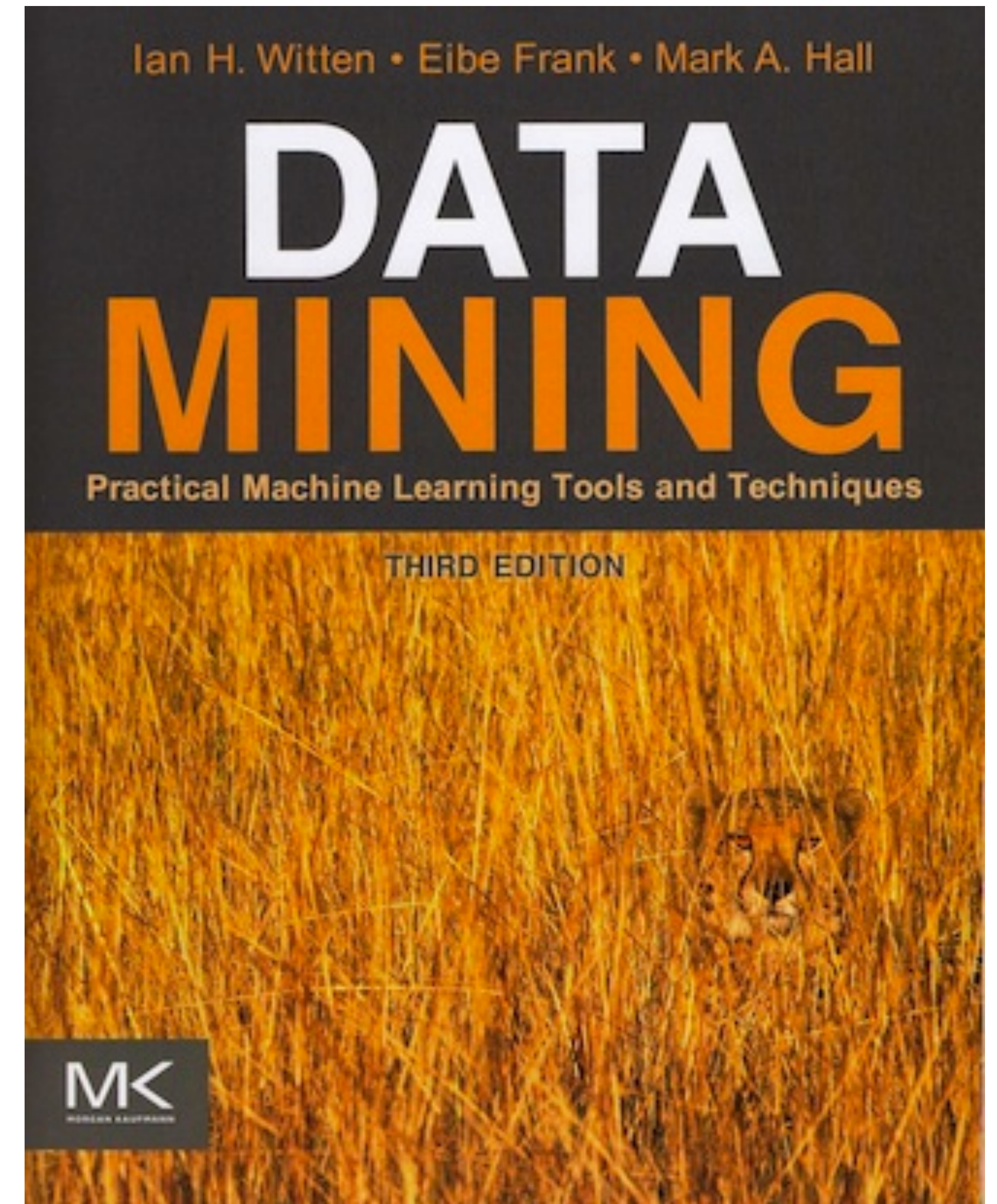
Witten, Ian H., Frank, Eibe, and Hall, Mark A.. 2011. *Data Mining : Practical Machine Learning Tools and Techniques with Java Implementations*. San Francisco: Elsevier Science & Technology.

CHAPTER

Algorithms: The Basic
Methods

4

4.2 STATISTICAL MODELING





Naïve Bayes

Lecturas recomendadas

Estimating Continuous Distributions in Bayesian Classifiers

George H. John
Computer Science Dept.
Stanford University
Stanford, CA 94305

gjohn@CS.Stanford.EDU
<http://robotics.stanford.edu/~gjohn/>

Pat Langley
Robotics Laboratory
Stanford University
Stanford, CA 94305

langley@CS.Stanford.EDU
<http://robotics.stanford.edu/~langley/>

Abstract

When modeling a probability distribution with a Bayesian network, we are faced with the problem of how to handle continuous variables. Most previous work has either solved the problem by discretizing, or assumed that the data are generated by a single Gaussian. In this paper we abandon the normality as-

