

Ejemplo Naive Bayes

JAFH

Datos:

Tenemos un conjunto de observaciones sobre:

- Ambiente: lluvioso, nublado, soleado
- Temperatura: baja, media, alta
- Viento: si, no

Donde se decide si jugar o no sobre el

Atributo de decisión

- Jugar: si, no

Ambiente	Temperatura	Viento	Jugar
L	B	N	S
L	M	N	S
L	M	N	S
L	B	S	N
L	M	S	N
N	A	N	S
N	A	N	S
N	B	S	S
N	M	S	S
S	A	N	N
S	B	N	S
S	M	N	N
S	A	S	N
S	M	S	S

Formula Naive Bayes

$$P(c_j|x_1, \dots, x_n) = P(c_j) * \prod_{i=1}^n P(x_i|c_j)$$

x_i =[Ambiente, Temperatura, Viento]

C_j =[jugar, nojugar]=[JS,JN]

Ambiente	Temperatura	Viento	Jugar
L	B	N	S
L	M	N	S
L	M	N	S
L	B	S	N
L	M	S	N
N	A	N	S
N	A	N	S
N	B	S	S
N	M	S	S
S	A	N	N
S	B	N	S
S	M	N	N
S	A	S	N
S	M	S	S

Calcular $P(x_i | c_j)$ para $P(x_i = \text{Ambiente} | C_j = \text{JN})$

$$P(x_{\text{ambiente}} | c_{jn})$$

$P(\text{Ambiente} | \text{JN})$:
Renglones de JN=5

$$P(\text{Ambiente} = \text{S} | \text{Jugar} = \text{N}) = 3/5 = 0.6$$

$$P(\text{Ambiente} = \text{N} | \text{Jugar} = \text{N}) = 0/5 = 0$$

$$P(\text{Ambiente} = \text{L} | \text{Jugar} = \text{N}) = 2/5 = 0.4$$

Ambiente	Temperatura	Viento	Jugar
S	A	N	N
S	B	N	S
S	M	N	N
S	A	S	N
S	M	S	S
N	A	N	S
N	A	N	S
N	B	S	S
N	M	S	S
L	B	N	S
L	M	N	S
L	M	N	S
L	B	S	N
L	M	S	N

Calcular $P(x_j|c_j)$ para $P(x_i=\text{Ambiente} | c_j=\text{JS})$

$$P(x_{\text{ambiente}}|c_{js})$$

$P(\text{Ambiente} | \text{JS})$:
Renglones de JS=9

$$P(\text{Ambiente}=\text{S} | \text{Jugar}=\text{S})=2/9=0.22$$

$$P(\text{Ambiente}=\text{N} | \text{Jugar}=\text{S})=4/9=0.44$$

$$P(\text{Ambiente}=\text{L} | \text{Jugar}=\text{S})=3/9=0.33$$

Ambiente	Temperatura	Viento	Jugar
S	A	N	N
S	B	N	S
S	M	N	N
S	A	S	N
S	M	S	S
N	A	N	S
N	A	N	S
N	B	S	S
N	M	S	S
L	B	N	S
L	M	N	S
L	M	N	S
L	B	S	N
L	M	S	N

Si tabulamos estos datos

	Ambiente			
		S	N	L
Jugar	N	0.6	0	0.4
	S	0.222	0.444	0.333

$P(\text{Ambiente} \mid \text{JN})$:

Renglones de JN=5

$$P(\text{Ambiente}=\text{S} \mid \text{Jugar}=\text{N})=3/5=0.6$$

$$P(\text{Ambiente}=\text{N} \mid \text{Jugar}=\text{N})=0/5=0$$

$$P(\text{Ambiente}=\text{L} \mid \text{Jugar}=\text{N})=2/5=0.4$$

$P(\text{Ambiente} \mid \text{JS})$:

Renglones de JS=9

$$P(\text{Ambiente}=\text{S} \mid \text{Jugar}=\text{S})=2/9=0.22$$

$$P(\text{Ambiente}=\text{N} \mid \text{Jugar}=\text{S})=4/9=0.44$$

$$P(\text{Ambiente}=\text{L} \mid \text{Jugar}=\text{S})=3/9=0.33$$

Calcular $P(x_j|c_j)$ para $P(x_i=\text{viento} | c_j=\text{JN})$

$$P(x_{\text{viento}}|c_{jn})$$

$P(\text{viento} | \text{JN})$:

