Page No.
Date

Naive Bayes classification

Ques Consider the following dataset & predict the class of new instance X using Naive Bayes classification algorithm

ID	Refund	Marital status	Amount	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes.

Where X = (Refund = No, Marital status = Married, Income = 120K).

PREDICT ???

Understanding

Evade is class.

Class is either Yes or No.

Page No.
Date

So the question to understand is that data given says
Read the data as
If ID = 4 is married with 120K amount belongs to class No.

similarly
ID = 4 is single & belongs to class Yes.

WHAT TO PREDICT IS TO BE READ AS.

Instance X belongs to which class
if
REFUND = No

MARITAL STATUS = MARRIED

AMOUNT = 120K.

So the data becomes as follows.

Page No. _____
Date

ID	Refund	Marital Status	Amount	Evade
1	Yes	Single	12K	NO
2	NO	Married	100K	No
3	NO	Single	70K	No
4	Yes	Married	120K	No
5	NO	Divorced	95K	Yes
6	NO	Married	60K	No
7	Yes	Divorced	220K	No
8	NO	Single	85K	Yes
9	No	Married	75K	No
10	No.	Single	90K	Yes.
⟹ PREDICT	11	NO	MARRIED	120K	??

* Bayesian classification

Problem statement
- Given features $X_1, X_2, \cdots X_n$.
- Predict a label Y.

* Bayes's Classifier.

- A probabilistic frame work for solving classification problems.

- Conditional probability

$$P(C/A) = \frac{P(A,C)}{P(A)} \quad \text{---} 1$$

Page No.
Date

$$P(A|C) = \frac{P(A,C)}{P(C)} \quad \text{— 2}$$

Read above as
Probability of C given A. → 1
Probability of A given C → 2

* ~~Bayes theorem~~:

* Remember: Bayes classifier uses conditional Probability

Bayes Theorem:-

$$P(C|A) = \frac{P(A|C)\,P(C)}{P(A)}$$

Probability of C given A is calculated as
(Probability of A given C) * (Probability of C)
Probability of A.

Page No.
Date

In general, compute the posterior probability $P(C \mid A_1, A_2, \ldots A_n)$ for all values of C using the Bayes Theorem.

Choose the value of C that maximizes $P(C \mid A_1, A_2 \ldots \ldots A_n)$

* Class

$$P(C) = \frac{N_C}{N}$$

~~Eg~~.

Count the no. of NO's & Yes' from the data set

$n(No) = 7$, $n(Yes) = 3$.

No. of NO's No. of Yes.

Probability of NO's $= \frac{7}{10} = P(No)$

Probability of Yes' $= \frac{3}{10} = P(Yes)$

Attributes here are Marital status
Refund
~~Business~~

Attributes here are discrete by nature. Means they are either True / false or Yes/No.

Page No.
Date

so for discrete attributes

$$P(A_i \mid C_k) = \frac{|A_{ik}|}{N_{ic}}$$

where $|A_{ik}|$ is the number of instances having attributes A_i's and belongs to class C_k.

e.g.

P(status = Married / No)

Means probability of instance who are married and belongs to class No

so the probability will be.

$$P(\text{Married} \mid \text{No}) = \frac{4}{7}.$$

from the dataset

n(Married) = 4.
n(No) = 7.

similarly
P(Refund = Yes / Yes)

Page No.
Date

Means probability of instance refund with Yes belonging to class Yes.

$$P(\text{Refund} = \text{Yes} \mid \text{Yes}) = \frac{0}{3} = 0$$

** simply count the instance with the class stated **.

so the question was to predict the CLASS for

Refund = No
Marital status = Married
Amount = 120K.

for this Naive Bayes probability

$P(X \mid \text{class} = \text{No})$

$= P(\text{Refund} = \text{No} \mid \text{Class} = \text{No}) *$
$P(\text{Married} \mid \text{Class} = \text{No}) *$
$P(\text{Amount} = 120\text{K} \mid \text{Class} = \text{No})$

→ Means probability of invoiced instance X to belong class No.

→ Calculated as

Page No.
Date

Similarly

Probability of instance X belonging to class Yes. is given as.

$P(X \mid \text{class} = \text{Yes})$

$= P(\text{Refund} = \text{No} \mid \text{class} = \text{Yes}) *$
$P(\text{Married} \mid \text{class} = \text{Yes}) *$
$P(\text{Income} = 120\text{K} \mid \text{class} = \text{Yes})$

** Calculate for both class i.e. YES & NO

- Probability of Yes & No.

- If Whichever Probability with class YES or No comes greater the prediction falls for that class

CALCULATION.

$P(\text{Refund} = \text{NO} \mid \text{class} = \text{NO})$
counts Refund as No with class NO. from the given dataset

Page No.
Date

$n(\text{Refund} = \text{No}) = 4.$
$n(\text{Class} = \text{No}) = 7.$

$\Rightarrow P(\text{Refund} = \text{No} \mid \text{Class} = \text{No}) = \frac{4}{7}$

$P(\text{Married} \mid \text{Class} = \text{No})$

$n(\text{Married}) = 4.$
$n(\text{Class} = \text{No}) = 7.$

$\Rightarrow P(\text{Married} \mid \text{Class} = \text{No}) = \frac{4}{7}.$

$P(\text{Account Income} = 120\text{K} \mid \text{Class} = \text{No})$

* From dataset, Account is not discrete. i.e. the values are distributive so we have to go for probability distribution. For this we need to find Normal distribution.

