Bayesian Classifier

Bayesian Classifier

- Statistical Classifier
- Predict Class Membership Probabilities
- Based on Bayes Theorem
- High Accuracy and Speed in large Databases
- Prior Probability: Probability of X, P(X)
- Posterior Probability: Probability of X when consider some condition Y, P(X|Y)
- A Simple Bayesian Classifier Naive Bayesian Classifier
 - Class Conditional Independence

MLE vs MAP

Maximum Likelihood Estimation

$$egin{aligned} heta_{MLE} &= rg\max_{ heta} f(X_1, X_2, \ldots, X_n \,|\, heta) \ &= rg\max_{ heta} \sum_i \log f(X_i \,|\, heta) \end{aligned}$$

Maximum A Priori

$$egin{aligned} heta_{MAP} &= rg \max_{ heta} f(heta \,|\, X_1, X_2, \ldots, X_n) \ &= rg \max_{ heta} \left(\log g(heta) + \sum_i \log f(X_i \,|\, heta)
ight) \end{aligned}$$

Conditional Independence

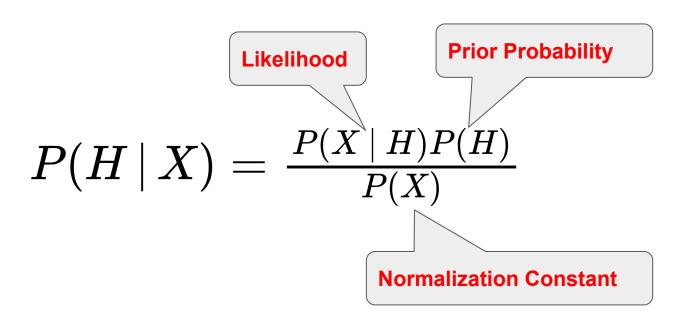
A and B are independent if,

$$P(A\cap B)=P(A) imes P(B) \ orall_{a,b}: P(A=a\cap B=b)=P(A=b) imes P(B=b)$$

A and B are conditionally independent given C if,

$$P(A,B\,|\,C) = P(A\,|\,C) imes P(B\,|\,C) \ orall_{a,b,c}: P(A=a\cap B=b\,|\,C=c) = P(A=b\,|\,C=c) imes P(B=b\,|\,C=c)$$

Bayes Theorem



Naive Bayes Classifier

- Attributes are conditionally independent
- ullet Consider *n*-dimensional attribute vector, $X=(X_1,X_2,\ldots,X_n)$
- ullet Consider m classes, $C=(C_1,C_2,\ldots,C_m)$
- ullet Naive Bayes predicts that a tuple belongs to some class C_i for given condition $oldsymbol{\mathcal{X}}$ if and only if,

$$P(C_i \mid X) > P(C_j \mid X) \ for \ 1 \leq j \leq m, \ j \neq i$$

ullet Goal: maximize $P(C_i \,|\, X)$

Naive Bayes Equations

Likelihood:
$$P(X \mid C_i) = \prod_{k=1}^n P(X_k \mid C_i)$$
 $= P(X_1 \mid C_i) imes P(X_2 \mid C_i) imes \dots imes P(X_n \mid C_i)$

Categorical Attribute:

$$P(X_k \,|\, C_i) = rac{|D_{X_k,C_i}|}{D_{C_i}}$$

Continuous-valued Attribute: Gaussian Distribution

$$egin{aligned} P(X_k \,|\, C_i) &= g(X_k, \mu_{C_i}, \sigma_{C_i}) \ g(X, \mu, \sigma) &= rac{1}{\sqrt{2\pi}\sigma} e^{-rac{(X-\mu)^2}{2\sigma^2}} \end{aligned}$$

Sample Data: D

ID	age	income	student	creditRating	buyComputer
1	youth	high	yes	fair	no
2	youth	low	no	excellent	no
3	middleaged	medium	no	fair	yes
4	middleaged	medium	no	fair	yes
5	senior	medium	yes	fair	yes
6	senior	low	no	excellent	no
7	midleaged	high	yes	excellent	yes
8	youth	medium	no	fair	no
9	youth	low	yes	fair	yes
10	senior	high	yes	fair	yes
11	youth	high	yes	excellent	yes
ID	age	income	student	creditRating	buyComputer
1	youth	low	no	fair	???

Sample Data: Problem Definition

ID	age	income	student	creditRating	buyComputer
1	youth	low	no	fair	???

