Discrete Maths

24

Bayes' Theorem

- Important theorem associated with conditional probability.

Let $A_1, A_2, \ldots, A_n$ be mutually disjoint events (that is mutually exclusive which can not occur at the same time), Also, $P(A_i) \neq 0$
$i = 1, 2, \ldots, n$

For any arbitrary event E which is a subset of $\bigcup_{i=1}^{n} A_i$ such that

$P(E) > 0$, $\Rightarrow \Rightarrow$

We have

$$P\left(A_i/E\right) = \frac{P(A_i)\cdot P\left(E/A_i\right)}{\sum_{i=1}^{n} P(A_i)\cdot P\left(E/A_i\right)}$$

$$i = 1, 2, \ldots, n$$

Here, simplicity wise

$$P(E)\cdot P\left(A_i/E\right) = P(A_i)\cdot P\left(E/A_i\right)$$

because,

Given

$\because E \subset \bigcup_{i=1}^{n} A_i$

$E = E \cap \left(\bigcup_{i=1}^{n} A_i\right)$

$= \bigcup_{i=1}^{n} E \cap A_i$

$$P(E) = P\left[\bigcup_{i=1}^{n} E \cap A_i\right]$$

$$= \sum_{i=1}^{n} P(E \cap A_i)$$

$\because P(E/A_i) = \frac{P(E \cap A_i)}{P(A_i)}$

$$= \sum_{i=1}^{n} P(A_i) \cdot P(E/A_i)$$

In other words

$$P(A) \cdot P(B/A) = P(B) \cdot P(A/B)$$

Observations (Not to confuse with steps)

(A) The probabilities $P(A_1), P(A_2), \ldots, P(A_n)$ are called as the 'a priori probabilities' because they exist before we get any information from experiment itself.

(B) The probabilities $P(E/A_i)$ that is

$$P(E/A_1) =$$

$$P(E/A_2) =$$

$$P(E/A_3) = \quad \ldots$$

are called likelyhoods

) because they indicate how likely event E under ~~consideration~~ is to occur, given each and every a priori probability ~~(one at a time)~~

(c) The probabilities $P(A_i/E)$, $i = 1, 2, .., n$ are called "Posterior probabilities" that is

$P(A_1/E)$, $P(A_2/E)$, $P(A_3/E)$ . . . , $P(A_n/E)$

because they are computed after the results of the experiments are known.

(Pg. 350 onward Data Mining C&T by Han Kamber Pei)

Bayes' theorem is useful in that it provides a way of calculating the posterior probabilities.

i.e $P(H/X)$ from $P(H)$, $P(X)$ and $P(X/H)$

$$P(H/X) = \frac{P(H) \cdot P(X/H)}{P(X)}$$

, here, X is considered evidence.

~~Let H be some~~ hypothesis such as that ~~the data tuple~~ X belongs to a specified class.

naive ~~Bayesian~~ classification
(simple)

Assumption

the attributes' values are conditionally independent of one another, given the class label of tuple.

i.e. having trees have no connection to rainfall.

or rainfall has no connection ~~to earth~~ water level, etc.

In real world this assumption is not applicable every time.

But, it would be ~~extremely~~ computationally expensive to compute

$P(X/C_i)$ for large data sets.

To simplify,
naive assumption of class-conditional-independence is made.

likelyhoods

$$P(X/C_i) = \prod_{k=1}^{n} P(x_k / C_i)$$

test il

$$P(X/C_i) = P(x_1/C_i) \times P(x_2/C_i) \times P(x_3/C_i) \times \ldots \times P(x_n/C_i)$$

$P(E/A_i)$

∵ product rule apply only when events are independent. Otherwise, it fails.

pg. 338 page. 350

Table

Class-Labeled Training Tuples from the *AllElectronics* Customer Database

RID	*age*	*income*	*student*	*credit_rating*	*Class: buys_computer*
1	youth	high	no	fair	no
2	youth	high	no	excellent	no
3	middle_aged	high	no	fair	yes
4	senior	medium	no	fair	yes
5	senior	low	yes	fair	yes
6	senior	low	yes	excellent	no
7	middle_aged	low	yes	excellent	yes
8	youth	medium	no	fair	no
9	youth	low	yes	fair	yes
10	senior	medium	yes	fair	yes
11	youth	medium	yes	excellent	yes
12	middle_aged	medium	no	excellent	yes
13	middle_aged	high	yes	fair	yes
14	senior	medium	no	excellent	no

X = (age = youth, income = medium, student = yes, credit_rating = fair)

The-Morgan-Kaufmann-Series-in-Data-Management-Systems
-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mining.
Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011

will he buy computer

✓ Yes No

0.028 0.007

Example

3 candidates X, Y, Z

Chances of win 4 : 2 : 3

Event E Introduce computer education

$P(E/A_X) = 0.3$, Probability of X, if selected, introduction of comp. edu.

