Introduction to Data Science

Navie Bayes Classification

Arijit Mondal

Dept. of Computer Science & Engineering

Indian Institute of Technology Patna

arijit@iitp.ac.in

Probability

- Rules for probability

Probability

- Rules for probability
  - Let A and B are two events

Probability

- Rules for probability
  - Let A and B are two events
  - Then, $P(A \wedge B) = P(A) + P(B) - P(A \vee B)$

AND

OR

Probability

- Rules for probability
  - Let A and B are two events
  - Then, $P(A \wedge B) = P(A) + P(B) - P(A \vee B)$
  - If A and B are independent, then $P(A \wedge B) = P(A) \cdot P(B)$

Probability

- Rules for probability
  - Let A and B are two events
  - Then, $P(A \wedge B) = P(A) + P(B) - P(A \vee B)$
  - If A and B are independent, then $P(A \wedge B) = P(A) \cdot P(B)$
  - $P(A \wedge B) = P(A) \cdot P(B|A) = P(B)P(A|B)$

Probability

- Rules for probability
  - Let A and B are two events
  - Then, $P(A \wedge B) = P(A) + P(B) - P(A \vee B)$
  - If A and B are independent, then $P(A \wedge B) = P(A) \cdot P(B)$
  - $P(A \wedge B) = P(A) \cdot P(B|A) = P(B)P(A|B)$
  - Bayes theorem: $P(A|B) = \dfrac{P(A) \cdot P(B|A)}{P(B)}$

Naive Bayes

- Let $X = (x_1, \ldots, x_n)$ need to be classfied into one of m classes $C_1, \ldots, C_m$

Naive Bayes

- Let $X = (x_1, \ldots, x_n)$ need to be classfied into one of m classes $C_1, \ldots, C_m$
- We compute the probability of each possible class given X

$p(c_i | X)$

$p(c_1 | X)$

$p(c_2 | X)$

$\boxed{c_r \quad \text{max}}$

$p(c_m | X)$

Naive Bayes

Height, weight

- Let $X = (x_1, \ldots, x_n)$ need to be classfied into one of m classes $C_1, \ldots, C_m$
- We compute the probability of each possible class given X
- By Basyes theorem, $P(C_i|X) = \frac{P(C_i) \cdot P(X|C_i)}{P(X)}$

$$P(C_j|X) = \frac{P(C_j) \cdot P(X|C_j)}{P(X)}$$

$X_1 - C_1$

$X_2 - C_2$

$\vdots$

$X_p - C_K$

$$\frac{C_i}{|\gamma|}$$

Naive Bayes

- Let $X = (x_1, \ldots, x_n)$ need to be classfied into one of m classes $C_1, \ldots, C_m$
- We compute the probability of each possible class given X
- By Basyes theorem, $P(C_i|X) = \dfrac{P(C_i) \cdot P(X|C_i)}{P(X)}$
- Final class will be $\cancel{P(C_i|X)}$ $= \arg\max_{i=1,\ldots,m} P(C_i) \cdot P(X|C_i)$

$C(X)$

Naive Bayes

- Let $X = (x_1, \ldots, x_n)$ need to be classfied into one of m classes $C_1, \ldots, C_m$
- We compute the probability of each possible class given X
- By Basyes theorem, $P(C_i|X) = \dfrac{P(C_i) \cdot P(X|C_i)}{P(X)}$

$X =$

- Final class will be $P(C_i|X) = \arg\max_{i=1,\ldots,m} P(C_i) \cdot P(X|C_i)$
- Assuming independent world (naive assumption)

Naive Bayes

- Let $X = (x_1, \ldots, x_n)$ need to be classfied into one of m classes $C_1, \ldots, C_m$

H, W

M, F

- We compute the probability of each possible class given X
- By Basyes theorem, $P(C_i|X) = \dfrac{P(C_i) \cdot P(X|C_i)}{P(X)}$
- Final class will be $\cancel{P(C_i|X)} = \arg\max_{i=1,\ldots,m} P(C_i) \cdot P(X|C_i)$

C(X)

- Assuming independent world (naive assumption)
- Under the assumption, final class will be $\cancel{P(C_i|X)} = \arg\max_{i=1,\ldots,m} \boxed{P(C_i) \prod_{j=1}^{n} P(x_j|C_i)}$

C(X)

log

<< 1

Example

- Given the following observations

Day	Outlook	Temp	Humidity	Beach?
1	Sunny	High	High	Yes
2	Sunny	High	Normal	Yes
3	Sunny	Low	Normal	No
4	Sunny	Mild	High	Yes
5	Rain	Mild	Normal	No
6	Rain	High	High	No
7	Rain	Low	Normal	No
8	Cloudy	High	High	No
9	Cloudy	High	Normal	Yes
10	Cloudy	Mild	Normal	No

image source: Data Science Design Manual

Example

- Given the following observations
- Find $P(\text{Beach}|(\text{Sunny}, \text{Mild}, \text{High})) = ?$

$= P(B) \; P(S, M, H \mid B)$

$= P(B) \; P(S|B) \; P(M|B) \; P(H|B)$

$\frac{4}{10} \cdot \frac{3}{4} \quad \frac{1}{4} \quad \frac{2}{4} = 0.03$

Day	Outlook	Temp	Humidity	Beach?
1	Sunny	High	High	Yes
2	Sunny	High	Normal	Yes
3	Sunny	Low	Normal	No
4	Sunny	Mild	High	Yes
5	Rain	Mild	Normal	No
6	Rain	High	High	No
7	Rain	Low	Normal	No
8	Cloudy	High	High	No
9	Cloudy	High	Normal	Yes
10	Cloudy	Mild	Normal	No

$3/4, \; \boxed{\cancel{1/2}}$

image source: Data Science Design Manual

Example

- Given the following observations
- Find $P(\text{Beach}|(\text{Sunny}, \text{Mild}, \text{High})) =?$ 0.03
- Find $P(\text{NoBeach}|(\text{Sunny}, \text{Mild}, \text{High})) =?$

P(NB) . P(S|NB) P(M|NB) P(H|NB)

≈ 0.01

Day	Outlook	Temp	Humidity	Beach?
1	Sunny	High	High	Yes
2	Sunny	High	Normal	Yes
3	Sunny	Low	Normal	No
4	Sunny	Mild	High	Yes
5	Rain	Mild	Normal	No
6	Rain	High	High	No
7	Rain	Low	Normal	No
8	Cloudy	High	High	No
9	Cloudy	High	Normal	Yes
10	Cloudy	Mild	Normal	No

1/10 2^{10} 0/1

image source: Data Science Design Manual