Bayesian classifier. Despite the simplifying assumptions that underlie the naive Bayesian classifier, experiments on real-world data have repeatedly shown it to be competitive with much more sophisticated induction algorithms. For example, Clark & Niblett (1989) report naive Bayes producing accuracies comparable to those for rule-induction methods in medical domains, and Langley, Iba & Thompson (1992) found that it outperformed an algorithm for decision-tree induction

John, G. H., & Langley, P. (1995, August). *Estimating continuous distributions in Bayesian classifiers*. In Proceedings of the Eleventh conference on Uncertainty in artificial intelligence (pp. 338-345). Morgan Kaufmann Publishers Inc



Modelos Estadísticos

- Una técnica para diseñar un clasificador es utilizar **todos los atributos** y permitirles que hagan contribuciones a la decisión considerando que:
 - Son **igualmente importantes**.
 - Son **independientes** entre sí en relación a la clase predictora.
- Esto por supuesto no es realista. Lo que hacen interesantes a los conjuntos de datos reales es que **todos sus atributos:**
 - No son **igualmente importantes** y
 - No son **independientes**.



Modelos Estadísticos

outlook	temperature	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	mild	normal	FALSE	yes
sunny	mild	normal	TRUE	yes
overcast	mild	high	TRUE	yes
overcast	hot	normal	FALSE	yes
rainy	mild	high	TRUE	no



Modelos Estadísticos

outlook	temperature	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	mild	normal	FALSE	yes
sunny	mild	normal	TRUE	yes
overcast	mild	high	TRUE	yes
overcast	hot	normal	FALSE	yes
rainy	mild	high	TRUE	no



Modelos Estadísticos

outlook	temperature	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	mild	normal	FALSE	yes
sunny	mild	normal	TRUE	yes
overcast	mild	high	TRUE	yes
overcast	hot	normal	FALSE	yes
rainy	mild	high	TRUE	no

Valor			Total
sunny			5
overcast			
rainy			



Modelos Estadísticos

outlook	temperature	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	mild	normal	FALSE	yes
sunny	mild	normal	TRUE	yes
overcast	mild	high	TRUE	yes
overcast	hot	normal	FALSE	yes
rainy	mild	high	TRUE	no

Valor			Total
sunny			5
overcast			4
rainy			



Modelos Estadísticos

outlook	temperature	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	mild	normal	FALSE	yes
sunny	mild	normal	TRUE	yes
overcast	mild	high	TRUE	yes
overcast	hot	normal	FALSE	yes
rainy	mild	high	TRUE	no

Valor			Total
sunny			5
overcast			4
rainy			5
			14



Modelos Estadísticos

outlook	temperature	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	mild	normal	FALSE	yes
sunny	mild	normal	TRUE	yes
overcast	mild	high	TRUE	yes
overcast	hot	normal	FALSE	yes
rainy	mild	high	TRUE	no

Valor	yes	no	Total
sunny		3	5
overcast			4
rainy			5
			14



Modelos Estadísticos

outlook	temperature	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	mild	normal	FALSE	yes
sunny	mild	normal	TRUE	yes
overcast	mild	high	TRUE	yes
overcast	hot	normal	FALSE	yes
rainy	mild	high	TRUE	no

Valor	yes	no	Total
sunny	2	3	5
overcast			4
rainy			5
			14



Modelos Estadísticos

outlook	temperature	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	mild	normal	FALSE	yes
sunny	mild	normal	TRUE	yes
overcast	mild	high	TRUE	yes
overcast	hot	normal	FALSE	yes
rainy	mild	high	TRUE	no

Valor	yes	no	Total
sunny	2	3	5
overcast	4		4
rainy			5
			14



Modelos Estadísticos

outlook	temperature	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	mild	normal	FALSE	yes
sunny	mild	normal	TRUE	yes
overcast	mild	high	TRUE	yes
overcast	hot	normal	FALSE	yes
rainy	mild	high	TRUE	no

Valor	yes	no	Total
sunny	2	3	5
overcast	4	0	4
rainy			5
			14



Modelos Estadísticos

outlook	temperature	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	mild	normal	FALSE	yes
sunny	mild	normal	TRUE	yes
overcast	mild	high	TRUE	yes
overcast	hot	normal	FALSE	yes
rainy	mild	high	TRUE	no

Valor	yes	no	Total
sunny	2	3	5
overcast	4	0	4
rainy	3		5
			14



Modelos Estadísticos

outlook	temperature	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	mild	normal	FALSE	yes
sunny	mild	normal	TRUE	yes
overcast	mild	high	TRUE	yes
overcast	hot	normal	FALSE	yes
rainy	mild	high	TRUE	no

Valor	yes	no	Total
sunny	2	3	5
overcast	4	0	4
rainy	3	2	5
	9	5	14



Modelos Estadísticos

outlook	temperature	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	mild	normal	FALSE	yes
sunny	mild	normal	TRUE	yes
overcast	mild	high	TRUE	yes
overcast	hot	normal	FALSE	yes
rainy	mild	high	TRUE	no