Here,

$$X = (age = youth, income = low, student = no, creditRating = fair)$$

Find,
$$= \max(P(C_{yes} \,|\, X), P(C_{no} \,|\, X))$$

$$= \max \left(rac{P(X \,|\, C_{yes})P(C_{yes})}{P(X)}, rac{P(X \,|\, C_{no})P(C_{no})}{P(X)}
ight)$$

$$= \max(P(X | C_{yes}) P(C_{yes}), P(X | C_{no}) P(C_{no}))$$

 $= \max(P(X \,|\, C_{yes}) P(C_{yes}), P(X \,|\, C_{no}) P(C_{no}))$

ID	age	income	student	creditRating	buyComputer
1	youth	low	no	fair	???

$$P(buyComputer = yes)$$

$$P(age = youth \,|\, buyComputer = yes)$$

$$P(income = low \,|\, buyComputer = yes)$$

$$P(student = no \mid buyComputer = yes)$$

$$P(creditRating = fair | buyComputer = yes)$$

ID	age	income	student	creditRating	buyComputer
1	youth	high	yes	fair	no
2	youth	low	no	excellent	no
3	middleaged	medium	no	fair	yes
4	middleaged	medium	no	fair	yes
5	senior	medium	yes	fair	yes
6	senior	low	no	excellent	no
7	midleaged	high	yes	excellent	yes
8	youth	medium	no	fair	no
9	youth	low	yes	fair	yes
10	senior	high	yes	fair	yes
11	youth	high	yes	excellent	yes
12	middleaged	medium	no	excellent	yes
13	middleaged	low	yes	fair	yes
14	senior	high	yes	excellent	no

$$P(buyComputer = yes) = \frac{9}{14}$$

$$= \frac{9}{14}$$

$$P(age = youth \,|\, buyComputer = yes)$$

$$P(income = low \,|\, buyComputer = yes$$

$$P(student = no \,|\, buyComputer = yes$$

$$P(creditRating = fair \, | \, buyCompute$$

	I D	age	income	student	creditRatin g	buyComputer
	1	youth	high	yes	fair	no
	2	youth	low	no	excellent	no
	3	middleaged	medium	no	fair	yes
	4	middleaged	medium	no	fair	yes
	5	senior	medium	yes	fair	yes
	6	senior	low	no	excellent	no
	7	midleaged	high	yes	excellent	yes
	8	youth	medium	no	fair	no
	9	youth	low	yes	fair	yes
	1 0	senior	high	yes	fair	yes
s	1 1	youth	high	yes	excellent	yes
s	1 2	middleaged	medium	no	excellent	yes
	1 3	middleaged	low	yes	fair	yes
	1	senior	high	yes	excellent	no

$$P(buyComputer = yes) = \frac{9}{14}$$

$$=\frac{9}{14}$$

$$P(age = youth \mid buyComputer = yes)$$

$$P(income = low \,|\, buyComputer = yes$$

$$P(student = no \,|\, buyComputer = yes$$

$$P(creditRating = fair \, | \, buyCompute$$

	I D	age	income	student	creditRatin g	buyComputer
	1	youth	high	yes	fair	no
	2	youth	low	no	excellent	no
	3	middleaged	medium	no	fair	yes
	4	middleaged	medium	no	fair	yes
	5	senior	medium	yes	fair	yes
	6	senior	low	no	excellent	no
	7	midleaged	high	yes	excellent	yes
	8	youth	medium	no	fair	no
	9	youth	low	yes	fair	yes
	1 -0	$\frac{2}{9}$ senior	high	yes	fair	yes
s	1	youth	high	yes	excellent	yes
s	1 2	middleaged	medium	no	excellent	yes
Ŀ €	1 3	middleaged	low	yes	fair	yes
	1 4	senior	high	yes	excellent	no

Necessary Statistics,

$$P(buyComputer = yes) = \frac{9}{14}$$

$$=\frac{9}{14}$$

 $P(age = youth \mid buyComputer = yes)$

 $P(income = low \,|\, buyComputer = yes$

P(student = no | buyComputer = ye)

