NAIVE BAYESIAN CLASSIFICATION

Pradnya Deorukhkar
MSc(Computer Science)
Sem II
pradnya.15d@gmail.com

Mrs. Dipali Meher MCS,
MPhil, NET
mailtomeher@gmail.com

Agenda

- Introduction
- Solved Example
- Advantage
- Disadvantages
- References

Introduction

Naive Bayesian classification algorithm is very simple which assumes that the classification attributes are independent and they do not have any correlation between them. Let D be a training set of tuples and their associated class labels, and each tuple is represented by an n-D attribute vector X= (x1, x2, …, xn). Suppose there are m classes C1, C2, …, Cm. Classification is to derive the maximum P(Ci|X).

This can be derived from Bayes' theorem

$$P(Ci/X) = \frac{P(X/Ci) * P(Ci)}{P(X)}$$

EXAMPLE 1

AGE	INCOME	STUDENT	CREDIT_RATING	BUYS_COMP
<=30	High	No	Fair	No
<=30	High	No	Excellent	No
31…40	High	No	Fair	Yes
>40	Medium	No	Fair	Yes
>40	Low	Yes	Fair	Yes
>40	Low	Yes	Excellent	No
31…40	Low	Yes	Excellent	Yes
<=30	Medium	No	Fair	No
<=30	Low	Yes	Fair	Yes
>40	Medium	Yes	Fair	Yes
<=30	Medium	Yes	Excellent	Yes
31…40	Medium	No	Excellent	Yes
31…40	High	Yes	Fair	Yes
>40	Medium	No	Excellent	No

C1 : buys_computer = “yes”

C2 : buys_computer = “no”

Data sample X = (age <=30, Income = medium, Student = yes Credit_rating = Fair)

A data sample is given to us here and we have to find whether the person buys a computer or no using Naive Bayesian classification. So first we calculate the probability of buys_computer. So the numbers of yes values divided upon total records that is 14 gives us the probability of buys_computer with yes and same with no.

P(Ci) : P(buys_computer = “yes”) = 9/14 = 0.643

P(buys_computer = “no”) = 5/14= 0.357

The total number of records in the table are 14.

Compute P(X|Ci) for each class

P(age = “<=30” | buys_computer = “yes”) = 2/9 = 0.222

P(age = “<= 30” | buys_computer = “no”) = 3/5 = 0.6

P(income = “medium” | buys_computer = “yes”) = 4/9 = 0.444 P(income = “medium” | buys_computer = “no”) = 2/5 = 0.4

P(student = “yes” | buys_computer = “yes) = 6/9 = 0.667

P(student = “no” | buys_computer = “no”) = 1/5 = 0.2

P(credit_rating = “fair” | buys_computer = “yes”) = 6/9 = 0.667 P(credit_rating = “fair” | buys_computer = “no”) = 2/5 = 0.4

Multiplying all the probabilities with yes values and no values from above separately.

P(X|Ci) :
P(X|buys_computer = “yes”)
= 0.222 x 0.444 x 0.667 x 0.667
= 0.044
P(X|buys_computer = “no”)
= 0.6 x 0.4 x 0.2 x 0.4
= 0.019
Now, multiply these probabilities with the above calculated P(Ci)

P(X|Ci)*P(Ci) :
P(X|buys_computer = “yes”) * P(buys_computer = “yes”) =
0.044 x 0.643=0.028
P(X|buys_computer = “no”) * P(buys_computer = “no”) =
0.019 x 0.357=0.007

The maximum value is 0.028.
Therefore, X belongs to class (“buys_computer = yes”)

Example 2

CHILLS	RUNNY NOSE	HEADACHE	FEVER	FLU?
Y	N	Mild	Y	N
Y	Y	No	N	Y
Y	N	Strong	Y	Y
N	Y	Mild	Y	Y
N	N	No	N	N
N	Y	Strong	Y	Y
N	Y	Strong	N	N
Y	Y	Mild	Y	Y

C1 :Flu = “yes”

C2 : Flu = “no”

Data sample X = (chills=Y, runnynose=N, headache=mild, fever=Y)

A data sample is given to us here and we have to find whether the person has flu or no using Naive Bayesian classification. So first we calculate the probability of flu. So the numbers of yes values divided upon total records that is 8 gives us the probability of flu with yes and same with no.