Valor	yes	no	Total
hot	2	2	4
mild	4	2	6
cool	3	1	4
	9	5	14



Modelos Estadísticos

outlook	temperature	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	mild	normal	FALSE	yes
sunny	mild	normal	TRUE	yes
overcast	mild	high	TRUE	yes
overcast	hot	normal	FALSE	yes
rainy	mild	high	TRUE	no

Valor	yes	no	Total
high	3	4	7
normal	6	1	7
	9	5	14



Modelos Estadísticos

outlook	temperature	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	mild	normal	FALSE	yes
sunny	mild	normal	TRUE	yes
overcast	mild	high	TRUE	yes
overcast	hot	normal	FALSE	yes
rainy	mild	high	TRUE	no

Valor	yes	no	Total
FALSE	6	2	8
TRUE	3	3	6
	9	5	14



Modelos Estadísticos

outlook	temperature	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	mild	normal	FALSE	yes
sunny	mild	normal	TRUE	yes
overcast	mild	high	TRUE	yes
overcast	hot	normal	FALSE	yes
rainy	mild	high	TRUE	no

yes	no	Total
9	5	14



Modelos Estadísticos

outlook

Valor	yes	no	Total
sunny	2	3	5
overcast	4	0	4
rainy	3	2	5
	9	5	14

temperature

Valor	yes	no	Total
hot	2	2	4
mild	4	2	6
cool	3	1	4
	9	5	14

play

yes	no	Total
9	5	14

humidity

Valor	yes	no	Total
high	3	4	7
normal	6	1	7
	9	5	14

windy

Valor	yes	no	Total
FALSE	6	2	8
TRUE	3	3	6
	9	5	14



Modelos Estadísticos

outlook				
Valor	yes	no	P(E yes)	P(E no)
sunny	2	3	2/9	3/5
overcast	4	0	4/9	0/5
rainy	3	2	3/9	2/5
	9	5		

temperature				
Valor	yes	no	P(E yes)	P(E no)
hot	2	2	2/9	2/5
mild	4	2	4/9	2/5
cool	3	1	3/9	1/5
	9	5		

play			
yes	no	P(yes)	P(no)
9	5	9/14	5/14

humidity				
Valor	yes	no	P(E yes)	P(E no)
high	3	4	3/9	4/5
normal	6	1	6/9	1/5
	9	5		

windy				
Valor	yes	no	P(E yes)	P(E no)
FALSE	6	2	6/9	2/5
TRUE	3	3	3/9	3/5
	9	5		



Modelos Estadísticos

outlook				
Valor	yes	no	P(E yes)	P(E no)
sunny	2	3	2/9	3/5
overcast	4	0	4/9	0/5
rainy	3	2	3/9	2/5
	9	5		

humidity				
Valor	yes	no	P(E yes)	P(E no)
high	3	4	3/9	4/5
normal	6	1	6/9	1/5
	9	5		

temperature				
Valor	yes	no	P(E yes)	P(E no)
hot	2	2	2/9	2/5
mild	4	2	4/9	2/5
cool	3	1	3/9	1/5
	9	5		

windy				
Valor	yes	no	P(E yes)	P(E no)
FALSE	6	2	6/9	2/5
TRUE	3	3	3/9	3/5
	9	5		

play			
yes	no	P(yes)	P(no)
9	5	9/14	5/14

outlook	temperature	humidity	windy	play
sunny	cool	high	TRUE	?

$P(\textit{yes} \mid E) = ?$

$P(\textit{no} \mid E) = ?$



Modelos Estadísticos

outlook				
Valor	yes	no	P(E yes)	P(E no)
sunny	2	3	2/9	3/5
overcast	4	0	4/9	0/5
rainy	3	2	3/9	2/5
	9	5		

humidity				
Valor	yes	no	P(E yes)	P(E no)
high	3	4	3/9	4/5
normal	6	1	6/9	1/5
	9	5		

temperature				
Valor	yes	no	P(E yes)	P(E no)
hot	2	2	2/9	2/5
mild	4	2	4/9	2/5
cool	3	1	3/9	1/5
	9	5		

windy				
Valor	yes	no	P(E yes)	P(E no)
FALSE	6	2	6/9	2/5
TRUE	3	3	3/9	3/5
	9	5		

play			
yes	no	P(yes)	P(no)
9	5	9/14	5/14

outlook	temperature	humidity	windy	play
sunny	cool	high	TRUE	?

$$P(\text{yes} | E) = \frac{2}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{9}{14}$$

$$P(\text{yes} | E) = 0.005291005$$



Modelos Estadísticos

outlook				
Valor	yes	no	P(E yes)	P(E no)
sunny	2	3	2/9	3/5
overcast	4	0	4/9	0/5
rainy	3	2	3/9	2/5
	9	5		

humidity				
Valor	yes	no	P(E yes)	P(E no)
high	3	4	3/9	4/5
normal	6	1	6/9	1/5
	9	5		

temperature				
Valor	yes	no	P(E yes)	P(E no)
hot	2	2	2/9	2/5
mild	4	2	4/9	2/5
cool	3	1	3/9	1/5
	9	5		

windy				
Valor	yes	no	P(E yes)	P(E no)
FALSE	6	2	6/9	2/5
TRUE	3	3	3/9	3/5
	9	5		

play			
yes	no	P(yes)	P(no)
9	5	9/14	5/14

outlook	temperature	humidity	windy	play
sunny	cool	high	TRUE	?