					PID III	J. O
	I D	age	income	student	creditRatin g	buyComputer
	1	youth	high	yes	fair	no
	2	youth	low	no	excellent	no
	3	middleaged	medium	no	fair	yes
	4	middleaged	medium	no	fair	yes
	5	senior	medium	yes	fair	yes
	6	senior	low	no	excellent	no
	7	midleaged	high	yes	excellent	yes
	8	youth	medium	no	fair	no
	9	youth	low	yes	fair	yes
	1 0	$\frac{2}{9}$ senior	high	yes	fair	yes
s	1 1	$\stackrel{\text{(outn)}}{=} \frac{2}{9}$	high	yes	excellent	yes
s	1 2	middleaged	medium	no	excellent	yes
$t\epsilon$	1 3	middleaged	low	yes	fair	yes
	1 4	senior	high	yes	excellent	no

Necessary Statistics,

$$P(buyComputer = yes) = \frac{9}{14}$$

$$=\frac{9}{14}$$

 $P(age = youth \mid buyComputer = yes)$

 $P(income = low \,|\, buyComputer = ye.$

P(student = no | buyComputer = ye)

	I D	age	income	student	creditRatin g	buyComputer
	1	youth	high	yes	fair	no
	2	youth	low	no	excellent	no
	3	middleaged	medium	no	fair	yes
	4	middleaged	medium	no	fair	yes
	5	senior	medium	yes	fair	yes
	6	senior	low	no	excellent	no
	7	midleaged	high	yes	excellent	yes
	8	youth	medium	no	fair	no
	9	youth	low	yes	fair	yes
	1 —0	$\frac{2}{9}$ senior	high	yes	fair	yes
s	1	$\stackrel{\text{outn}}{=} \frac{2}{9}$	high	yes	excellent	yes
s	1 2	middleaged	$=\frac{3}{0}$	no	excellent	yes
$t\epsilon$	1 3	middleaged	IOW	yes	fair	yes
	1 4	senior	high	yes	excellent	no

Necessary Statistics,

$$P(buyComputer = yes) = \frac{9}{14}$$

$$=\frac{9}{14}$$

 $P(age = youth \mid buyComputer = yes)$

 $P(income = low \,|\, buyComputer = yes$

P(student = no | buyComputer = ye)

	I D	age	income	student	creditRatin g	buyComputer
	1	youth	high	yes	fair	no
	2	youth	low	no	excellent	no
	3	middleaged	medium	no	fair	yes
	4	middleaged	medium	no	fair	yes
	5	senior	medium	yes	fair	yes
	6	senior	low	no	excellent	no
	7	midleaged	high	yes	excellent	yes
	8	youth	medium	no	fair	no
Ì	9	youth	low	yes	fair	yes
	1	$\frac{2}{9}$ senior	high	yes	fair	yes
s	1 1	$\stackrel{\text{outn}}{=} \frac{2}{9}$	high	yes	excellent	yes
s	1 2	middleaged	$=\frac{3}{9}$	no	excellent	yes
$t\epsilon$	1	middleaged	iow	$=\frac{4^{y}}{0}$;s	fair	yes
	1 4	senior	high	yes	excellent	no

$$P(buyComputer = yes) \equiv \frac{9}{14}$$
 $P(age = youth \mid buyComputer = yes) \equiv \frac{2}{9}$
 $P(income = low \mid buyComputer = yes) \equiv \frac{2}{9}$
 $P(student = no \mid buyComputer = yes) \equiv \frac{3}{9}$
 $P(creditRating = fair \mid buyComputer = yes) \equiv \frac{4}{9}$

$$P(X | buyComputer = yes) imes P(buyComputer = yes) = rac{2}{9} imes rac{2}{9} imes rac{3}{9} imes rac{4}{9} imes rac{9}{14} = rac{8}{1701}$$

ID	age	income	student	creditRating	buyComputer
1	youth	low	no	fair	???

$$P(buyComputer = no)$$

$$P(age = youth \,|\, buyComputer = no)$$

$$P(income = low \,|\, buyComputer = no$$

$$P(student = no \,|\, buyComputer = no$$

$$P(creditRating = fair \, | \, buyComputer = no)$$

	ID	age	income	student	creditRating	buyComputer
	1	youth	high	yes	fair	no
	2	youth	low	no	excellent	no
	3	middleaged	medium	no	fair	yes
	4	middleaged	medium	no	fair	yes
	5	senior	medium	yes	fair	yes
	6	senior	low	no	excellent	no
	7	midleaged	high	yes	excellent	yes
	8	youth	medium	no	fair	no
	9	youth	low	yes	fair	yes
8	10	senior	high	yes	fair	yes
)	11	youth	high	yes	excellent	yes
4	12	middleaged	medium	no	excellent	yes
)	13	middleaged	low	yes	fair	yes
8	14	senior	high	yes	excellent	no

