Bayes Classifier

Conditional Probability

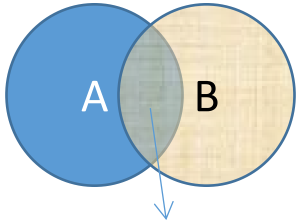
- Let A & B are 2 events of a random experiment and P(B) ≠ 0 i.e., n(B) ≠ 0

  n be the total number of elementary events in the experiments Let **n(A)** is no. of events favorable to A

  n(B) is no. of events favorable to B, out of which n (A∩B) is favorable to A also.

  n (A ∩ B) / n(B) is the proportion of elementary events that are favorable to A, among the elementary events favorable to B and is called conditional probability of A, given B has already occurred and is denoted by **P(A|B)**.

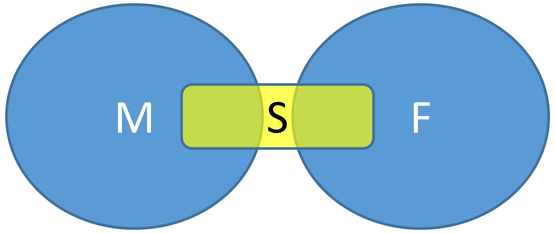
  So, **P(A|B) = n (A ∩ B) /n(B)**
  = P (A ∩ B) /P(B)



$(A \cap B)$

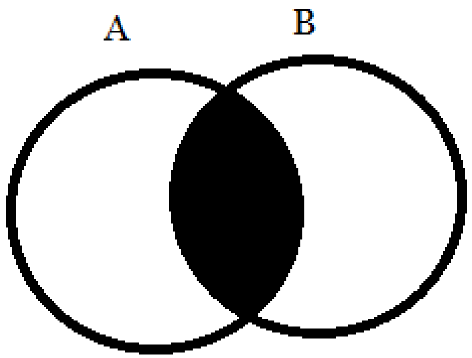
Example

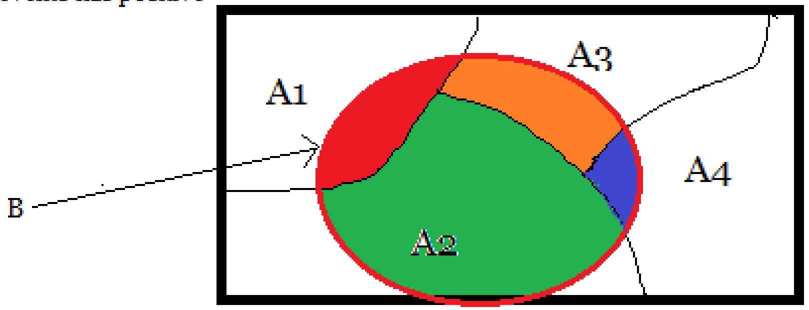
- In a group of 20 males and 5 females, 10 males and 3 females are service holders. What is the probability that a person selected at random from the group is a service holder given that the selected person is a male.
- Let A = person is a service holder and B = person is male
- We have to find out P(A|B) = $P(A \cap B)/P(B)$



- A ∩ B means selected person is male and service holder. So no of cases favorable to A ∩ B = 10
- P(A ∩ B) = 10/25 = 1/5 P(B) = 20/25 = 4/5
- P(A|B) = (1/5)/(4/5) = 1/2

Let A1, A2, ….,An be exhaustive and mutually exclusive and such that each of these events has positive probability. Then, for any event B





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

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

$P(A|B)^* \, P(B) = P(B|A)^* \, P(A)$ …………………… (i)

$P(A|B) = P(B|A)^* \, P(A) \;/\; P(B)$ ……………………(ii) *

$P(B) = P(A_1 \wedge B) + P(A_2 \wedge B) + P(A_3 \wedge B) + P(A_4 \wedge B)$ *

$$= \sum_{i=1}^{n} P(Ai \cap B) = \sum_{i=1}^{n} P(Ai)P(B \mid Ai) \quad \text{…………………(iii)}$$

$$P(Ai \mid B) = \frac{P(Ai)P(B \mid Ai)}{\sum_{j=1}^{n} P(Aj)P(B \mid Aj)}$$

Example

- The probability of X, Y, Z becoming the principal of a certain college are respectively .3, .5 and .2. The probability that student-aid-fund will be introduced in the college if X, Y, Z become principal is .4, .6 and .1 respectively. Given student-aid-fund is introduced, find the probability Y has been appointed as principal?