$$P(no | E) = \frac{3}{5} \times \frac{1}{5} \times \frac{4}{5} \times \frac{3}{5} \times \frac{5}{14}$$

$$P(no | E) = 0.020571429$$



Modelos Estadísticos

outlook

Valor	yes	no	P(E yes)	P(E no)
sunny	2	3	2/9	3/5
overcast	4	0	4/9	0/5
rainy	3	2	3/9	2/5
	9	5		

temperature

Valor	yes	no	P(E yes)	P(E no)
hot	2	2	2/9	2/5
mild	4	2	4/9	2/5
cool	3	1	3/9	1/5
	9	5		

play

yes	no	P(yes)	P(no)
9	5	9/14	5/14

humidity

Valor	yes	no	P(E yes)	P(E no)
high	3	4	3/9	4/5
normal	6	1	6/9	1/5
	9	5		

windy

Valor	yes	no	P(E yes)	P(E no)
FALSE	6	2	6/9	2/5
TRUE	3	3	3/9	3/5
	9	5		

outlook	temperature	humidity	windy	play
sunny	cool	high	TRUE	?

$$P(\text{yes} | E) = 0.005291005$$

$$P(\text{no} | E) = 0.020571429$$

$$P(\text{yes} | E) = \frac{0.005291005}{0.005291005 + 0.020571429}$$

$$P(\text{yes} | E) = 20.45 \%$$



Modelos Estadísticos

outlook				
Valor	yes	no	P(E yes)	P(E no)
sunny	2	3	2/9	3/5
overcast	4	0	4/9	0/5
rainy	3	2	3/9	2/5
	9	5		

temperature				
Valor	yes	no	P(E yes)	P(E no)
hot	2	2	2/9	2/5
mild	4	2	4/9	2/5
cool	3	1	3/9	1/5
	9	5		

play			
yes	no	P(yes)	P(no)
9	5	9/14	5/14

humidity				
Valor	yes	no	P(E yes)	P(E no)
high	3	4	3/9	4/5
normal	6	1	6/9	1/5
	9	5		

windy				
Valor	yes	no	P(E yes)	P(E no)
FALSE	6	2	6/9	2/5
TRUE	3	3	3/9	3/5
	9	5		

outlook	temperature	humidity	windy	play
sunny	cool	high	TRUE	?

$$P(\text{yes} | E) = 0.005291005$$

$$P(\text{no} | E) = 0.020571429$$

$$P(\text{no} | E) = \frac{0.020571429}{0.005291005 + 0.020571429}$$

$$P(\text{no} | E) = 79.55 \%$$



Modelos Estadísticos

outlook

Valor	yes	no	P(E yes)	P(E no)
sunny	2	3	2/9	3/5
overcast	4	0	4/9	0/5
rainy	3	2	3/9	2/5
	9	5		

temperature

Valor	yes	no	P(E yes)	P(E no)
hot	2	2	2/9	2/5
mild	4	2	4/9	2/5
cool	3	1	3/9	1/5
	9	5		

play

yes	no	P(yes)	P(no)
9	5	9/14	5/14

humidity

Valor	yes	no	P(E yes)	P(E no)
high	3	4	3/9	4/5
normal	6	1	6/9	1/5
	9	5		

windy

Valor	yes	no	P(E yes)	P(E no)
FALSE	6	2	6/9	2/5
TRUE	3	3	3/9	3/5
	9	5		

outlook	temperature	humidity	windy	play
sunny	cool	high	TRUE	?

$$P(\text{yes} | E) = 0.005291005$$

$$P(\text{no} | E) = 0.020571429$$

$$P(\text{yes} | E) = 20.45 \%$$

$$P(\text{no} | E) = 79.55 \%$$



Regla de Bayes

- Este método sencillo e intuitivo se basa en la regla de la **probabilidad condicional de Bayes**.
- La **regla de Bayes** dice que si se tiene una hipótesis H y la evidencia E que lleva a esa hipótesis, entonces:

$$P(H|E) = \frac{P(E|H) \times P(H)}{P(E)}$$

- Donde:
 - $P(A)$ es la probabilidad del evento A .
 - $P(A|B)$ es la probabilidad del evento A condicionada al evento B .



