Bayes theorem

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

→ Conditional probability

* The probability of the occurrence of an event A given that an event B has already occurred is called the conditional probability of A given B.

* Denoted by $P(A|B)$

* $P(A|B) = \frac{P(A\cap B)}{P(B)}$ if $P(B) \neq 0$

independent event
$P(A\cap B) = P(A) \cdot P(B)$.
$P(A|B) = P(A)$

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

$$P(A\cap B) = P(B\cap A)$$

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

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

Examples

Problem 1 : Consider a set of patients coming for treatment in a certain clinic. Let A denote the event that a "Patient has liver disease" and B the event that a "Patient is an alcoholic". It is known from experience that 10% of the patients entering the clinic have liver disease and 5% of the patients are alcoholic. Also among those patients diagnosed with liver disease, 7% are alcoholics. Given that a patient is alcoholic, what is the poobability that he will have liver disease ?

Ans)

A – Patient has liver disease.

B – Patient is an alchoholic.

$P(A) = 10\% = 0.1$

$P(B) = 5\% = 0.05$

$P(B|A) = 7\% = 0.07$

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

$$= \frac{0.07 \times 0.1}{0.05} = \underline{\underline{0.14}}$$

Problem 2: Three factories A, B, C of an electric bulb manufacturing company produce respectively 35%, 35% and 30% of the total output. Approximately 1.5%, 1% and 2% of the bulbs produced by these factories are known to be defective. If a randomly selected bulb manufactured by the company was found to be defective, What is the probability that the bulb was manufactures in factory A?

Ans) A – Manufacturing a bulb from factory A

B – Manufacturin a bulb from factory B

C – Manufacturi a bulb from factory C.

D – A bulb is defective.

$$P(A) = 35\% = .35$$
$$P(B) = 35\% = .35$$
$$P(C) = 30\% = .3$$
$$P(D|A) = 1.5\% = .015$$
$$P(D|B) = 1\% = 0.01$$
$$P(D|C) = 2\% = 0.02$$

$$P(A|D) = \frac{P(D|A) \cdot P(A)}{P(D|A)P(A) + P(D|B)P(B) + P(D|C)P(C)}$$

$$= \frac{.015 \times .35}{.015 \times .35 + .01 \times .35 + .02 \times .3} = \underline{0.356}$$

Naive Bayes Algorithm
(Bayesian Classifier)

Bayes theorem

Assumption

All features are independent - The presence or absence of a feature does not influence the presence or absence of any other feature.

f_1 f_2

$p(f_1|c)$
$p(f_2|c)$

Class conditional independence - It means that if the class is known, knowing one feature does not give additional ability to predict another feature.

How to choose the most appropriate class label of an instance, $X = (x_1, x_2, \ldots x_n)$?

Suppose we have a training set consisting of N examples having n features. Let the features be named as $(F_1, F_2, \ldots, F_n)$ and feature vector is of the form $(f_1, f_2 \ldots f_n)$. A class label is associated with each example. Let the class labels are $\{c_1, c_2 \ldots c_p\}$.

Let our instance be $X = (x_1, x_2, \ldots x_n)$

To determine the most appropriate class label, we compute the following conditional probabilities:

$$P(c_1|X),\ P(c_2|X),\ \cdots,\ P(c_p|X)$$

and choose the maximum among them. Let the maximum probability be $P(c_i|X)$, then we choose c_i as the most appropriate class label for X.

How to find the above conditional probabilities.?

We apply bayes theorem here:

$$\text{i.e } P(c_k|X) = \frac{P(X|c_k)\,P(c_k)}{P(X)}$$

X is given as $X = (x_1, x_2, \cdots, x_n)$.

So we can write

$$P(c_k|X) = \frac{P(x_1|c_k)\,P(x_2|c_k)\cdots\cdots P(x_n|c_k)\,P(c_k)}{P(x_1)\,P(x_2)\cdots\cdots P(x_n)}$$

Since the denominator $P(x_1)P(x_2)\cdots P(x_n)$ is independent of the class labels, we have

$$P(c_k|X) \propto P(x_1|c_k)\,P(x_2|c_k)\cdots P(x_n|c_k)\cdot P(c_k)$$

Hence we need to find the value $P(x_1|c_k)\,P(x_2|c_k)\cdots P(x_n|c_k)\cdot P(c_k)$, $k = 1, 2, \cdots p$.

where $P(C_k) = \dfrac{\text{No. of examples with class label } C_k}{\text{Total no. of examples.}}$

$$P(x_j|C_k) = \frac{\text{No. of examples with } j^{th} \text{ feature equal to } x_j \text{ and class label } C_k}{\text{No. of examples with class label } C_k}.$$

Naive bayes Algorithm

Let the training set having n features $F_1, F_2, \ldots F_n$ and feature vector $(f_1, f_2, \cdots f_n)$. Let the class labels $\{C_1, C_2, \ldots C_p\}$. and Let the instance having feature vectors $X = (x_1, x_2 \ldots x_n)$.