Renglones de JN=5

$$P(\text{viento}=\text{N} | \text{Jugar}=\text{N})=2/5=0.4$$

$$P(\text{viento}=\text{S} | \text{Jugar}=\text{N})=3/5=0.6$$

Ambiente ▾	Temperatura ▾	Viento ▾↑	Jugar ▾
S	A	N	N
S	B	N	S
S	M	N	N
N	A	N	S
N	A	N	S
L	B	N	S
L	M	N	S
L	M	N	S
S	A	S	N
S	M	S	S
N	B	S	S
N	M	S	S
L	B	S	N
L	M	S	N

Calcular $P(x_j|c_j)$ para $P(x_i=\text{viento} | c_j=\text{JS})$

$$P(x_{\text{viento}}|c_{js})$$

$P(\text{viento} | \text{JS})$:
Renglones de JS=9

$$P(\text{viento}=\text{N} | \text{Jugar}=\text{S})=6/9=0.667$$

$$P(\text{viento}=\text{S} | \text{Jugar}=\text{S})=3/9=0.333$$

Ambiente ▼	Temperatura ▼	Viento ▼↑	Jugar ▼
S	A	N	N
S	B	N	S
S	M	N	N
N	A	N	S
N	A	N	S
L	B	N	S
L	M	N	S
L	M	N	S
S	A	S	N
S	M	S	S
N	B	S	S
N	M	S	S
L	B	S	N
L	M	S	N

Tabulando viento

		Viento	
		S	N
Jugar	N	0.6	0.4
	S	0.333	0.667

$P(\text{viento} \mid \text{JN})$:
Renglones de JN=5

$P(\text{viento}=\text{N} \mid \text{Jugar}=\text{N})=2/5=0.4$
 $P(\text{viento}=\text{S} \mid \text{Jugar}=\text{N})=3/5=0.6$

$P(\text{viento} \mid \text{JS})$:
Renglones de JS=9

$P(\text{viento}=\text{N} \mid \text{Jugar}=\text{S})=6/9=0.667$
 $P(\text{viento}=\text{S} \mid \text{Jugar}=\text{S})=3/9=0.333$

De la misma forma calculamos los demás datos...

Datos tabulados

Ambiente		S	N	L
Jugar	N	0.6	0	0.4
	S	0.222	0.444	0.333
Temperatura		A	M	B
Jugar	N	0.4	0.4	0.2
	S	0.222	0.444	0.333
Viento		S	N	
Jugar	N	0.6	0.4	
	S	0.333	0.667	

Calculamos:

$$P(c_j|x_1, \dots, x_n) = P(c_j) * \prod_{i=1}^n P(x_i|c_j)$$

Clasificador bayesiano simple

			P(A J)	P(T J)	P(V J)	
		a priori	Ambiente	Temperatura	Viento	P(J=S)*P(A J)
P(J=S A=S,T=A,V=N) ~	0.64	0.22	0.22	0.67	0.021	
P(J=S A=S,T=M,V=N) ~	0.64	0.22	0.44	0.67	0.042	
P(J=S A=S,T=B,V=N) ~	0.64	0.22	0.33	0.67	0.032	
P(J=S A=S,T=A,V=S) ~	0.64	0.22	0.22	0.33	0.011	
P(J=S A=S,T=M,V=S) ~	0.64	0.22	0.44	0.33	0.021	
P(J=S A=S,T=B,V=S) ~	0.64	0.22	0.33	0.33	0.016	
P(J=S A=N,T=A,V=N) ~	0.64	0.44	0.22	0.67	0.042	
P(J=S A=N,T=M,V=N) ~	0.64	0.44	0.44	0.67	0.085	
P(J=S A=N,T=B,V=N) ~	0.64	0.44	0.33	0.67	0.063	
P(J=S A=N,T=A,V=S) ~	0.64	0.44	0.22	0.33	0.021	
P(J=S A=N,T=M,V=S) ~	0.64	0.44	0.44	0.33	0.042	
P(J=S A=N,T=B,V=S) ~	0.64	0.44	0.33	0.33	0.032	
P(J=S A=L,T=A,V=N) ~	0.64	0.33	0.22	0.67	0.032	
P(J=S A=L,T=M,V=N) ~	0.64	0.33	0.44	0.67	0.063	
P(J=S A=L,T=B,V=N) ~	0.64	0.33	0.33	0.67	0.048	
P(J=S A=L,T=A,V=S) ~	0.64	0.33	0.22	0.33	0.016	
P(J=S A=L,T=M,V=S) ~	0.64	0.33	0.44	0.33	0.032	
P(J=S A=L,T=B,V=S) ~	0.64	0.33	0.33	0.33	0.024	