Regla de Bayes

- Si la **hipótesis H** es que sí se juegue en el día;
- La **evidencia E** es la combinación de valores particulares para la nueva instancia (*outlook=sunny, temperature=cool, humidity=high, windy=true*);
- Vamos a llamar a esas piezas de evidencias de **E1, E2, E3 y E4**
- Vamos a asumir también que esas evidencias son **independientes**.
- Dados estos datos, la probabilidad combinada se puede obtener por medio de la siguiente fórmula:

$$P(\text{yes} | E) = \frac{P(E_1 | \text{yes}) \times P(E_2 | \text{yes}) \times P(E_3 | \text{yes}) \times P(E_4 | \text{yes}) \times P(\text{yes})}{P(E)}$$



Regla de Bayes

- Si la **hipótesis H** es que sí se juegue en el día;
- La **evidencia E** es la combinación de valores particulares para la nueva instancia (*outlook=sunny, temperature=cool, humidity=high, windy=true*);
- Vamos a llamar a esas piezas de evidencias de **E1, E2, E3 y E4**
- Vamos a asumir también que esas evidencias son **independientes**.
- Dados estos datos, la probabilidad combinada se puede obtener por medio de la siguiente fórmula:

$$P(\text{yes} | E) = \frac{P(\text{outlook} = \text{sunny} | \text{yes}) \times P(\text{temperature} = \text{cool} | \text{yes}) \times P(\text{humidity} = \text{high} | \text{yes}) \times P(\text{windy} = \text{true} | \text{yes}) \times P(\text{yes})}{P(E)}$$



Regla de Bayes

- No preocuparse por el **denominador** ($P(E)$) pues será eliminado al **normalizar** (hacer que las probabilidades sumen 1).
- La $P(\text{yes})$ al final de la fórmula, es la probabilidad sin ninguna evidencia, esta es llamada de **probabilidad anterior** de la hipótesis H .

$$P(\text{yes} | E) = \frac{P(E_1 | \text{yes}) \times P(E_2 | \text{yes}) \times P(E_3 | \text{yes}) \times P(E_4 | \text{yes}) \times P(\text{yes})}{P(E)}$$

$$P(\text{no} | E) = \frac{P(E_1 | \text{no}) \times P(E_2 | \text{no}) \times P(E_3 | \text{no}) \times P(E_4 | \text{no}) \times P(\text{no})}{P(E)}$$



Naïve Bayes

- Este método anteriormente visto es conocido como **Naïve Bayes**.
 - Está basado en las **reglas de Bayes**.
 - Asume ingenuamente **independencia**.
- Pero a pesar del nombre despectivo.
 - Funciona de manera **muy eficaz** en conjuntos de datos reales.
 - Funciona bien cuando se combina con algunos de los procedimientos de **selección de atributos**.



Naïve Bayes

Valores sin frecuencia

- Las cosas van mal al clasificador Naïve Bayes si el valor de un atributo particular, **no se encuentra** en el conjunto de entrenamiento.
- Analicemos en el conjunto de entrenamiento el valor del atributo **outlook = overcast**. Está siempre relacionado con la salida yes.
 - La probabilidad de *outlook = overcast* dado *No* sería 0 (**$P(\text{outlook}=\text{overcast}|\text{no})=0$**).
 - Y dado que las otras probabilidades se multiplican por esta probabilidad, la **probabilidad final sería 0**.
 - Las probabilidades que son cero tienen un **veto** sobre los otros. Esto no es una buena idea.
 - Pero el error se puede arreglar fácilmente por pequeños **ajustes** en el método de cálculo de las probabilidades de frecuencias.

outlook	play
sunny	no
sunny	no
overcast	yes
rainy	yes
rainy	yes
rainy	no
overcast	yes
sunny	no
sunny	yes
rainy	yes
sunny	yes
overcast	yes
overcast	yes
rainy	no



Naïve Bayes

Valores sin frecuencia

outlook				
Valor	yes	no	P(E yes)	P(E no)
sunny	2	3	2/9	3/5
overcast	4	0	4/9	0/5
rainy	3	2	3/9	2/5
	9	5		

temperature				
Valor	yes	no	P(E yes)	P(E no)
hot	2	2	2/9	2/5
mild	4	2	4/9	2/5
cool	3	1	3/9	1/5
	9	5		

play			
yes	no	P(yes)	P(no)
9	5	9/14	5/14

humidity				
Valor	yes	no	P(E yes)	P(E no)
high	3	4	3/9	4/5
normal	6	1	6/9	1/5
	9	5		

windy				
Valor	yes	no	P(E yes)	P(E no)
FALSE	6	2	6/9	2/5
TRUE	3	3	3/9	3/5
	9	5		



Naïve Bayes

Valores sin frecuencia

sin compensación

outlook				
Valor	yes	no	P(E yes)	P(E no)
sunny	2	3	2/9	3/5
overcast	4	0	4/9	0/5
rainy	3	2	3/9	2/5
	9	5		

con compensación

outlook				
Valor	yes	no	P(E yes)	P(E no)
sunny	3	4	3/12	4/8
overcast	5	1	5/12	1/8
rainy	4	3	4/12	3/8
	12	8		



Naïve Bayes

Valores sin frecuencia

sin compensación

outlook				
Valor	yes	no	P(E yes)	P(E no)
sunny	2	3	0.22	0.6
overcast	4	0	0.44	0.0
rainy	3	2	0.34	0.4
	9	5		

con compensación

outlook				
Valor	yes	no	P(E yes)	P(E no)
sunny	3	4	0.25	0.500
overcast	5	1	0.42	0.125
rainy	4	3	0.33	0.375
	12	8		



Naïve Bayes

Valores ausentes

- **En el entrenamiento:**

- Si falta un valor en una instancia de entrenamiento, simplemente **no se incluye** en los conteos de frecuencia.
- Las relaciones de probabilidad se basan en el **número de valores que realmente** ocurren en lugar de en el número total de instancias.