P(Ci) : P(Flu = “yes”) = 5/8 = 0.62

P(Flu = “no”) = 3/10= 0.37

The total number of records in the table are 8.

Compute P(X|Ci) for each class

P(chills="Y" | Flu = "yes") = 3/5 = 0.6

P(chills ="Y" |Flu = "no") = 1/3=0.33

P(runny nose="N" | Flu = "yes") = 1/5 = 0.2

P(runny nose="N" |Flu = "no") =2/3 = 0.66

P(headache="mild" | Flu = "yes") = 2/5 = 0.4

P(headache="mild" |Flu = "no") = 1/3 = 0.33

P(fever="Y" | Flu = "yes") = 4/5= 0.8

P(fever="Y" |Flu = "no") = 1/3=0.33

Multiplying all the probabilities with yes values and no values from above separately.

P(X|Ci) :

P(X|Flu= “yes”)

= 0.6 x 0.4 x 0.2 x 0.8

= 0.0384

P(X|Stolen = “no”)

= 0.33x 0.66 x 0.33 x 0.33

= 0.024

Now, multiply these probabilities with the above calculated P(Ci)

P(X|Ci)*P(Ci) :

P(X|Flu = “yes”) * P(Flu = “yes”) = 0.0384 x 0.62=0.023

P(X|Flu = “no”) * P(Flu = “no”) = 0.024 x0.37=0.0088

The maximum value is 0.023.

Therefore, X belongs to class (“Flu=yes”)

Example 3

NO	COLOR	TYPE	ORIGIN	STOLEN?
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

C1 : Stolen = “yes”

C2 : Stolen = “no”

Data sample X = (color=red, type=SUV, Origin=Domestic)

A data sample is given to us here and we have to find whether the person has stolen or no using Naive Bayesian classification. So first we calculate the probability of stolen. So the numbers of yes values divided upon total records that is 10 gives us the probability of stolen with yes and same with no.

P(Ci) : P(Stolen = “yes”) = 5/10 = 0.5

P(Stolen = “no”) = 5/10= 0.5

The total number of records in the table are 10.

Compute P(X|Ci) for each class

P(color="red" | Stolen = "yes") = 3/5 = 0.6
P(color="red" | Stolen = "no") = 2/5 = 0.4

P(type="SUV" | Stolen = "yes") = 1/5 = 0.2
P(type="SUV" | Stolen = "no") = 3/5 = 0.6

P(origin="Domestic" | Stolen = "yes") =2 /5 = 0.4
P(origin="Domestic"| Stolen = "no") = 3/5 = 0.6

Multiplying all the probabilities with yes values and no values from above separately.

P(X|Ci) :

P(X|Stolen = “yes”)

= 0.6 x 0.4 x 0.2

= 0.048

P(X|Stolen = “no”)

= 0.6 x 0.4 x 0.4

= 0.096

Now, multiply these probabilities with the above calculated P(Ci)

P(X|Ci)*P(Ci) :

P(X|Stolen = “yes”) * P(Stolen = “yes”) = 0.048 x 0.5=0.024

P(X|Stolen = “no”) * P(Stolen = “no”) = 0.096 x0.5=0.048

The maximum value is 0.048.

Therefore, X belongs to class (“Stolen= no”)

Example 4

OUTLOOK	TEMP	HUMIDITY	WINDY	PLAY GOLF?
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No

C1 : Play Golf = “yes”

C2 : Play Golf = “no”

Data sample X = (outlook=rainy, temp=hot, humidity=high, windy=false)

A data sample is given to us here and we have to find whether the person has play golf or no using Naive Bayesian classification. So first we calculate the probability of play golf. So the numbers of yes values divided upon total records that is 14 gives us the probability of playing golf with yes and same with no.

P(Ci) : P(Play golf = “yes”) = 9/14 = 0.643

P(Play golf = “no”) = 5/14= 0.357

The total number of records in the table are 14.

Compute P(X|Ci) for each class

P(outlook="rainy" | Play Golf = "yes") = 2/9 = 0.222
P(outlook="rainy" |Play Golf = "no") = 3/5 = 0.6

P(temp="hot"| Play Golf = "yes") = 2/9 = 0.222
P(temp="hot" |Play Golf = "no") = 2/5 = 0.4

P(humidity="high"| Play Golf = "yes") = 3/9 = 0.33
P(humidity="high"| Play Golf = "no") = 4/5 = 0.8

P(windy="false"| Play Golf = "yes") = 6/9 = 0.666
P(windy="false" |Play Golf = "no") = 2/5 = 0.4

Multiplying all the probabilities with yes values and no values from above separately.