  Let A1, A2 and A2 are events of appoint of X, Y , Z as principal respectively. So, P(A1) = .3, P(A2) =.5 and P(A3) = .2. Let B be the event of getting student fund. So. P(B|A1)=.4 P(B|A2)=.6 and P(B|A3)=.1 and

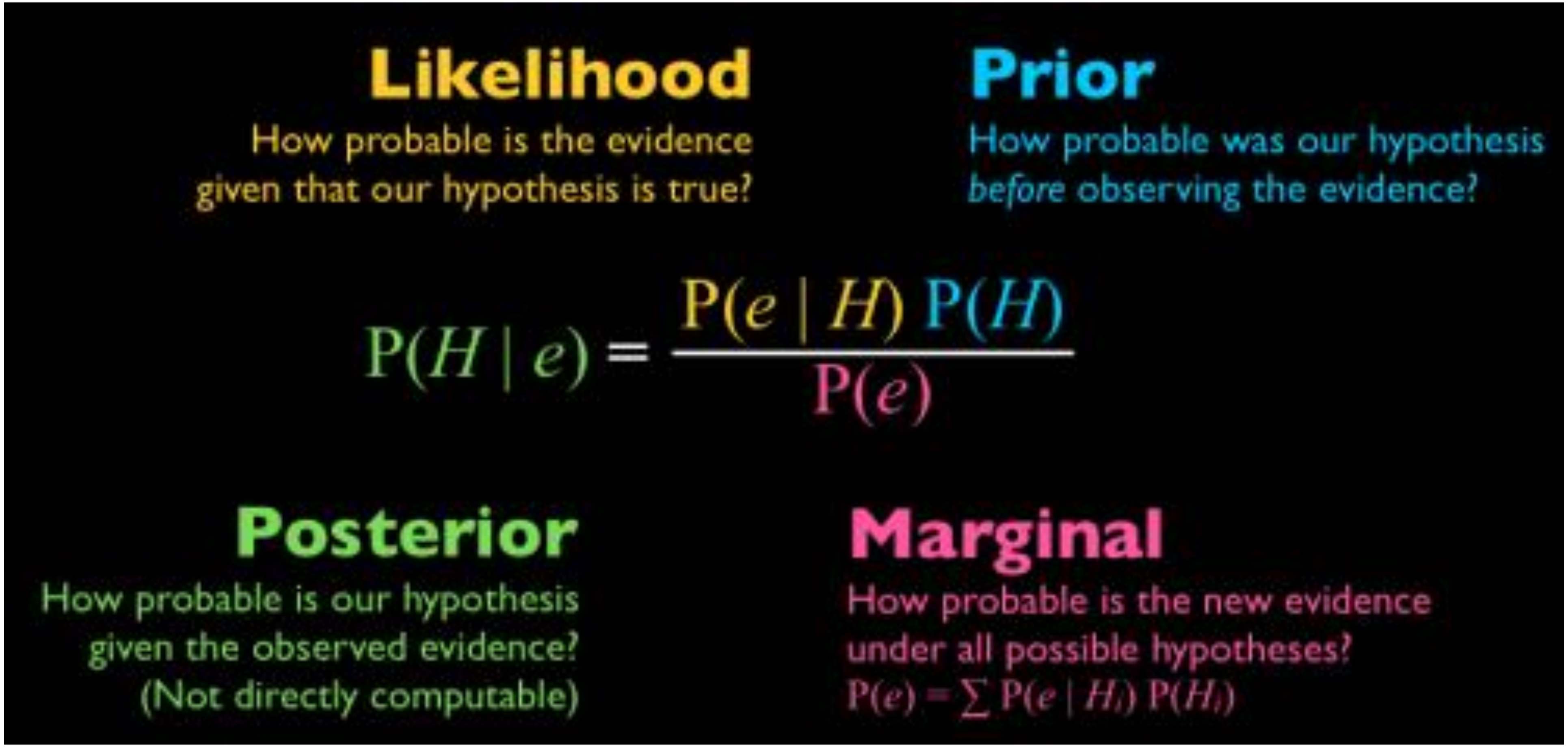
  we have to find out P(A2|B)