$P(E/A_Y) = 0.5$ Y

$P(E/A_Z) = 0.8$ Z

What is the probability that there was computer education?

$$P(E) = \sum_{i=1}^{n} P(A_i) \cdot P(E/A_i)$$

$$= \frac{4}{9} \cdot \frac{3}{10} + \frac{2}{9} \cdot \frac{5}{10} + \frac{3}{9} \cdot \frac{8}{10} = \frac{23}{45}$$

Example

3 different machines

M_1	M_2	M_3	
20%	30%	50%	items produced
1%	2%	3%	defective

~~Suppose~~, one item is selected at random and it is found defective.

What is the probability that this item was produced by machine M_3?

Any learning or improvements to business process?

Solution

E The event that the selected item is defective

A_1 selected item was produced by M_1

A_2 " M_2

A_3 " M_3

We need to find

$$P\left(A_3 / E\right).$$

$P(A_1) = 0.2$

$P(A_2) = 0.3$

$P(A_3) = 0.5$

$P(E/A_1) = 0.01$

$P(E/A_2) = 0.02$

$P(E/A_3) = 0.03$

Using Bayes' theorem

$$P(A_3 | E) = \frac{P(A_3) \cdot P(E/A_3)}{\sum_{i=1}^{n} P(A_i) \cdot P(E/A_i)}$$

$$= \frac{0.5 \times 0.03}{(0.2 \times 0.01) + (0.3 \times 0.02) + (0.5 \times 0.03)}$$

$$= 0.652$$

that the probability that the selected defective item was produced by M_3 is 0.65 65%

The machine which generates more defective products shall be given less overall production.

This will reduce the probability of item being defective itself.

see if you can prove this.

Generally speaking

$$P(A) \cdot P(B/_A) = P(B) \cdot P(A|_B)$$

Above is helpful for determining second thing when you are aware of first. that is any one can be likelyhood and hence the other become posteriori.

Without loss of generality,

If you know that person A is in room#3 then only speaking of person B sitting with person A or both are sitting with each other makes more sense.

C_1 and C_2 Priori probabilities.

$$P(\overset{C_1}{\text{buys\_computer}}_{\text{yes}}) = \frac{9}{14} = 0.643$$

$$P(\overset{C_2}{\text{buys\_computer}}_{\text{No}}) = 5/14 = 0.357$$

To compute $P(X/C_i)$ likelyhoods

Given computer_buys_yes, age = ~~youth~~ ? 2/9

income = medium ? 4/9

student = yes ? 6/9

credit_rating = fair ? 6/9

computer_buys_No

Given computer_buys_No,

age = youth ? $3/5$

income = medium ? $2/5$

student = Yes ? $1/5$

credit_rating = fair ? $2/5$

$$P(X \mid buys\_comp = Yes) = \frac{2}{9} \times \frac{4}{9} \times \frac{6}{9} \times \frac{6}{9}$$
$$= 0.044$$

$$P(X \mid buys\_comp\_No) = \frac{3}{5} \times \frac{2}{5} \times \frac{1}{5} \times \frac{2}{5}$$
$$= 0.019$$

Class conditional independence.

[∵ rule of product probab. Independent events]

Posteriori

$$\cancel{P\left(\text{buys\_computer} = \text{"yes"} / X_{tuple}\right)} = \frac{P(\text{buys\_comp} = \text{yes}) * P(X / \text{buys\_comp} = \text{yes})}{P(X)}$$

$$= \frac{0.0643 \times 0.044}{P(X)} = \frac{0.028}{P(X)}$$

$$P\left(\text{buys\_computer} = \text{"No"} / X_{tuple}\right) = \frac{0.357 \times 0.019}{P(X)} = \frac{0.007}{P(X)}$$

It is okey not to find $P(X)$, because we want to just compare.

$0.028 > 0.007$. Hence, tuple X is more likely to ~~buy computer~~ YES

This can also be seen as
filling missing values for given
tuple based on probability concepts.

Here, for tuple X
we figured mathematically
missing value of column computer_buys
as "yes".