

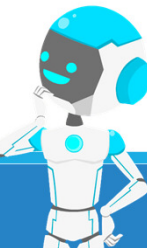
Probabilidade para Classificação

Como usar probabilidade para prever um fato?

Opção: Teorema de Bayes

Mas como criar um modelos a partir de probabilidade?

Olhando dados históricos e calculado a chances da classe serem influenciadas por determinados atributos!



Como?

Duas formas mais
conhecidas:

Naive Bayes

Redes Baysianas



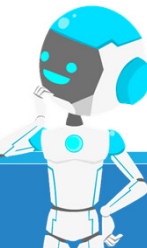
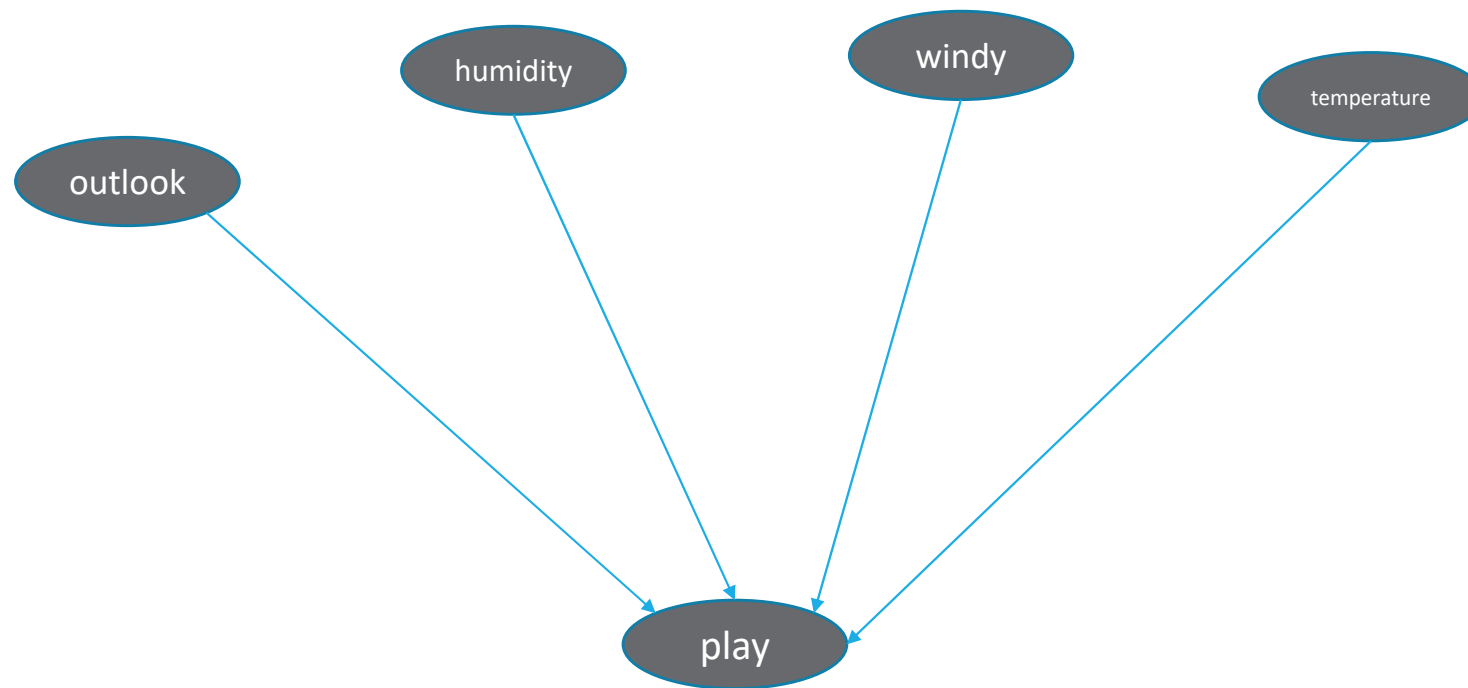
Naive Bayes

