Naïve Bayesian Classifier: Example

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
30…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

Classes:

C1:buys_computer= 'yes'; C2:buys_computer= 'no'

Unknown sample:

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

Naïve Bayes (Summary)

- Advantages :
  - Easy to implement
  - Good results obtained in most of the cases
- Disadvantages
  - Assumption: class conditional independence , therefore loss of accuracy
  - Practically, dependencies exist among variables
  - E.g., hospitals: patients: Profile: age, family history etc
    Symptoms: fever, cough etc., Disease: lung cancer, diabetes etc
  - Dependencies among these cannot be modeled by Naïve Bayesian Classifier

$P(B.C="yes" \mid A_1 A_2 A_3 A_4)$

$\propto P(A_1 A_2 A_3 A_4 \mid B.C="yes") P(B.C="yes")$

$= P(A_1|yes) \cdot P(A_2|yes) \cdot P(A_3|yes) \cdot P(A_4|yes) \cdot P(yes)$

A C E G

● Example ●

- P(Buy computer = yes) = $9/14 = 0.643$
  P(Buy computer = NO) = $5/14 = 0.357$

A. P(Age = "≤30" | B.C = "YES") = $2/9 = 0.222$

B. P(Age = "≤30" | B.C = 'NO") = $3/5 = 0.6$

C. P(Income = "medium" | B.C = "YES") = $4/9 = 0.444$

b. P(medium | NO) = $2/5 = 0.4$

E. P(Student = "yes" | B.C = "Yes") = $6/9 = 0.667$

F. P(Student | no) = $1/5 = 0.2$

G. P(Fair | yes) = $6/9 = 0.667$

H. P(Fair | no) = $2/5 = 0.4$

Therefore, P(yes | Given) $\propto (0.222 \times 0.444 \times 0.667 \times 0.667) \times 0.643$

$\vee$ $= 0.028$

P(no | Given) $\propto 0.007$.

$\therefore$ Decision $\Rightarrow$ "YES" (Buy computer)