$$P(buyComputer = no) = \frac{5}{14}$$

$$=\frac{5}{14}$$

$$P(age = youth \mid buyComputer = no)$$

$$P(income = low \,|\, buyComputer = no)$$

$$P(student = no \,|\, buyComputer = no)$$

$$P(creditRating = fair \, | \, buyCompute$$

	I D	age	income	student	creditRatin g	buyComputer
	1	youth	high	yes	fair	no
	2	youth	low	no	excellent	no
	3	middleaged	medium	no	fair	yes
	4	middleaged	medium	no	fair	yes
	5	senior	medium	yes	fair	yes
	6	senior	low	no	excellent	no
	7	midleaged	high	yes	excellent	yes
	8	youth	medium	no	fair	no
	9	youth	low	yes	fair	yes
	1	senior	high	yes	fair	yes
))	1	youth	high	yes	excellent	yes
) [C	1 2	middleaged	medium	no	excellent	yes
t	1 2 %	=no	low	yes	fair	yes
	1 4	senior	high	yes	excellent	no

$$P(buyComputer = no) = \frac{5}{14}$$

$$=\frac{5}{14}$$

$$P(age = youth | buyComputer = no)$$

$$P(income = low \,|\, buyComputer = no)$$

$$P(student = no \,|\, buyComputer = no)$$

$$P(creditRating = fair \, | \, buyCompute$$

	I D	age	income	student	creditRatin g	buyComputer
	1	youth	high	yes	fair	no
	2	youth	low	no	excellent	no
	3	middleaged	medium	no	fair	yes
	4	middleaged	medium	no	fair	yes
	5	senior	medium	yes	fair	yes
	6	senior	low	no	excellent	no
	7	midleaged	high	yes	excellent	yes
	8	youth	medium	no	fair	no
	9	youth	low	yes	fair	yes
	1 0	$\frac{3^{\text{sc}}}{5}$ nior	high	yes	fair	yes
))	1	youth	high	yes	excellent	yes
) [C	1 2	middleaged	medium	no	excellent	yes
$t^{(i)}$	1 e r	=no	low	yes	fair	yes
	1 4	senior	high	yes	excellent	no

$$P(buyComputer = no) = \frac{5}{14}$$

$$=\frac{5}{14}$$

$$P(age = youth | buyComputer = no)$$

$$P(income = low \, | \, buyComputer = no)$$

$$P(student = no \,|\, buyComputer = no)$$

$$P(creditRating = fair \, | \, buyCompute$$

	I D	age	income	student	creditRatin g	buyComputer
	1	youth	high	yes	fair	no
	2	youth	low	no	excellent	no
	3	middleaged	medium	no	fair	yes
	4	middleaged	medium	no	fair	yes
	5	senior	medium	yes	fair	yes
	6	senior	low	no	excellent	no
	7	midleaged	high	yes	excellent	yes
	8	youth	medium	no	fair	no
	9	youth	low	yes	fair	yes
	1 0	$\frac{3^{s}}{5}$ nior	high	yes	fair	yes
)	1	$\stackrel{\text{youth}}{=} \frac{2}{5}$	high	yes	excellent	yes
) [C	1) 2	middleaged	medium	no	excellent	yes
$t^{(i)}$	1 e r	=no	low	yes	fair	yes
	1 4	senior	high	yes	excellent	no

Necessary Statistics,

$$P(buyComputer = no) = \frac{5}{14}$$

$$=\frac{5}{14}$$

 $P(age = youth \mid buyComputer = no)$

 $P(income = low \,|\, buyComputer = no$

 $P(student = no \,|\, buyComputer = no)$

	I D	age	income	student	creditRatin g	buyComputer
	1	youth	high	yes	fair	no
	2	youth	low	no	excellent	no
	3	middleaged	medium	no	fair	yes
	4	middleaged	medium	no	fair	yes
	5	senior	medium	yes	fair	yes
	6	senior	low	no	excellent	no
	7	midleaged	high	yes	excellent	yes
	8	youth	medium	no	fair	no
	9	youth	low	yes	fair	yes
	1 0	$\frac{3^{\epsilon}}{5}$	high	yes	fair	yes
o)	1	y_0 utn $\frac{2}{5}$	high -	yes	excellent	yes
) (C	1 2	middleaged	$=\frac{5}{5}$	no	excellent	yes
t	1 2 %	=no	iow	= 2 ^c 5 yes	fair	yes
	1 4	senior	high	yes	excellent	no