Outlook	Temperature	Humidity	windy
---------	-------------	----------	-------

Jogar:
YES/NO



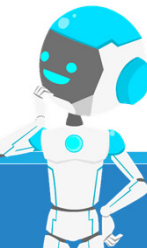
Probabilidad



Probabilidade Condicional da Classe

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
9/14	5/14
0,64	0,35



Probabilidade Condicional dos Atributos com a Classe

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	Yes	No
outlook	Sunny	2/9	3/5	0,22	0,6
	Overcast	4/9	0/5	0,44	0
	Rainy	3/9	2/5	0,33	0,4
Temperature	Hot	2/9	2/5	0,22	0,4
	Mil	4/9	2/5	0,44	0,4
	Cool	3/9	1/5	0,33	0,2
Humidity	High	3/9	4/5	0,33	0,8
	Normal	6/9	1/5	0,66	0,2
windy	TRUE	3/9	3/5	0,33	0,6
	FALSE	6/9	2/5	0,66	0,4
		9/14	5/14	0,64	0,35



Modelo

		Yes	No	Yes	No
outlook	Sunny	2/9	3/5	0,22	0,6
	Overcast	4/9	0/5	0,44	0
	Rainy	3/9	2/5	0,33	0,4
Temperature	Hot	2/9	2/5	0,22	0,4
	Mil	4/9	2/5	0,44	0,4
	Cool	3/9	1/5	0,33	0,2
Humidity	High	3/9	4/5	0,33	0,8
	Normal	6/9	1/5	0,66	0,2
windy	TRUE	3/9	3/5	0,33	0,6
	FALSE	6/9	2/5	0,66	0,4
		9/14	5/14	0,64	0,35

Cálculo da Probabilidade Posterior



Faz-se o cálculo da probabilidade posterior para cada classe



A classe que tiver o maior valor, “vence”

Cálculo da Probabilidade Posterior

outlook	temperature	humidity	windy
sunny	hot	high	FALSE

Probabilidade YES

$P(\text{yes}) * P(\text{sunny}|\text{yes}) * P(\text{hot}|\text{yes}) * P(\text{High}|\text{Yes}) * P(\text{FALSE}|\text{yes})$

$$0,64 * 0,22 * 0,22 * 0,33 * 0,66 = 0,006747$$

Probabilidade NO

$P(\text{no}) * P(\text{sunny}|\text{no}) * P(\text{hot}|\text{No}) * P(\text{High}|\text{no}) * P(\text{FALSE}|\text{no})$

$$0,35 * 0,6 * 0,4 * 0,8 * 0,4 = 0.03$$

		Yes	No	Yes	No
outlook	Sunny	2/9	3/5	0,22	0,6
	Overcast	4/9	0/5	0,44	0
	Rainy	3/9	2/5	0,33	0,4
Temperature	Hot	2/9	2/5	0,22	0,4
	Mil	4/9	2/5	0,44	0,4
	Cool	3/9	1/5	0,33	0,2
Humidity	High	3/9	4/5	0,33	0,8
	Normal	6/9	1/5	0,66	0,2
windy	TRUE	3/9	3/5	0,33	0,6
	FALSE	6/9	2/5	0,66	0,4
		9/14	5/14	0,64	0,35



Probabilidade Posterior

outlook	temperature	humidity	windy
Rainy	Cool	normal	TRUE

Probabilidade YES

$P(\text{yes}) * P(\text{rainy}|\text{yes}) * P(\text{cool}|\text{yes}) * P(\text{normal}|\text{Yes}) * P(\text{TRUE}|\text{yes})$

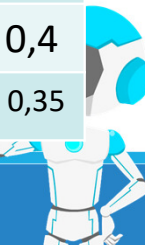
$$0,64 * 0,33 * 0,33 * 0,66 * 0,33 = 0,01518$$

Probabilidade NO

$P(\text{no}) * P(\text{rainy}|\text{no}) * P(\text{cool}|\text{No}) * P(\text{normal}|\text{no}) * P(\text{TRUE}|\text{no})$

$$0,35 * 0,4 * 0,2 * 0,2 * 0,6 = 0,00336$$

		Yes	No	Yes	No
outlook	Sunny	2/9	3/5	0,22	0,6
	Overcast	4/9	0/5	0,44	0
	Rainy	3/9	2/5	0,33	0,4
Temperature	Hot	2/9	2/5	0,22	0,4
	Mil	4/9	2/5	0,44	0,4
	Cool	3/9	1/5	0,33	0,2
Humidity	High	3/9	4/5	0,33	0,8
	Normal	6/9	1/5	0,66	0,2
windy	TRUE	3/9	3/5	0,33	0,6
	FALSE	6/9	2/5	0,66	0,4
		9/14	5/14	0,64	0,35



Valores
contínuos

“Discretização” dos atributos.
Por exemplo, transformar
idades em “Criança”, “Adulto”,
“Idoso”

Converter na probabilidade
segundo a distribuição normal
(**Gaussian naive Bayes**)