  - By Bayes theorem,

    $$P(Ai \mid B) = \frac{P(Ai)P(B \mid Ai)}{\sum_{j=1}^{n} P(Aj)P(B \mid Aj)}$$

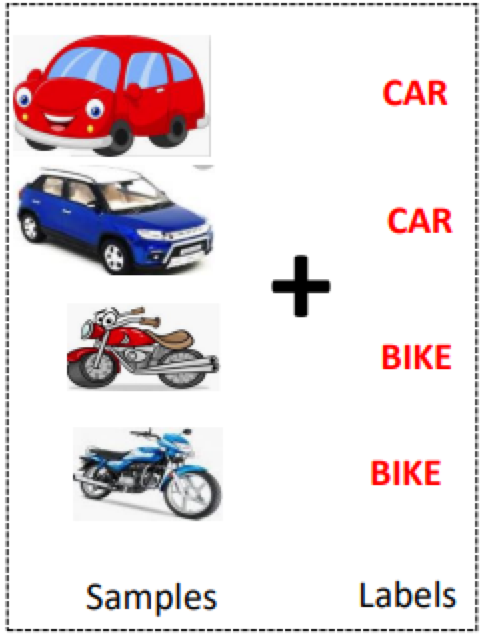
  - P(A2|B)= $\dfrac{P(A2)P(B \mid A2)}{\sum_{j=1}^{3} P(Aj)P(B \mid Aj)}$

    $$= \frac{.5*.6}{.3*.4+.5*.6+.2*.1} = \frac{.3}{.44}$$



Hence, Bayes Theorem can be written as: **posterior = likelihood * prior / evidence**

Classification



Training Dataset

f () = CAR/BIKE

Given a dataset $\boldsymbol{D = \{ x_1, x_2\, x_3 \dots x_n \}}$ and set of class labels $\boldsymbol{C = \{ c_1\, c_2\, c_3 \dots c_k \}}$, the task of classification to devise a mapping function $\boldsymbol{f : D \rightarrow C}$.

Bayesian Classifier

#Wheel	Height	Class Label
4	H	CAR
4	H	CAR
4	H	CAR
2	L	BIKE
2	L	BIKE
2	L	BIKE
4	L	BIKE
2	H	CAR

Pr(CAR| 4,H) = 100%

Pr(BIKE| 4,L) = 100%

Pr(CAR| 2,H) = 100%

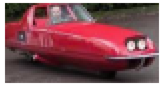
Pr(BIKE| 2,L) = 100%

Pr(CAR| 4,L) = 0%

Pr(BIKE|4,H) = 0%

Pr(CAR| 2,L) = 0%

Pr(BIKE| 2,H) = 0%

 {2 H} ?

$Pr(c_i|x), \forall c_i \epsilon C$

$$class = \arg\max_{c_i} \Pr(c_i|x)$$

Bayesian Classifier

	#Wheel	Height	Class Label
	4	H	CAR
	4	H	CAR
	4	H	CAR
	2	L	BIKE
	2	L	BIKE
	2	L	BIKE
	4	L	BIKE
	2	H	CAR

Pr(CAR| 4,H) = 100%

Pr(BIKE| 4,L) = 100%

Pr(CAR| 2,H) = 100%

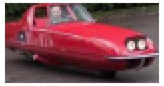
Pr(BIKE| 2,L) = 100%

Pr(CAR| 4,L) = 0%

Pr(BIKE|4,H) = 0%

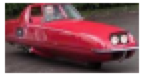
Pr(CAR| 2,L) = 0%

Pr(BIKE| 2,H) = 0%

 {2 H} ?

$Pr(c_i|x), \forall c_i \epsilon C$

$$\textbf{\textit{class}} = \arg\max_{c_i} \Pr(c_i|x)$$

$Pr(CAR \,|$  $)$

$Pr(CAR\,|\{2, H\}) = 1$

$Pr(BIKE\,|\{2, H\}) = 0$

$\textbf{\textit{class}}$ = CAR

Bayes Rule

Likelihood

Prior

$$\Pr(c_i|x) = \frac{\Pr(c_i, x)}{\Pr(x)} = \frac{\Pr(x|c_i)\Pr(c_i)}{\Pr(x)}$$

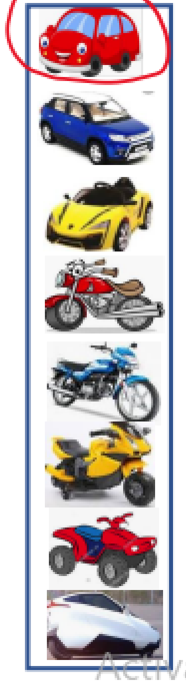
$$= \frac{\Pr(x|c_i)\Pr(c_i)}{\Pr(x|c_1)\Pr(c_1) + \Pr(x|c_2)\Pr(c_2) + \ \ldots + \Pr(x|c_k)\Pr(c_k)}$$

