BITS PILANI, DUBAI CAMPUS
DUBAI INTERNATIONAL ACADEMIC CITY, DUBAI

SECOND SEMESTER 2023 – 2024
COURSE: BITS F464 (Machine Learning)
COMPONENT: Practice Tutorial 3 **DATE:** 23rd October 2023

Q1: Consider a below dataset of 14 training examples of the target concept "Buys_computer", where each instance is described as attributes Age, Income, Student,Credit_rating.

Age	Income	Student	Credit_rating	Buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31 to 40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31 to 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 to 40	medium	no	excellent	yes
31 to 40	high	yes	fair	yes
>40	medium	no	excellent	no

Apply Navie Bayes Classifier on given training data to classify the novel instance X = (age <=30, Income = medium, Student = yes, Credit_rating = fair) in a correct class. Show the steps by step procedure of calculating all posterior and prior probabilities to classify the instance in a correct class (Training and Testing Phase).

Answer:

Training Phase:
P(Ci): P(buys_computer = "yes") = 9/14 = 0.643
P(buys_computer = "no") = 5/14= 0.357
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 = "yes" | 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
Testing Phase:

X = (age <= 30 , income = medium, student = yes, credit_rating = fair)
P(X|Ci) : P(X|buys_computer = "yes") = 0.222 × 0.444 × 0.667 × 0.667 = 0.044
P(X|buys_computer = "no") = 0.6 × 0.4 × 0.2 × 0.4 = 0.019
P(X|Ci)*P(Ci) : P(X|buys_computer = "yes") * P(buys_computer = "yes") = 0.028
P(X|buys_computer = "no") * P(buys_computer = "no") = 0.007
Therefore, X belongs to class ("buys_computer = yes")

Q2: Consider the following set of training examples,

Instance	**Classification**	**A1**	**A2**
1	+	T	T
2	+	T	T
3	-	T	F
4	+	F	F
5	-	F	T
6	-	F	T

What is the information gain of A2 relative to these training examples?

Answer:
Gain(S,A2) = 0

Q3: Given a following observed datapoints *(X: x_i)* with two classes A and B,

Class A	**Class B**
10	14
11	11
8	9
10	13

The prior probabilities of class A and B are 0.4 and 0.6 respectively. What will be the class of following unlabelled data points (x1 and x2) using Gaussian for classification? (Show all steps of calculations)

i. X1 = 5

ii. X2 = 12

Answer:

i. X1 belongs to class B

ii. X2 belongs to class B

Q4: Consider a following dataset which shows willingness of certain customer to eat a chocolate of certain manufactures with specific flavor,

Manufacturer	**Flavor**	**Willingness to eat**
Lindt	Strawberry	Yes
Lindt	Milk	Yes
Cadbury	Strawberry	Yes
Cadbury	Milk	No
Milka	Strawberry	No
Milka	Milk	No

i. In above dataset, which feature (Manufacturer or flavor) has the larger information gain with willingness to eat?

ii. Also draw the decision tree for predicting willingness to eat which maximizes the information gain?

Answer:

i. Manufacture have larger information gain than flavor.

ii. Decision tree