Probability

$P(A) = \frac{T^A}{T}$

$P(A)$ - the measure of likelihood that an event A will occur

T^A - all possible results associated with the event A

T - all possible results

Conditional probability

$$P(C|A) = \frac{P(C \cap A)}{P(A)}$$ - conditional probability that a patient has a disease C, if he has symptoms A

$$P(A|C) = \frac{P(A \cap C)}{P(C)}$$ - conditional probability that a patient has symptoms A, if he has a disease C

$P(C \cap A)$ - probability that a patient has a disease C and symptoms A

$P(C)$ - probability that a patient has a disease C

$P(A)$ - probability of symptoms

Bayes theorem

$$P(C|A) = \frac{P(C \cap A)}{P(A)}$$

$$P(A|C) = \frac{P(A \cap C)}{P(C)}$$

$$P(C|A) = \frac{P(A|C) * P(C)}{P(A)}$$

Conditional probability table

Table describing conditional probabilities of diseases, where the given symptom was observed:

	influenza C_1	cold C_2	pneumonia C_3	allergy C_4
headache A_1	$P(C_1\|A_1)$	$P(C_2\|A_1)$	$P(C_3\|A_1)$	$P(C_4\|A_1)$
cough A_2	$P(C_1\|A_2)$	$P(C_2\|A_2)$	$P(C_3\|A_2)$	$P(C_4\|A_2)$
sneeze A_3	$P(C_1\|A_3)$	$P(C_2\|A_3)$	$P(C_3\|A_3)$	$P(C_4\|A_3)$
temperature A_4	$P(C_1\|A_4)$	$P(C_2\|A_4)$	$P(C_3\|A_4)$	$P(C_4\|A_4)$

$$\sum_{i=1}^{n} P(A_i) = 1 \qquad \sum_{j=1}^{m} P(C_j|A_i) = 1 \qquad P(C_j) = \sum_{i=1}^{n} P(A_i) * P(C_j|A_i)$$

$$P(A_i|C_j) = \frac{P(A_i) * P(C_j|A_i)}{P(C_j)} \qquad P(C_j|A_i) = \frac{P(C_j) * P(A_i|C_j)}{P(A_i)}$$

More general Bayes Theorem formula

Bayes theorem has the more general form for many diseases and many symptoms:

$$P(C_j|A_{i1} \cap \ldots \cap A_{ik}) = \frac{P(C_j) * P(A_{i1}|C_j) * \ldots * P(A_{ik}|C_j)}{\sum_{l=1}^{n} P(C_l) * P(A_{i1}|C_l) * \ldots * P(A_{ik}|C_l)}$$

Bayes Theorem: the comparison of equivalent sets and events

Ω - a space of independent elementary observed results; $A \in 2^\Omega \Rightarrow A' \in 2^\Omega$ - complementarity; $A, B \in 2^\Omega \Rightarrow A \cup B \in 2^\Omega$ - additivity	F - the independent rule set such that $a \in F \Leftrightarrow b \notin F - \{\mathbf{0}, a\}$ this means $b \wedge \neg a = \mathbf{0}$
$(2^\Omega, \cup, \cap, ', \Omega, \phi)$	$(F, \vee, \wedge, \neg, \mathbf{1}, \mathbf{0})$
$P(\phi) = 0 \qquad P(\Omega) = 1$	$P(\mathbf{0}) = 0 \qquad P(\mathbf{1}) = 1$
$A \cap A' = \phi \quad A \cup A' = \Omega$	$a \wedge \neg a = \mathbf{0} \quad a \vee \neg a = \mathbf{1}$
$\forall A, B \in 2^\Omega \quad A \cap B = \phi$ $P(A \cup B) = P(A) + P(B)$	$\forall a, b \in F \quad a \wedge b = \mathbf{0}$ $P(a \vee b) = P(a) + P(b)$
$\forall A \in 2^\Omega \quad P(A) + P(A') = 1$	$\forall a \in F \quad P(a) + P(\neg a) = 1$
$A \subseteq B \qquad P(A) \leq P(B)$	$(a \Rightarrow b) = \mathbf{1} \qquad P(a) \leq P(b)$