Marginalization

#Wheel	Height	Class Label
4	H	CAR
4	H	CAR
4	H	CAR
2	L	BIKE
2	L	BIKE
2	L	BIKE
4	L	BIKE
2	H	CAR

Bayesian Classifier

$$\Pr(c_i|x) = \Pr(c_i \mid \{w_1, w_2 w_3 \dots w_k\}) = \frac{\Pr(\{w_1, w_2 w_3 \dots w_k\}|c_i)\Pr(c_i)}{\Pr(\{w_1, w_2 w_3 \dots w_k\})}$$



#Wheel	Height	Class Label
4	H	CAR
4	H	CAR
4	H	CAR
2	L	BIKE
2	L	BIKE
2	L	BIKE
4	L	BIKE
2	H	CAR

$\Pr(CAR|$  $)$

$$= \Pr(CAR \mid \{4, H\}) = \frac{\Pr(\{4, H\}|CAR)\Pr(CAR)}{\Pr(\{4, H\})}$$

$$= \frac{0.75 \times 0.5}{0.375} = 1$$

$\Pr(BIKE|$  $)$

$$= \Pr(BIKE \mid \{4, H\}) = \frac{\Pr(\{4, H\}|BIKE)\Pr(BIKE)}{\Pr(\{4, H\})}$$

$$= \frac{0 \times 0.5}{0.375} = 0$$

Ans: CAR

Bayesian Classifier

$$\Pr(c_i|x) = \Pr(c_i \mid \{w_1, w_2 w_3 \dots w_k\}) = \frac{\Pr(\{w_1, w_2 w_3 \dots w_k\}|c_i)\Pr(c_i)}{\Pr(\{w_1, w_2 w_3 \dots w_k\})}$$

$\Pr(CAR|$  $)$

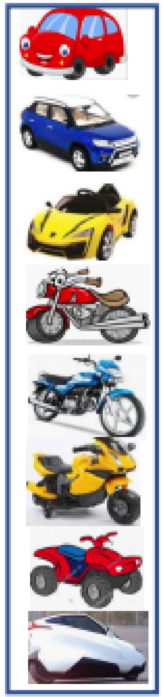
$= \Pr(CAR \mid \{4, H\})$

$$= \frac{\Pr(\{4, H\}|CAR)\Pr(CAR)}{\Pr(\{4, H\})}$$

$\Pr(BIKE|$  $)$

$= \Pr(BIKE \mid \{4, H\})$

$$= \frac{\Pr(\{4, H\}|BIKE)\Pr(BIKE)}{\Pr(\{4, H\})}$$



#Wheel	Height	Class Label
4	H	CAR
4	H	CAR
4	H	CAR
2	L	BIKE
2	L	BIKE
2	L	BIKE
4	L	BIKE
2	H	CAR

Bayesian Classifier

$$\Pr(c_i|x) \;= \Pr(c_i \mid \{w_1, w_2 w_3 \ldots w_k\}) \;= \frac{\Pr(\{w_1, w_2 w_3 \ldots w_k\}|c_i)\Pr(c_i)}{\Pr(\{w_1, w_2 w_3 \ldots w_k\})}$$

$\Pr(CAR|$  $)$

$= \Pr(CAR \mid \{4, H\})$

$\sim \Pr(\{4, H\}|CAR)\Pr(CAR)$

$\Pr(BIKE|$  $)$

$= \Pr(BIKE \mid \{4, H\})$

$\sim \Pr(\{4, H\}|BIKE)\Pr(BIKE)$

#Wheel	Height	Class Label
4	H	CAR
4	H	CAR
4	H	CAR
2	L	BIKE
2	L	BIKE
2	L	BIKE
4	L	BIKE
2	H	CAR

Bayesian Classifier

$$\Pr(c_i|x) \;\; = \Pr(c_i \mid \{w_1, w_2, w_3 \dots w_k\}) = \frac{\Pr(\{w_1, w_{2,} w_3 \dots w_k\}|c_i)\Pr(c_i)}{\Pr(\{w_1, w_2, w_3 \dots w_k\})}$$

If ***k*** (the number of classes) is **small,**

estimating **likelihood** $\Pr(\{w_1, w_{2,} w_3 \dots w_k\}|c_i)$ is **feasible**.

