For Non Numerical

Naïve Bays classifier

No.E - 08365

Likelyhood

$$P(c|x) = \frac{P(x|c)\,P(c)}{P(x)}$$

$P(c)$ ← class prior probability

$P(x)$ ← Predictor prior probability

$P(c|x)$ ↓ Posterior probability

The steps to compute posterior probability

1) Construct the frequency table
2) Transform the frequency table into likelyhood.
3) The class with highest posterior is the outcome.

$\prod_{n}$

Step 1 — Construction of the frequency table

Outlook or weather	Temp	Humidity	Windy	Class Play Golf
Rainy	Hot	high	False	No
Rainy	Hot	High	True	No
Sunny	Mild	high	False	Yes.
Sunny	Cool	Normal	False	Yes.
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes.
Sunny	Hot	High	True	No
Rainy	Cool	High	True	No
Rainy	Cool	Normal	False	Yes
Sunny	Hot	High	True	No

(a) First we have to calculate the prior from the Class

Prior of Play Golf Yes $= \frac{5}{11} = 0.454$

$\therefore P(Y) = 0.454$

Prior of Play Golf No $= \frac{6}{11} = 0.545$

$P(N) = 0.545$

(b)

Freaquency tables of the attributes

		Play Golf	
		Yes	No
Outlook	Sunny	$3+1/5 = 0.8$	$2+1/6 = 0.5$
	Rainy	$2+1/5 = 0.6$	$4+1/6 = 0.83$

In this we have to calculate probability of yes in sunny day

probability of No in sunny day

probability of yes in Rainy day

probability of No in Rainy day

For that we have to divide the Total ~~number~~ number of sunny day when ~~their is~~ Play Golf with the Total number of Yes

Similarly

		Play Golf	
		Yes	No
Temp^r	Hot	$0+1/5 = 0.2$	$2+1/6 = 0.83$
	mild	$2+1/5 = 0.6$	$1+1/6 = 0.33$
	cool	$3+1/5 = 0.8$	$1+1/6 = 0.33$

Humidity	Play Golf: Yes	Play Golf: No
High	$\frac{3+1}{5}=0.9$	$\frac{6+1}{6}=1.9$
~~Low~~ Normal	$\frac{4+1}{5}=1$	$\frac{0+1}{6}=0.14$

Windy	Play Golf: Yes	Play Golf: No
False	$\frac{5+1}{5}=1.2$	$\frac{2+1}{6}=0.5$
True	$\frac{0+1}{5}=0.2$	$\frac{4+1}{6}=0.83$

(c) Evaluate the posterior for a given problem

Posterior play Golf when

outlook	Temp	Humidity	windy
Sunny	~~High~~ Mild	Normal / ~~Low~~	False

Now the Likelyhood with this problem will be

Likely hood $Yes = P(\text{sunny}|Yes) \cdot P(\text{mild}|Yes) \cdot (\text{Normal}|Yes) \cdot (\text{False} \cdot Yes) \times \{P(Yes)\}$

$= 0.8 \times 0.6 \times 1 \times 1.2 \times 0.454$

$= 3.6 \times 0.454$

$= 1.6$

Likely hood $No = P(\text{sunny}|No) \times P(\text{mind}|No)(\text{Normal}|No) \cdot (\text{False}|No) \times \{P(No)\}$

$= 0.5 \times 3.3 \times 0.14 \times 0.5 \times 0.54$

$= 0.07124$

D) Evaluate the Normalized posterior

$$P(yes|x) = \frac{P(x|yes)\,P(yes)}{P(x)}$$

$$= \frac{1.6}{(1.6 + 0.071)}$$

$$= 0.958$$

$$P(No|x) = \frac{P(x|No)\cdot P(No)}{P(x)}$$

$$= \frac{0.071}{(1.6 + 0.071)}$$

$$= 0.042$$

Naive Bays classifier For Numerical problem

No.E - 08364

DEPARTMENT OF PHYSICS
Estd.:2011
Date:
RAJIV GANDHI UNIVERSITY

RAJIV GANDHI UNIVERSITY

SUPPLEMENTARY SHEET (4 PAGES)