Bayes rule

$$P(h|e) = \frac{P(e|h)P(h)}{P(e)}$$

where h means hypothesis and e denotes an event.
Such a rule is just an another form of the usual rule:

$$e \Rightarrow h$$

Bayes Theorem

$$\exists\ H = \{h_1, \ldots, h_n\}, \text{ where}$$

$$\forall i \neq j \quad h_i \wedge h_j = \mathbf{0} \quad \bigcup_{i=1}^{n} h_i = \mathbf{1}, \quad P(h_i) > 0, \quad i = 1, \ldots, n$$

$$\exists\ \{e_1, \ldots, e_m\}, \text{ where}$$

$$P(e_1, \ldots, e_m | h_i) = \prod_{j=1}^{m} P(e_j | h_i),\ \ i = 1, \ldots, n \Leftrightarrow$$

$$\Leftrightarrow \forall e_j, h_i \quad e_j \text{ conditionally independent on } h_i$$

$$P(h_i | e_1, \ldots, e_m) = \frac{P(e_1, \ldots, e_m | h_i) P(h_i)}{\sum_{k=1}^{n} P(e_1, \ldots, e_m | h_k) P(h_k)}$$

$$P(h_i | e_1, \ldots, e_m) = \frac{\prod_{j=1}^{m} P(e_j | h_i)}{\sum^{n} \prod^{m} P(e_j | h_k) P(h_k)} P(h_i)$$

PROSPECTOR modifications (1976)

An additional assumption:

$$P(e_1, \dots, e_m | \neg h_i) = \prod_{j=1}^{m} P(e_j | \neg h_i), \ \ i = 1, \dots, n$$

New Bayes rule: $P(\neg h|e) = \dfrac{P(e|\neg h)P(\neg h)}{P(e)}$ or

$$\frac{P(h|e)}{P(\neg h|e)} = \frac{P(e|h)}{P(e|\neg h)} \frac{P(h)}{P(\neg h)}$$

$O(h) = \dfrac{P(h)}{P(\neg h)}$ - a chance a priori

$O(h|e) = \dfrac{P(h|e)}{P(\neg h|e)}$ - a chance a posteriori

A reliability coefficient: $\lambda = \dfrac{P(e|h)}{P(e|\neg h)} \Rightarrow O(h|e) = \lambda O(h)$

Further PROSPECTOR modifications

$$\text{In a general case: } O(h_i|e_1,\ldots,e_m) = O(h_i)\prod_{k=1}^{m}\lambda_{k_i},$$

$$\text{where } \lambda_{k_i} = \frac{P(e_k|h_i)}{P(e_k|\neg h_i)}$$

$$\overline{\lambda} = \frac{P(\neg e|h)}{P(\neg e|\neg h)} \Rightarrow O(h|\neg e) = \overline{\lambda}O(h)$$

Coefficients λ i $\overline{\lambda}$ are defined a priori. λ denotes observation sufficiency e (especially for $\lambda \gg 1$) and $\overline{\lambda}$ denotes necessity e (especially for $0 \leq \overline{\lambda} \leq 1$).

Bayes model disadvantages

- Assumptions are not accomplished.
- Ignorance is hidden in a priori probabilities.
- Probabilities are known only for elementary observed independently events, but not for their sets.
- Probabilities are for both negative and positive events at the same time.

Naive Bayes classifier assumptions

- Each instance x described by attribute values $a(x) = \langle a_1(x), a_2(x) \ldots a_n(x) \rangle$, where $a_i(x)$ is the given value of the attribute a_i $\left(a_i(x) \in \{a_{ij}\},\ j \in (1 \ldots A_i)\right)$.
- Attribute values $a_i(x)$ of instances x are conditionally independent given the target class C_k.
- It is so called Naive Bayes assumption:
  $P(a(x)|C_k) = \prod_i P(a_i(x)|C_k)$
  which is usually not true, but incorrect class probabilities very often permit correct classification.
- Conditional probabilities of attribute values $a_i(x)$ given the class C_k are $P(a_i(x)|C_k) = P_{T^{C_k}}(a_i(x)) = \frac{|T^{C_k}_{a_i(x)}|}{|T^{C_k}|}$.
- $P(C_k|a(x)) = \dfrac{P(C_k) \prod_i P(a_i(x)|C_k)}{\sum_{C_l \in C} P(C_l) \prod_i P(a_i(x)|C_l)}$

- The final Naive Bayes classifier hypothesis $h(x)$ predicting the correct class is just the greatest conditional probability:
  $$P(C_k|a(x)) = \frac{P(C_k)P(a(x)|C_k)}{\sum_{C_l \in C} P(C_l)P(a(x)|C_l)}$$
- $$P(C_k|a(x)) = \frac{P(C_k)\prod\limits_i P(a_i(x)|C_k)}{\sum\limits_{C_l \in C} P(C_l)\prod\limits_i P(a_i(x)|C_l)}$$
- $$h(x) = \arg\max_{C_k \in C} P(C_k|a(x))$$
- In a case of not present values in training instances to prevent prediction errors the number of values A_i of the attribute a_i is added to conditional probability:
  $$P(a_i(x)|C_k) = P_{T^{C_k}}(a_i(x)) = \frac{|T^{C_k}_{a_i(x)}|+1}{|T^{C_k}|+A_i}.$$