- **En la predicción:**

- Los valores ausente no son ningún problema. Simplemente **no se consideran en la productoria**.
- Supongamos que deseamos hacer la siguiente predicción:

outlook	temperature	humidity	windy	play
?	cool	high	TRUE	?



Naïve Bayes

Valores ausentes

outlook

Valor	yes	no	P(E yes)	P(E no)
sunny	2	3	2/9	3/5
overcast	4	0	4/9	0/5
rainy	3	2	3/9	2/5
	9	5		

temperature

Valor	yes	no	P(E yes)	P(E no)
hot	2	2	2/9	2/5
mild	4	2	4/9	2/5
cool	3	1	3/9	1/5
	9	5		

play

yes	no	P(yes)	P(no)
9	5	9/14	5/14

humidity

Valor	yes	no	P(E yes)	P(E no)
high	3	4	3/9	4/5
normal	6	1	6/9	1/5
	9	5		

windy

Valor	yes	no	P(E yes)	P(E no)
FALSE	6	2	6/9	2/5
TRUE	3	3	3/9	3/5
	9	5		

outlook	temperature	humidity	windy	play
?	cool	high	TRUE	?

$$P(\text{yes} | E) = ?$$

$$P(\text{no} | E) = ?$$



Naïve Bayes

Valores ausentes

outlook

Valor	yes	no	P(E yes)	P(E no)
sunny	2	3	2/9	3/5
overcast	4	0	4/9	0/5
rainy	3	2	3/9	2/5
	9	5		

temperature

Valor	yes	no	P(E yes)	P(E no)
hot	2	2	2/9	2/5
mild	4	2	4/9	2/5
cool	3	1	3/9	1/5
	9	5		

play

yes	no	P(yes)	P(no)
9	5	9/14	5/14

humidity

Valor	yes	no	P(E yes)	P(E no)
high	3	4	3/9	4/5
normal	6	1	6/9	1/5
	9	5		

windy

Valor	yes	no	P(E yes)	P(E no)
FALSE	6	2	6/9	2/5
TRUE	3	3	3/9	3/5
	9	5		

outlook	temperature	humidity	windy	play
?	cool	high	TRUE	?

$$P(\text{yes} | E) = \frac{3}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{9}{14}$$

$$P(\text{yes} | E) = 0.023809524$$



Naïve Bayes

Valores ausentes

outlook				
Valor	yes	no	P(E yes)	P(E no)
sunny	2	3	2/9	3/5
overcast	4	0	4/9	0/5
rainy	3	2	3/9	2/5
	9	5		

humidity				
Valor	yes	no	P(E yes)	P(E no)
high	3	4	3/9	4/5
normal	6	1	6/9	1/5
	9	5		

temperature				
Valor	yes	no	P(E yes)	P(E no)
hot	2	2	2/9	2/5
mild	4	2	4/9	2/5
cool	3	1	3/9	1/5
	9	5		

windy				
Valor	yes	no	P(E yes)	P(E no)
FALSE	6	2	6/9	2/5
TRUE	3	3	3/9	3/5
	9	5		

play			
yes	no	P(yes)	P(no)
9	5	9/14	5/14

outlook	temperature	humidity	windy	play
?	cool	high	TRUE	?

$$P(no | E) = \frac{1}{5} \times \frac{4}{5} \times \frac{3}{5} \times \frac{5}{14}$$

$$P(no | E) = 0.053571429$$



Naïve Bayes

Valores ausentes

outlook				
Valor	yes	no	P(E yes)	P(E no)
sunny	2	3	2/9	3/5
overcast	4	0	4/9	0/5
rainy	3	2	3/9	2/5
	9	5		

humidity				
Valor	yes	no	P(E yes)	P(E no)
high	3	4	3/9	4/5
normal	6	1	6/9	1/5
	9	5		

temperature				
Valor	yes	no	P(E yes)	P(E no)
hot	2	2	2/9	2/5
mild	4	2	4/9	2/5
cool	3	1	3/9	1/5
	9	5		

windy				
Valor	yes	no	P(E yes)	P(E no)
FALSE	6	2	6/9	2/5
TRUE	3	3	3/9	3/5
	9	5		

play			
yes	no	P(yes)	P(no)
9	5	9/14	5/14

outlook	temperature	humidity	windy	play
?	cool	high	TRUE	?