$$P(buyComputer = no) = \frac{5}{14}$$
 $P(age = youth \mid buyComputer = no) = \frac{3}{5}$
 $P(income = low \mid buyComputer = no) = \frac{2}{5}$
 $P(student = no \mid buyComputer = no) = \frac{3}{5}$
 $P(creditRating = fair \mid buyComputer = no) = \frac{2}{5}$

$$P(X | buyComputer = no) imes P(buyComputer = no) = rac{3}{5} imes rac{2}{5} imes rac{3}{5} imes rac{3}{5} imes rac{5}{14} = rac{18}{875}$$

Sample Data: Prediction

$$P(X | buyComputer = yes) \times P(buyComputer = yes) < P(X | buyComputer = no) \times P(buyComputer = no)$$

buyComputer = no

$$P(X | buyComputer = no) imes P(buyComputer = no) = rac{3}{5} imes rac{2}{5} imes rac{3}{5} imes rac{5}{14}$$
 $= rac{18}{875}$
 $P(X | buyComputer = yes) imes P(buyComputer = yes) = rac{2}{9} imes rac{2}{9} imes rac{3}{9} imes rac{4}{9} imes rac{9}{14}$
 $= rac{8}{1701}$

Zero Probability

$$egin{aligned} &= P(age = middleaged \,|\, buyComputer = no) \ &= rac{0}{5} \end{aligned}$$

X = (age = middleaged, income = low,

I D	age	income	student	creditRatin g	buyComputer
1	youth	high	yes	fair	no
2	youth	low	no	excellent	no
3	middleaged	medium	no	fair	yes
4	middleaged	medium	no	fair	yes
5	senior	medium	yes	fair	yes
6	senior	low	no	excellent	no
7	midleaged	high	yes	excellent	yes
8	youth	medium	no	fair	no
9	youth	low	yes	fair	yes
st^{1}	udent =	no,cr	edit Ra	ating =	fair)
1	youth	high	yes	excellent	yes
1 2	middleaged	medium	no	excellent	yes
1 3	middleaged	low	yes	fair	yes
1 4	senior	high	yes	excellent	no ZT

Laplacian Correction

Add an extra count for each different value.

For attribute age,

- → count(youth | no) = 4 (Actually 3)
- → count(middleaged | no) = 1 (Actually 0)
- → count(senior | no) = 3 (Actually 2)
- \rightarrow count(no) = 8 (Actually 5)

$$egin{aligned} &= P(age = middleaged \, | \, buyComputer = no) \ &= rac{1}{8} \end{aligned}$$

$$X=(age=middleaged,income=low,% \label{eq:X} X=(age=middleaged,income=low,% \label{eq:X} X=(age=middleaged,in$$

	I D	age	income	student	creditRatin g	buyComputer
	1	youth	high	yes	fair	no
	2	youth	low	no	excellent	no
	3	middleaged	medium	no	fair	yes
	4	middleaged	medium	no	fair	yes
	5	senior	medium	yes	fair	yes
	6	senior	low	no	excellent	no
	7	midleaged	high	yes	excellent	yes
	8	youth	medium	no	fair	no
	9	youth	low	yes	fair	yes
	1 0	senior	high	yes	fair	yes
	1 1	youth	high	yes	excellent	yes
s	¹ ₽ ≀	udent=	medium no, cr	$red\overset{\scriptscriptstyleno}{it}Rd$	$ating = % rac{{{ ext{excellent}}}{{{ ext{ord}}}} = % rac{{{ ext{excellent}}}}{{{ ext{ord}}}} = % rac{{{ ext{excellent}}}}}{{{ ext{ord}}}} = % rac{{{ ext{excellent}}}}{{{ ext{ord}}}} = % rac{{{ ext{excellent}}}}{{{ ext{ord}}}} = % rac{{{ ext{excellent}}}}{{{ ext{ord}}}} = % rac{{{ excellent}}}}{{{ ext{ord}}}} = % rac{{{ ext{excellent}}}}{{{ ext{ord}}}} = % rac{{{ ext{ord}}}}{{{ ext{ord}}}} = % ra$	$=fair^{^{\mathrm{yes}}})$
	1 3	middleaged	low	yes	fair	yes
	1 4	senior	high	yes	excellent	no

Thank You