P(X|Ci) :
P(X|play golf = “yes”)
= 0.222 x 0.222 x 0.333 x 0.666
= 0.0109
P(X|play golf = “no”)
= 0.6 x 0.4 x 0.8 x 0.4
= 0.0768
Now, multiply these probabilities with the above calculated P(Ci)

P(X|Ci)*P(Ci) :
P(X|play golf = “yes”) * P(play golf = “yes”) =
0.0109 x 0.643=0.007
P(X|play golf = “no”) * P(play golf= “no”) =
0.0768 x 0.357=0.0274

The maximum value is 0.0274.
Therefore, X belongs to class (“play golf = no”)

Example 5

PATIENT	DISEASE	SUGAR LEVEL	SURVIVAL CHANCES
Small	Serious	High	Yes
Medium	Normal	Low	Yes
Senior	Lifetime	Normal	Yes
Small	Lifetime	High	No
Small	Normal	High	Yes
Senior	Serious	Normal	No
Medium	Serious	Low	Yes
Senior	Normal	Low	No
Medium	Lifetime	Normal	Yes
Medium	Serious	High	No
Senior	Normal	Low	No

C1 : survival chances = “yes”

C2 : survival chances = “no”

Data sample X = (age=senior, disease=normal, sugar level=normal)

A data sample is given to us here and we have to find whether the person has survival chances or no using Naive Bayesian classification. So first we calculate the probability of survival chances. So the numbers of yes values divided upon total records that is 11 gives us the probability of survival chances with yes and same with no.

P(Ci) : P(survival chances = “yes”) = 6/11 = 0.545

P(survival chances = “no”) = 5/11= 0.455

The total number of records in the table are 11.

Compute P(X|Ci) for each class

P(age="senior" | survival chances = "yes") = 1/6 = 0.166

P(age="senior" | survival chances = "no") = 3/5 = 0.6

P(disease=" normal" | survival chances = "yes") = 2/6 = 0.33

P(disease=" normal" | survival chances = "no") = 2/5 = 0.4

P(sugar level="normal" | survival chances = "yes") = 2/6 = 0.33

P(sugar level="normal" | survival chances = "no") = 1/5 = 0.2

Multiplying all the probabilities with yes values and no values from above separately.

P(X|Ci) :
P(X|survival chances = “yes”)
= 0.166 x 0.33 x 0.33
= 0.018
P(X|survival chances = “no”)
= 0.6 x 0.4 x 0.2
= 0.048
Now, multiply these probabilities with the above calculated P(Ci)

P(X|Ci)*P(Ci) :
P(X|survival chances = “yes”) * P(survival chances = “yes”) =
0.018 x 0.545=0.0098
P(X|survival chances = “no”) * P(survival chances= “no”) =
0.048 x 0.455=0.02184

The maximum value is 0.02184.
Therefore, X belongs to class (“survival chances = no”)

Advantages

- Very simple, easy to implement and fast. Need less training data.
- It is highly scalable.
- It handles continuous and discrete data.
- Not sensitive to irrelevant features.
- It is useful in natural processing language.
- It can be easily updateable if we want to add training data.
- It is not a complicated algorithm.

Disadvantages

- The first disadvantage is that the Naive Bayes classifier makes an assumption on the shape of your data distribution, i.e. any two features are independent given the output class. Due to this, the result can be very bad.
- Another problem happens due to data scarcity. For any possible value of a feature, you need to estimate a likelihood value by an approach.
- This can result in probabilities going towards 0 or 1, which in turn leads to numerical instabilities and worse results will be seen.
- With Naive-Bayes, if you do have a class label and a certain attribute value together then the probability estimate will be zero.

References

1) www.ijsrd.com/articles/IJSRDV4I70116.pdf
2) *https://www.researchgate.net*
3) *GEEKS FOR GEEKS*
4) Indian Journal of Science and Technology, Vol 8(16), DOI: 10.17485/ijst/2015/v8i16/62055, July 2015.
5) www.ijraset.com

THANK YOU