$$P(\text{yes} | E) = 0.023809524$$

$$P(\text{no} | E) = 0.034285714$$

$$P(\text{yes} | E) = \frac{0.023809524}{0.023809524 + 0.034285714}$$

$$P(\text{yes} | E) = 40.98 \%$$



Naïve Bayes

Valores ausentes

outlook				
Valor	yes	no	P(E yes)	P(E no)
sunny	2	3	2/9	3/5
overcast	4	0	4/9	0/5
rainy	3	2	3/9	2/5
	9	5		

humidity				
Valor	yes	no	P(E yes)	P(E no)
high	3	4	3/9	4/5
normal	6	1	6/9	1/5
	9	5		

temperature				
Valor	yes	no	P(E yes)	P(E no)
hot	2	2	2/9	2/5
mild	4	2	4/9	2/5
cool	3	1	3/9	1/5
	9	5		

windy				
Valor	yes	no	P(E yes)	P(E no)
FALSE	6	2	6/9	2/5
TRUE	3	3	3/9	3/5
	9	5		

play			
yes	no	P(yes)	P(no)
9	5	9/14	5/14

outlook	temperature	humidity	windy	play
?	cool	high	TRUE	?

$$P(\text{yes} | E) = 0.023809524$$

$$P(\text{no} | E) = 0.034285714$$

$$P(\text{no} | E) = \frac{0.034285714}{0.023809524 + 0.034285714}$$

$$P(\text{no} | E) = 59.02 \%$$



Naïve Bayes

Valores ausentes

outlook				
Valor	yes	no	P(E yes)	P(E no)
sunny	2	3	2/9	3/5
overcast	4	0	4/9	0/5
rainy	3	2	3/9	2/5
	9	5		

humidity				
Valor	yes	no	P(E yes)	P(E no)
high	3	4	3/9	4/5
normal	6	1	6/9	1/5
	9	5		

temperature				
Valor	yes	no	P(E yes)	P(E no)
hot	2	2	2/9	2/5
mild	4	2	4/9	2/5
cool	3	1	3/9	1/5
	9	5		

windy				
Valor	yes	no	P(E yes)	P(E no)
FALSE	6	2	6/9	2/5
TRUE	3	3	3/9	3/5
	9	5		

play			
yes	no	P(yes)	P(no)
9	5	9/14	5/14

outlook	temperature	humidity	windy	play
?	cool	high	TRUE	?

$P(yes | E) = 0.023809524$

$P(no | E) = 0.034285714$

$P(yes | E) = 40.98 \%$

$P(no | E) = 59.02 \%$



Naïve Bayes

Valores numéricos

- Los valores numéricos se manejan generalmente asumiendo que tienen una distribución de **probabilidad normal** o gaussiana.
- Para los atributos nominales, calculamos las **frecuencias** como en los ejemplos anteriores.
- Para los atributos numéricos, calculamos la **media** y la **desviación estándar** para cada clase.
- La función de **densidad de probabilidad** para una distribución normal con media μ y desviación estándar σ viene dada por la expresión:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x - \mu)^2}{2\sigma^2}}$$



Naïve Bayes

Actividad en Weka

- Ejecutar el algoritmo **Naïve Bayes** usando el conjunto de datos **weather.nominal.arff** y discuta el modelo algoritmo resultante.



Naïve Bayes

Actividad en Weka

- Ejecutar el algoritmo **Naïve Bayes** usando el conjunto de datos **weather.numeric.arff** y discuta el modelo algoritmo resultante.



Naïve Bayes

Actividad en RapidMiner

- Ejecutar el algoritmo **Naïve Bayes** usando el conjunto de datos **weather.nominal.csv** y discuta el modelo algoritmo resultante.

Naïve Bayes

Actividad en RapidMiner

Parameters ✕

Read CSV

Import Configuration Wizard...

Configure this operator by means of a Wizard.

csv file

column separators

☐ trim lines

[Hide advanced parameters](#)

[Change compatibility \(9.2.001\)](#)