NORMAL DISTRIBUTION.

$$P(A_i \mid C_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} e^{\left(\frac{A_i - \mu_{ij}}{2\sigma_{ij}^2}\right)}$$

Page No.
Date

Where

μ_{ij} = mean.

σ_{ij}^2 = variance.

e.g. How to calculate mean & variance

suppose given data set

$X = (5, 8, 7, 6, 9)$

$$\sigma^2 = \frac{\sum (X_i - \bar{X})^2}{n-1}$$

$\bar{X}$ = mean of samples.

n = no. of samples.

$$\therefore \bar{X} = \frac{5+8+7+6+9}{5}$$

$$\boxed{\bar{X} = 7}$$

$$\sigma^2 = \frac{(5-7)^2 + (8-7)^2 + (7-7)^2 + (6-7)^2 + (9-7)^2}{5-1}$$

$$\boxed{\sigma^2 = 2.5}$$

Page No.
Date

In our case

$$P(\text{Amount} = 120K \mid \text{Class} = No) = \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} e^{\left(\frac{A_i - \mu_{ij}}{2\sigma_{ij}^2}\right)}$$

$A_i = 120K$

μ_{ij} = Mean of all the amounts belonging to class No

$$= \frac{125 + 100 + 70 + 120 + 60 + 220 + 75}{7}$$

Total no. of class No = 7.

$$\mu_{ij} = \frac{770}{7} = 110K.$$

(Amount corresponding to class No is taken for mean)

$$\sigma_{ij}^2 = \frac{(125-110)^2 + (100-110)^2 + (70-110)^2 + (120-110)^2 + (60-220)^2 + (75-110)^2 + (60-110)^2 + (220-110)^2}{7-1}$$

$$= \frac{(15)^2 + (-10)^2 + (-40)^2 + (10)^2 + (-50)^2 + (110)^2 - (-35)^2}{6}$$

Page No.
Date

$$= \frac{17850}{6}$$

$$\sigma_{ij}^2 = 2975$$

$\therefore$ P(Amount $\leq$ 120K / Class = No)

$$= \frac{1}{\sqrt{2\pi(2975)}} e^{\left(\frac{120-110}{2(2975)}\right)}$$

$$= 0.0079.$$

$\Rightarrow$ P(X / Class = No)

$$= \frac{4}{7} \times \frac{4}{7} \times 0.0072$$

$$= 0.0023.$$

So the conditional probability for instance X for class No is 0.0023.

Now, calculate for class Yes too

Page No.
Date

$P(X \mid Class = Yes)$

$= P(Refund = No \mid Class = Yes) *$
$P(Married \mid Class = Yes) *$
$P(Amount = 125k \mid Class = Yes)$

$= P(Refund = No \mid class = Yes)$

$= \frac{3}{3.} =$ $n(Yes) = 3.$

$P(Married \mid Class = Yes)$

$= \frac{0}{3.} = 0.$

$P(Amount = 125K \mid Class = Yes)$
$Ai = 125k$

$\mu_{ij} = \frac{95 + 85 + 90}{3} = 90.$

$\sigma_{ij}^{2} = \frac{(95-90)^2 + (85-90)^2 + (90-90)^2}{3.}$

$= 25$

$\Rightarrow P(Amount = 125K \mid Class = Yes)$

$= \frac{1}{\sqrt{2\pi\sigma_{ij}^{2}}} e^{\left(\frac{Ai - \mu_{ij}}{2\sigma_{ij}}\right)}$

Page No.
Date

$$= \frac{1}{\sqrt{2\pi}\,25} e^{\left(\frac{125-90}{2\times 25}\right)}$$

$$= \cancel{12\times 10^{-9}}$$

$$= 1.2\times 10^{-9}$$

~~Thus, P(X | Class = Yes)~~

Thus $P(X \mid \text{Class} = \text{Yes})$

$$= \frac{3}{3} \times \frac{0}{3} \times 1.2\times 10^{-9}$$

$$= 0$$

$P(X \mid No) \cdot P(No)$

$$= 0.0023 \times \frac{7}{10}$$

$=$

$P(X \mid Yes) \cdot P(Yes)$

$$= 0 \times \frac{3}{10}$$

$$= 0$$

Page No. _____
Date

$\therefore$ $P(No/X) > P(Yes/X)$

$\Rightarrow$ The prediction for the instance X is No

i.e. Refund = No.
Marital status = Married
Amount = 120K.
Evade/Class = No.

Assignment
(Do submit when the college reopens)

Ques For the above dataset find the prediction for instance X.

X = (Refund = Yes; Status = Single; Amount = 100K)

Ques Give the Naive Bayes algorithm

Ques What are the pros & cons. of Naive Bayes algorithm.