However, if ***k*** (the number of classes) is **very large,**

estimating **likelihood** $\Pr(\{w_1, w_{2,} w_3 \dots w_k\}|c_i)$ is **a very expensive task** over **a large dataset.**

#Wheel	Height	Class Label
4	H	CAR
4	H	CAR
4	H	CAR
2	L	BIKE
2	L	BIKE
2	L	BIKE
4	L	BIKE
2	H	CAR

$$\Pr(\{w_1, w_{2,} w_3 \dots w_k\}|c_i) \quad = \Pr(w_1|w_2, w_3, \dots, w_3, c_i). Pr(w_2|w_3, w_4, \dots, w_3, c_i) \dots . Pr(w_k|c_i)$$

Naïve Bayes Classifier

$$\Pr(c_i|x) \;\; = \Pr(c_i \mid \{w_1, w_2, w_3 \dots w_k\}) = \frac{\Pr(\{w_1, w_2, w_3 \dots w_k\}|c_i) \Pr(c_i)}{\Pr(\{w_1, w_2, w_3 \dots w_k\})}$$

To simplify the estimation, we make an **assumption**

- The features are **conditionally independent**.

Bayesian: $\Pr(\{w_1, w_2, w_3 \dots w_k\}|c_i) \;\; = \Pr(w_1|w_2, w_3, \dots, w_3, c_i). Pr(w_2|w_3, w_4, \dots, w_3, c_i) \dots . Pr(w_k|c_i)$

Naïve Bayes: $\Pr(\{w_1, w_2, w_3 \dots w_k\}|c_i) \;\; \sim \Pr(w_1|c_i). Pr(w_2|c_i) \dots . Pr(w_k|c_i) = \prod_{j=1}^{k} \Pr(w_j|c_i)$

Naïve Bayes Classifier

Now, with regards to our dataset, we can apply Bayes' theorem in following way:

$$P(y|X) = \frac{P(X|y)P(y)}{P(X)}$$

where, y is class variable and X is a dependent feature vector (of size *n*) where:

$$X = (x_1, x_2, x_3, ..., x_n)$$

Now, its time to put a naive assumption to the Bayes' theorem, which is, **independence** among the features. So now, we split **evidence** into the independent parts.

Now, if any two events A and B are independent, then,

```
P(A,B) = P(A)P(B)
```

Hence, we reach to the result:

$$P(y|x_1, ..., x_n) = \frac{P(x_1|y)P(x_2|y)...P(x_n|y)P(y)}{P(x_1)P(x_2)...P(x_n)}$$

which can be expressed as:

$$P(y|x_1, ..., x_n) = \frac{P(y)\prod_{i=1}^{n} P(x_i|y)}{P(x_1)P(x_2)...P(x_n)}$$

Now, as the denominator remains constant for a given input, we can remove that term:

$$P(y|x_1, ..., x_n) \propto P(y)\prod_{i=1}^{n} P(x_i|y)$$

Now, we need to create a classifier model. For this, we find the probability of given set of inputs for all possible values of the class variable *y* and pick up the output with maximum probability. This can be expressed mathematically as:

$$y = argmax_y P(y)\prod_{i=1}^{n} P(x_i|y)$$

So, finally, we are left with the task of calculating P(y) and P(x_i | y).

Naïve Bayes Classifier

$$\Pr(c_i|x) \;\; = \Pr(c_i \mid \{w_1, w_2, w_3 \dots w_k\}) = \frac{\Pr(\{w_1, w_2, w_3 \dots w_k\}|c_i)\Pr(c_i)}{\Pr(\{w_1, w_2, w_3 \dots w_k\})}$$

$$\sim \prod_{j=1}^{k} \Pr(w_j|c_i)\Pr(c_i)$$

#Wheel	Height	Class Label
4	H	CAR
4	H	CAR
4	H	CAR
2	L	BIKE
2	L	BIKE
2	L	BIKE
4	L	BIKE
2	H	CAR

$$\Pr(CAR|\{4, H\}) = \Pr(4|CAR) \times \Pr(H|CAR) \times \Pr(CAR)$$

$$= 0.75 \times 1 \times 0.5 = 0.375$$

$$\Pr(BIKE|\{4, H\}) = \Pr(4|BIKE) \times \Pr(H|BIKE) \times \Pr(BIKE)$$

$$= 0.25 \times 0 \times 0.5 = 0$$