Calculamos:

$$P(c_j|x_1, \dots, x_n) = P(c_j) * \prod_{i=1}^n P(x_i|c_j)$$

			P(A NJ)	P(T NJ)	P(V NJ)	P(J=N)*P(A N
P(J=N A=S,T=A,V=N) ~	0.36	0.60	0.40	0.40	0.40	0.034
P(J=N A=S,T=M,V=N) ~	0.36	0.60	0.40	0.40	0.40	0.034
P(J=N A=S,T=B,V=N) ~	0.36	0.60	0.20	0.40	0.40	0.017
P(J=N A=S,T=A,V=S) ~	0.36	0.60	0.40	0.40	0.60	0.051
P(J=N A=S,T=M,V=S) ~	0.36	0.60	0.40	0.40	0.60	0.051
P(J=N A=S,T=B,V=S) ~	0.36	0.60	0.20	0.40	0.60	0.026
P(J=N A=N,T=A,V=N) ~	0.36	0.00	0.40	0.40	0.40	0.000
P(J=N A=N,T=M,V=N) ~	0.36	0.00	0.40	0.40	0.40	0.000
P(J=N A=N,T=B,V=N) ~	0.36	0.00	0.20	0.40	0.40	0.000
P(J=N A=N,T=A,V=S) ~	0.36	0.00	0.40	0.40	0.60	0.000
P(J=N A=N,T=M,V=S) ~	0.36	0.00	0.40	0.40	0.60	0.000
P(J=N A=N,T=B,V=S) ~	0.36	0.00	0.20	0.40	0.60	0.000
P(J=N A=L,T=A,V=N) ~	0.36	0.40	0.40	0.40	0.40	0.023
P(J=N A=L,T=M,V=N) ~	0.36	0.40	0.40	0.40	0.40	0.023
P(J=N A=L,T=B,V=N) ~	0.36	0.40	0.20	0.40	0.40	0.011
P(J=N A=L,T=A,V=S) ~	0.36	0.40	0.40	0.40	0.60	0.034
P(J=N A=L,T=M,V=S) ~	0.36	0.40	0.40	0.40	0.60	0.034
P(J=N A=L,T=B,V=S) ~	0.36	0.40	0.20	0.40	0.60	0.017

Recordemos el algoritmo para calcular predicciones

Paso 1: Para cada clase c_j , calcular la probabilidad dados los atributos x_i :

$$P(c_j|x_1, \dots, x_n) = P(c_j) * \prod_{i=1}^n P(x_i|c_j)$$

Pass 2: seleccionar la clase con mayor probabilidad

Predicción: aplicando el algoritmo

Paso 1: Para cada clase

- $c=[js, jn]$,
- calcular la probabilidad dados los atributos
- $x_i=[\text{Ambiente}=\text{nublado}, \text{Temperatura}=\text{Media}, \text{Viento}=\text{No}]$
- $= P(C_i \mid A=N, T=M, V=N)$

Verificamos los cálculos para las clases JS y JN y tenemos los valores: 0.085 y 0.00

Paso 2: La clase con mayor probabilidad es J=S

$P(J=S \mid A=N, T=M, V=N) \sim$	0.64	0.44	0.44	0.67	0.085
$P(J=N \mid A=N, T=M, V=N) \sim$	0.36	0.00	0.40	0.40	0.000

Resultado

Dada la ocurrencia de los atributos x_i =[Ambiente=nublado, Temperatura= Media, Viento=No]

Nos da la predicción de:

Jugar igual a si.