LAB Manual

PART A

(PART A : TO BE REFFERED BY STUDENTS)

**Experiment No.07**

**Aim:** Implementation of Naïve Bayes Classifier

**Prerequisites:**

C/C++/Java /Python Programming

**Learning Outcomes:**

Concepts of Bayesian theorem and Classification

**Theory:**

**Bayesian Theorem**

* Given training data**X***, posteriori probability of a hypothesis* H*,* P(H|**X**)*,* follows the Bayes theorem

* Informally, this can be written as

posteriori = likelihood x prior/evidence

* Predicts **X** belongs to C2 iff the probability P(Ci|**X**) is the highest among all the P(Ck|X) for all the *k* classes
* Practical difficulty: require initial knowledge of many probabilities, significant computational cost

**Bayesian Classifier**

* 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 posteriori, i.e., the maximal P(Ci|**X**)
* This can be derived from Bayes’ theorem



* Since P(X) is constant for all classes, only



needs to be maximized

**Example**

Class:

C1:buys\_computer = ‘yes’

C2:buys\_computer = ‘no’

Data sample

X = (age <=30, Income = medium, Student = yes, Credit\_rating = Fair)

Training data:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| age | income | student | credit\_rating | buys\_computer |
| <=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 |

Solution:

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

**X = (age <= 30 , income = medium, student = yes, credit\_rating = fair)**

**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

**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”)**

PART B

(PART B : TO BE COMPLETED BY STUDENTS)

***(Students must submit the soft copy as per following segments within two hours of the practical slot. The soft copy must be uploaded on the Blackboard or emailed to the concerned lab in charge faculties at the end of the practical in case the there is no Black board access available)***

|  |  |
| --- | --- |
| Roll No. | Name: |
| Class : | Batch : |
| Date of Experiment: | Date of Submission |
| Grade : | Time of Submission: |
| Date of Grading: | |

**B.1 Software Code written by student:**

***(Paste your c/c++/java code completed during the 2 hours of practical in the lab here)***

**B.2 Input and Output:**

***(Paste your program input and output in following format, If there is error then paste the specific error in the output part. In case of error with due permission of the faculty extension can be given to submit the error free code with output in due course of time. Students will be graded accordingly.)***

**Input Data:**

**Output Data:**

**B.3 Observations and learning:**

***(Students are expected to comment on the output obtained with clear observations and learning for each task/ sub part assigned)***

**B.4 Conclusion:**

*(****Students must write the conclusions based on their learning)***

**B.5 Questions of Curiosity**

Q1.What are the issues in classification? Explain each with the help of an example.

Q2.Summarize all approaches used for classification with their advantages and limitations.