Naïve Bayes Intuition

Dr. Muhammad Wasim

The Idea of Naïve Bayes Classifier

Training Dataset

Example	age	income	student	credit_rating	Class:buys_computer
1	<=30	high	no	fair	no
2	<=30	high	no	excellent	no
3	31-40	high	no	fair	yes
4	>40	medium	no	fair	yes
5	>40	low	yes	fair	yes
6	>40	low	yes	excellent	no
7	31-40	low	yes	excellent	yes
8	<=30	medium	no	fair	no
9	<=30	low	yes	fair	yes
10	>40	medium	yes	fair	yes
11	<=30	medium	yes	excellent	yes
12	31-40	medium	no	excellent	yes
13	31-40	high	yes	fair	yes
14	>40	medium	no	excellent	no

Unseen Example

age	income	student	credit_rating	Class
<=30	medium	yes	fair	?

Why Naïve Bayes is called as Naïve?

$$P(yes|x_1, x_2, x_3, x_4) = \frac{P(x_1, x_2, x_3, x_4|yes) \times P(yes)}{P(x_1, x_2, x_3, x_4)}$$

$$P(no|x_1, x_2, x_3, x_4) = \frac{P(x_1, x_2, x_3, x_4|no) \times P(no)}{P(x_1, x_2, x_3, x_4)}$$

Naïve Assumption: Features are independent of each other

$$P(x_1, x_2, x_3, x_4|yes) = P(x_1|yes) \times P(x_2|yes) \times P(x_3|yes) \times P(x_4|yes)$$

Prior Probability and Conditional Probability

x_1	y
0	Yes
0	Yes
1	Yes
0	Yes
0	Yes
1	No
1	No
1	No
1	No
1	No

$P(yes)$ =

$P(no)$ =

$P(x_1 = 1|yes)$ =

$P(x_1 = 0|yes)$ =

$P(x_1 = 0|no)$ =

$P(x_1 = 1|no)$ =

A Working Example

Training Dataset

Example	age	income	student	credit_rating	Class
1	<=30	high	no	fair	no
2	<=30	high	no	excellent	no
3	31-40	high	no	fair	yes
4	>40	medium	no	fair	yes
5	>40	low	yes	fair	yes
6	>40	low	yes	excellent	no
7	31-40	low	yes	excellent	yes
8	<=30	medium	no	fair	no
9	<=30	low	yes	fair	yes
10	>40	medium	yes	fair	yes
11	<=30	medium	yes	excellent	yes
12	31-40	medium	no	excellent	yes
13	31-40	high	yes	fair	yes
14	>40	medium	no	excellent	no

$$P(yes) = \frac{9}{14} = 0.643 \qquad P(no) = \frac{5}{14} = 0.357$$

Unseen Example

age	income	student	credit_rating	Class
<=30	medium	yes	fair	?

$$P(age \leq 30 \,|yes) = \frac{2}{9} = 0.222$$

$$P(age \leq 30 \,|no) = \frac{3}{5} = 0.600$$

$$P(income = medium|yes) = \frac{4}{9} = 0.444$$

$$P(income = medium|no) = \frac{2}{5} = 0.400$$

$$P(student = yes|yes) = \frac{6}{9} = 0.667$$

$$P(student = yes|no) = \frac{1}{5} = 0.200$$

$$P(c.rating = "fair" \,|yes) = \frac{6}{9} = 0.667$$

$$P(c.rating = "fair"|no) = \frac{2}{5} = 0.400$$

A Working Example (Cont.)

$$P(X|yes) = P(age \leq 30\ |yes) \times P(income = medium|yes) \times P(student = yes|yes) \times P(c.rating = "fair"\ |yes)$$

$$= 0.222 \times 0.444 \times 0.667 \times 0.667 = 0.044$$

$$P(X|no) = P(age \leq 30\ |no) \times P(income = medium|no) \times P(student = yes|no) \times P(c.rating = "fair"\ |no)$$

$$= 0.600 \times 0.400 \times 0.200 \times 0.400 = 0.019$$

$$P(X|yes) \times P(yes) = 0.044 \times 0.643 = 0.028$$

$$P(X|no) \times P(no) = 0.019 \times 0.357 = 0.007$$

age	income	student	credit_rating	Class
<=30	medium	yes	fair	