Types of classification algorithm

i. Statistical ——— Naive Bayesian classification
ii. Distance
iii. Decision Tree ——— ID3, C4.5, CART
iv. Neural Network
v. Rule based

Naive Bayesian classification :

Bayesian classifiers are statistical classifiers they can predict class membership probabilities such as the probabilities that the given tuple belongs to a particular class.

Naive Bayesian classification is based on Bayes theorem.

$$P(H|X) = \frac{P(X|H)\,P(H)}{P(X)}$$

Predict a class label using naive bayesian classification for the following tuple.

X = (age = youth, income = medium, student = yes, credit rating = fair)

RID	age	income	student	credit rating	Class: buys comp
1	Youth	high	no	fair	no
2	Youth middle age	high	no	excellent	no
3	middle age	high	no	fair	Yes
4	senior	medium	no	fair	Yes
5	senior	low	Yes	fair	Yes
6	senior	low	Yes	excellent	no
7	middle age	low	Yes	excellent	Yes
8	Youth	medium	no	fair	no
9	Youth	low	Yes	fair	Yes
10	senior	medium	Yes	fair	Yes
11	Youth	medium	Yes	excellent	Yes
12	middle age	medium	no	excellent	Yes
13	middle age	high	Yes	fair	Yes
14	senior	medium	no	excellent	no

Step-1

In the given database the attributes are age, income, student & Credit-rating.

The class label attribute is buys-computer has 2 discrete values Yes & no.

Let C_1 correspond to the class buys-Computer = Yes

C_2 correspond to the class buys-computer = no

We need to maximize

$P(\text{buys\_computer} = \text{Yes} \mid X)$ or

$P(\text{buys\_computer} = \text{no} \mid X)$

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

$$\frac{P(X \mid \text{buys\_comp} = \text{Yes}) * P(\text{buys\_comp} = \text{Yes})}{P(X)}$$

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

$$\frac{P(X \mid \text{buys\_comp} = \text{No}) * P(\text{buys\_comp} = \text{No})}{P(X)}$$

$P(C_1)$

$P(\text{buys\_computer} = \text{Yes}) = 9/14 = 0.643$

$P(C_2) =$

$P(\text{buys\_comp} = \text{No}) = 5/14 = 0.357$

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

C_1

To calculate probability of

$P(X \mid C_i)$ for $i = 1, 2$

We calculate the following conditional probability

$P(\text{age} = \text{Youth} \mid \text{buys\_comp} = \text{Yes})$

$= 2/9 = P(\text{Youth} \cap \text{Yes})$

$P(\text{Yes})$

$= 0.22$

Cr P(age = Youth | buys_comp = no)

$= \frac{P(\text{Youth} \cap \text{no})}{P(\text{no})}$

$= 3/5 = 0.6$

P(income = medium | buys_comp = Yes)

$= \frac{P(\text{medium} \cap \text{Yes})}{P(\text{Yes})}$

$4/9 = 0.445$

P(income = medium | buys_comp = no)

$= \frac{P(\text{medium} \cap \text{no})}{P(\text{no})}$

$= 2/5 = 0.4$

P(student = Yes | buys_comp = Yes)

$\frac{P(\text{Yes} \cap \text{Yes})}{P(\text{Yes})} = \frac{6}{9} = 0.667$

P(student = no | buys_com = no)

$= \frac{1}{5} = 0.2$

4

$P(\text{Cr-rating} = \text{fair} \mid \text{buys-comp} = \text{Yes})$

$$= \frac{P(\text{fair} \cap \text{Yes})}{P(\text{Yes})} = \frac{6}{9} = 0.667$$

$P(\text{Cr-rating} = \text{fair} \mid \text{buys-comp} = \text{no})$

$$= \frac{P(\text{fair} \cap \text{no})}{P(\text{no})} = \frac{2}{5} = 0.4$$

$P(X \mid \text{buys-comp} = \text{Yes}) = P(\text{age} = \text{youth} \mid \text{buys-comp} = \text{Yes}) \times$
$P(\text{student} = \text{Yes} \mid \text{buys-comp} = \text{Yes}) \times$
$P(\text{income} = \text{Yes} \mid \text{buys-comp} = \text{Yes}) \times$
$P(\text{Cr-rating} = \text{fair} \mid \text{buys-comp} = \text{Yes})$

$= 0.222 \times 0.945 \times 0.667 \times 0.667$

$= 0.044$

$P(X \mid \text{buys-comp} = \text{No}) =$

$0.6 \times 0.4 \times 0.2 \times 0.4$

$= 0.019$

To find the class C_i that maximizes the given tuple X we calculate $P(C_i \mid X)$

$$P(C_i \mid X) = \frac{P(X \mid C_i) \times P(C_i)}{P(X)}$$

$$P(C_2|X) = \frac{P(X|C_2) * P(C_2)}{P(X)}$$

$P(X|C_1) * P(C_1)$

$= P(X|\text{buys\_com} = \text{yes}) * P(\text{yes})$

$= 0.044 \times 0.643$

$= 0.028$

$P(X|C_2) P(C_2)$

$= 0.019 \times 0.357 = 0.007$

$$P(B) = \sum_i P(B|A_i) P(A_i)$$

$P(X) = P(X|\text{yes}) P(\text{yes}) + P(X|\text{no}) * P(\text{no})$

$= 0.028 + 0.007$

$= 0.035$

$P(C_1|X) = \frac{0.028}{0.035} = 0.8$

$P(C_2|X) = \frac{0.007}{0.035} = 0.2$

∴ 0.028 is the max value the naive bayesian classifiers predicts buys_comp : yes for the tuple X

Date ....
Page ....

∴ 0.8 maps the class value C_1 the naive bayesian classifies predicts buys_com = yes for the tuple X.