Step 1: Compute the probabilities $P(C_k)$ for $k = 1, 2, \ldots p$.

Step 2: Form a table showing the conditional probabilities. $P(f_1|C_k), P(f_2|C_k) \ldots P(f_n|C_k)$ for all values of $f_1, f_2 \ldots f_n$ and for $k = 1, 2, \ldots P$.

Step 3: Compute the products

$$q_k = P(x_1|C_k)\, P(x_2|C_k) \ldots P(x_n|C_k)\, P(C_k)$$

for $K = 1, 2, \ldots P$.

Step 4: Find j such $q_j = \max(q_1, q_2, \ldots q_p)$

Step 5: Assign the class label C_j to the instance X.

Problem 1

Training set

Sl.No	Swim	Fly	Crawl	Class label
1	Fast	No	No	Fish
2	Fast	No	Yes	Animal
3	Slow	No	No	Animal
4	Fast	No	No	Animal
5	No	Short	No	Bird
6	No	Short	No	Bird
7	No	Rarely	No	Animal
8	Slow	No	Yes	Animal
9	Slow	No	No	Fish
10	Slow	No	Yes	Fish
11	No	Long	No	Bird
12	Fast	No	No	Bird

Let instance X = (Slow, Rarely, No).
Which class label is most appropriate for X ? [Use Bayesian Classifier]

Ans) Construction of frequency table

	Swim			Fly				Crawl		Total
	Fast	Slow	No	Short	Long	Rarely	No	Yes	No	
Animal	2	2	1	0	0	1	4	2	3	5
Bird	1	0	3	2	1	0	1	0	4	4
Fish	1	2	0	0	0	0	3	1	2	3

Step 1: Computation of probability of each class (C_k), $k = 1, 2, 3$.

$$P(C_k) = \frac{\text{No. of samples with class label } C_k}{\text{Total no. of samples.}}$$

$$P(\text{Animal}) = \frac{5}{12}$$

$$P(\text{Bird}) = \frac{4}{12}$$

$$P(\text{Fish}) = \frac{3}{12}$$

Step 2: Construction of conditional probability table.

$$P(f_j | C_k) = \frac{P(f_j \cap C_k)}{P(C_k)}$$

$$= \frac{\text{No. of samples with } f_j \text{ \& } C_k / \text{Total No. of samples}}{\text{No. of samples with } C_k / \text{Total no. of samples.}}$$

$$= \frac{\text{No. of samples with } f_j \text{ \& } C_k}{\text{No. of samples with } C_k.}$$

Class	Swim			Fly				Crawl	
	Fast	Slow	No	Short	Long	Rarely	No	Yes	No
Animal	2/5	2/5	1/5	0/5	0/5	1/5	4/5	2/5	3/5
Bird	1/4	0/4	3/4	2/4	1/4	0/4	1/4	0/4	4/4
Fish	1/3	2/3	0/3	0/3	0/3	0/3	3/3	1/3	2/3

Step 3: Computation of conditional probability $P(C_k | X)$, $k = 1, 2, 3$

$$X = (\text{Slow, Rarely, No})$$

$$P(\text{Animal} | X) = P(\overset{\text{swim}}{\text{Slow}} | \text{Animal}) \cdot P(\overset{\text{Fly}}{\text{Rarely}} | \text{Animal}) \cdot P(\overset{\text{crawl}}{\text{No}} | \text{Animal}) \cdot P(\text{Animal})$$

$$= \frac{2}{5} \times \frac{1}{5} \times \frac{3}{5} \times \frac{5}{12} = \underline{0.02}$$

$$P(Fish|X) = P\left(\begin{smallmatrix}swim\\slow\end{smallmatrix}\middle|Fish\right) P\left(\begin{smallmatrix}Fly\\rarely\end{smallmatrix}\middle|Fish\right)$$

$$P\left(\begin{smallmatrix}crawl\\No\end{smallmatrix}\middle|Fish\right) \cdot P(Fish)$$

$$= \frac{2}{3} \cdot \frac{0}{3} \cdot \frac{2}{3} \cdot \frac{3}{12}$$

$$= \underline{\underline{0}}$$

$$P(Bird|X) = P\left(\begin{smallmatrix}swim\\slow\end{smallmatrix}\middle|Bird\right) P\left(\begin{smallmatrix}Fly\\rarely\end{smallmatrix}\middle|Bird\right)$$

$$\cdot P\left(\begin{smallmatrix}crawl\\No\end{smallmatrix}\middle|Bird\right) P(Bird)$$

$$= \frac{0}{4} \cdot \frac{0}{4} \cdot \frac{4}{4} \cdot \frac{4}{12}$$

$$= \underline{\underline{0}}$$

Step 4: Maximum probability is

$$P(Animal|X) = 0.02$$

$\underline{\underline{\text{Animal}}} \longrightarrow X$ (slow, Rarely, No)