Data Science

Marc Tommasi

October 10, 2023

Outline

Outline

Conditional probability and Bayes

- Notations:
  - $P(A)$ the probability of an event \(A)
  - $P(A \cap B)$ or $P(A, B)$ the probability of having both events A and B.
  - $P(A \mid B)$ the probability of having the event \(A) knowing \(B).

Definition of Conditional probabilities

$$P(A|B) = \frac{P(A,B)}{P(B)} \text{ or } P(B|A) = \frac{P(A,B)}{P(A)}$$

So
Donc

$$P(A,B) = P(A|B)P(B) = P(B|A)P(A)$$

Bayes Theorem

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}$$

A little abuse of notations

- $P(X = \boldsymbol{x})$ is written $P(\boldsymbol{x})$. ($P(\boldsymbol{x})$ means that the random variable X has value $\boldsymbol{x}$ hence this is the probability of the event $X = \boldsymbol{x}$)
- *Example* : In the case where we have several attributes $P(X = (0,0))$, $P(X = (0,1))$, are written $P(0,0)$, $P(0,1)$, etc... also written by $P(\boldsymbol{x})$ or $P(x_1, x_2)$

Back to learning

Reminders

- The data are generated by a fixed but unknown joint probability of having a data description $\boldsymbol{x}$ and a class y, written $P(\boldsymbol{x}, y)$.
- We want to solve the problem of finding the best y when we observe a $\boldsymbol{x}$.

Bayes' rule

- The best rule is

$$\operatorname*{argmax}_{y} P(y \mid \boldsymbol{x})$$

- Bayes error: Error of this rule
- is the smallest error that can be made for this learning if the examples are described by $\boldsymbol{x}$.

Difficult to calculate

- $P(y \mid \boldsymbol{x})$ can't be calculated because P is unknown.
- If we apply the ERM principle, by Bayes' rule we have

$$P(y \mid \boldsymbol{x}) = \frac{P(y)P(\boldsymbol{x} \mid y)}{P(\boldsymbol{x})}$$

- We're looking for the value of y that maximizes this quantity. But $P(\boldsymbol{x})$ does not depend on y. Simply solve

$$\underset{y}{\operatorname{argmax}}\, P(y)P(\boldsymbol{x} \mid y)$$

- Problem: the calculation can't be done efficiently $\boldsymbol{x}$ and of the values it can take. Example: In the binary case with 5 attributes, we have 2^5 possibilities and therefore 2 times 2^5 quantities to be estimated for all cases of $P(\boldsymbol{x} \mid y)$.

The chain rule

$$\begin{aligned} P(x_1, x_2, \ldots x_n) &= P(x_1 \mid x_2, \ldots x_n) P(x_2, \ldots x_n) \\ &= P(x_1 \mid x_2, \ldots x_n) P(x_2, \mid x_3, \ldots x_n) P(x_3, \ldots x_n) \ldots \end{aligned}$$

- Obtained by applying $P(A, B) = P(A|B)P(B)$ repeatedly when B is an event that can be a conjunction of events.
- By recursion

$$P(\boldsymbol{x} \mid y) = \prod_{k=1}^{n} P(x_k \mid x_{k-1}, \ldots, x_1, y)$$

Conditional Independence and Naive Bayes

- If A and B are independent, then $P(A \mid B) = P(A)$.
- If we assume that all attributes are independent, i.e. if x_i and x_j are independent for every

$i \neq j$, then the product is written :

$$P(\boldsymbol{x} \mid y) = \prod_{k=1}^{n} P(x_k \mid y)$$

- To calculate each of these $P(x_k \mid y)$ just count!
- It's a strong approximation!
- But the calculation is low complexity.

$$\operatorname*{argmax}_{y} P(y) \prod_{k=1}^{n} P(x_k \mid y)$$