Parameters ✕

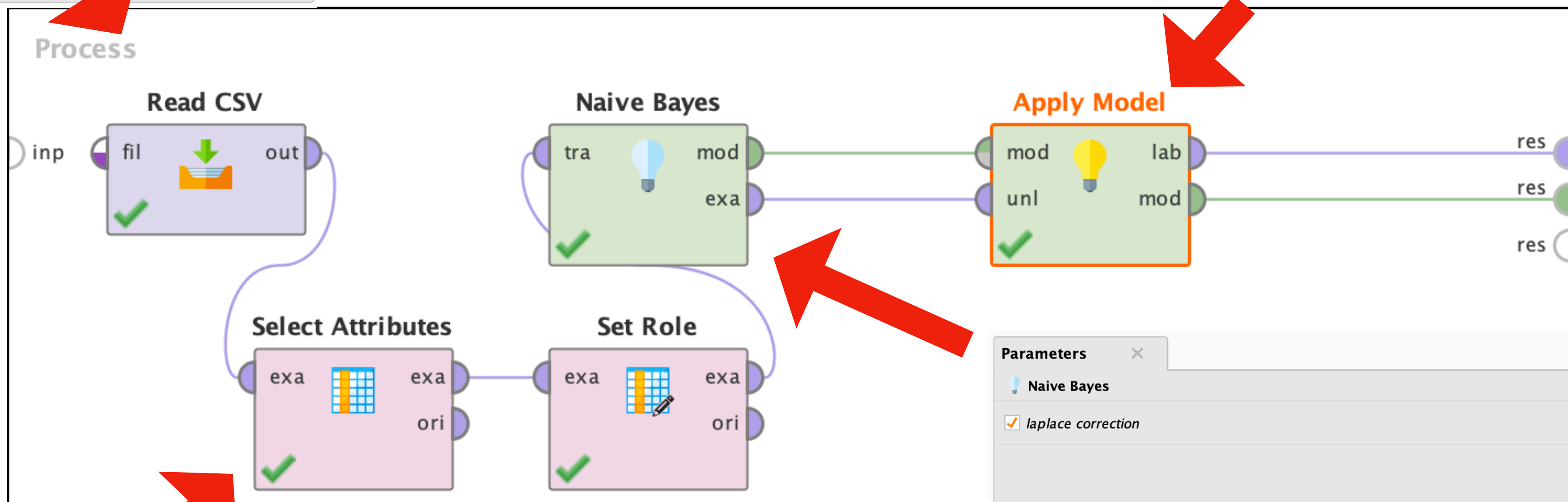
Apply Model

application parameters Edit List (0)...

☐ create view

[Hide advanced parameters](#)

[Change compatibility \(9.2.001\)](#)



Parameters ✕

Select Attributes

attribute filter type subset

attributes Select Attributes...

☐ invert selection

☒ include special attributes

Parameters ✕

Set Role

attribute name

target role

set additional roles Edit List (0)...

[Change compatibility \(9.2.001\)](#)

Parameters ✕

Naive Bayes

☒ laplace correction

[Hide advanced parameters](#)



Naïve Bayes

Actividad en RapidMiner

- Ejecutar el algoritmo **Naïve Bayes** usando el conjunto de datos **iris.data** y discutir el modelo algoritmo resultante.



Naïve Bayes

Actividad en Python

- Ejecutar el algoritmo **Naïve Bayes** usando el conjunto de datos **iris.data** y discutir el modelo algoritmo resultante.



Naïve Bayes

Actividad en Python

Cargamos el conjunto de datos

import pandas

archivo="iris.data"

columnas=['longitud-sépalo', 'ancho-sépalo', 'longitud-pétalo', 'ancho-pétalo', 'clase']

conjunto_de_datos = pandas.read_csv(archivo, names=columnas)

X = conjunto_de_datos.iloc[:,0:4].values

y = conjunto_de_datos.iloc[:,4].values

print(X)

print(y)

```
[ [5.1 3.5 1.4 0.2]
  [4.9 3.  1.4 0.2]
  [4.7 3.2 1.3 0.2]
  [4.6 3.1 1.5 0.2]
  [5.  3.6 1.4 0.2]
  [5.4 3.9 1.7 0.4]
  [4.6 3.4 1.4 0.3]
```

```
['Iris-setosa' 'Iris-setosa' 'Iris-setosa' 'Iris-setosa' 'Iris-setosa'
 'Iris-setosa' 'Iris-setosa' 'Iris-setosa' 'Iris-setosa' 'Iris-setosa'
 'Iris-setosa' 'Iris-setosa' 'Iris-setosa' 'Iris-setosa' 'Iris-setosa'
 'Iris-setosa' 'Iris-setosa' 'Iris-setosa' 'Iris-setosa' 'Iris-setosa']
```



Naïve Bayes

Actividad en Python

```
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import accuracy_score
from sklearn.metrics import classification_report
```

```
gnb = GaussianNB()
modelo = gnb.fit(X, y)
y_predecido = modelo.predict(X)
```

0.96

```
print(accuracy_score(y, y_predecido))
print(classification_report(y, y_predecido))
```

	precision	recall	f1-score	support
Iris-setosa	1.00	1.00	1.00	50
Iris-versicolor	0.94	0.94	0.94	50
Iris-virginica	0.94	0.94	0.94	50
micro avg	0.96	0.96	0.96	150
macro avg	0.96	0.96	0.96	150
weighted avg	0.96	0.96	0.96	150



Naïve Bayes

Actividad en Python

```
X_ejemplo=[(5.1,3.5,1.4,0.2)]  
y_predecido = modelo.predict(X_ejemplo)  
print(y_predecido)
```

```
[ 'Iris-versicolor' ]
```



Naïve Bayes

Asignación

- Ejecutar el algoritmo **Naïve Bayes** usando el conjunto de datos **weather.nominal.csv** y discuta el modelo algoritmo resultante.



Competencias a adquirir en la sesión

- Al finalizar la sesión el alumno comprenderá el funcionamiento del **aprendizaje inductivo**.
- Al finalizar la sesión el alumno implementará **modelos algoritmos de clasificación** usando conjuntos de datos.
- Al finalizar la sesión el alumno **entenderá** el funcionamiento del algoritmo **Naïve Bayes**.
- Al finalizar la sesión el alumno **aplicará** el algoritmo **Naïve Bayes** para obtener modelos algorítmicos.