TO BE FILLED IN BY THE CANDIDATE:- *Signature of Invigilator*

(Do not write Roll No. & Name)

1. Subject ..... 2. Paper .....

Suppose we have numerical values like this

Outlook	temp	humidity	wind	play
Sunny	85	85	False	No
Sunny	80	90	True	No
Overcast	83	86	False	Yes
Rainy	70	96	False	Yes
Rainy	68	80	False	Yes
Rainy	65	70	True	No
overcast	64	65	True	Yes
Sunny	72	95	False	No
Sunny	69	70	False	Yes
Rainy	75	80	False	Yes
Sunny	75	70	True	Yes
overcast	72	90	True	Yes
overcast	81	75	False	Yes
Rainy	71	91	True	No

Now calculating the Frequency

Prior

Play	
Yes	No
9	5

P(Y) = 9/14 P(N) = 5/14

Outlook

	Yes	No
Sunny	2/9	3/5
overcast	4/9	0/5
Rainy	3/9	2/5

Temperature

	Yes	No
~~Sunny~~	83	85
	70	80
	68	65
mean	73	74.6
Std dev	6.2	7.9

Humidity

	Yes	No
	86	85
	96	90
	80	70
mean	79.1	86.2
Std dev	10.2	9.7

Windy

	Yes	No
False	6/9	2/5
True	3/9	3/5

The Gaussian probability distribution OR Normal distribution is given by

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

where σ = is the Standard deviation

M = mean

σ^2 = is variance

This probability distribution function will be used for the calculation of the Likely hood

Let us now take an example

Outlook	Temper	Humidity	windy	play
Sunny	66	90	True	??

Likelyhood

$$Yes = \frac{2}{9} \times GPD(66) \times GPD(99) \times \frac{3}{9} \times \frac{9}{14}$$

$$= \frac{2}{9} \times \left\{ \frac{1}{6.2\sqrt{2\pi}} e^{-\frac{(66-73)^2}{2\times(6.2)^2}} \right\} \times \left\{ \frac{1}{10.2\sqrt{2\pi}} e^{-\frac{(90-79.1)^2}{2(10.2)^2}} \right\} \times \frac{3}{9} \times \frac{9}{14}$$

$$= 0.000036$$

~~P(yes) =~~

Likelyhood

$$No = \frac{3}{5} \times \left\{ \frac{1}{7.9\sqrt{2\pi}} e^{-\frac{(66-74.6)^2}{2(7.9)^2}} \right\} \times \left\{ \frac{1}{9.7\sqrt{2\pi}} e^{-\frac{(90-86.2)^2}{2(9.7)^2}} \right\} \times \frac{3}{5} \times \frac{5}{14}$$

$$= 0.000136$$

$$P(yes) = 0.000036 / (0.000036 + 0.000136) = 20.9\%$$

$$P(No) = 0.000136 / (0.000036 + 0.000136) = 79.1\%$$

Psudo code for determining the Naive Bays classification

1) Determine the Number of classes
2) Determine the priors of each classes
3) Determine the frequency of each of the attributes
4) Determine the likelyhood from the frequency
5) Determine the posterior probability

This code is for Non zero classes

Step 1

Determine the number of classes

No of class = 0

for i = 1 : length(class vector)
C

~~for j = 1 : length(c)~~

~~if~~

if Sum(c) ≠ 0 and c(i) ≠ 0

No of class = No of class + 1,

Z = C / c(i)

I = find C == 1

K(i) = c(i)

M(i) = I

L(i) = length(I) / length(c)

{Z == 1 = [0]} // assign zeros to each element where Z = 1

end

prior of K(i) = L(i)

Pseudo code to determine frequency of the attributes for a particular class

---

Step 1 Lets consider the attributes is a mxn matrix where each columns represent a particular attribute

Find the frequency of a particular attribute for a particular class

for I = 1 : No of column

For

Select column & M(i) MM = M(i)

find for j = 1 : No of class → class vector

z = find (C == Class(j)) → class element return

find

X = MM(z) // Find the element of MM which correspond to the class element (j)

A(j) = Mean(x)

B(j) = stdv(x)

Frequency of Attributes