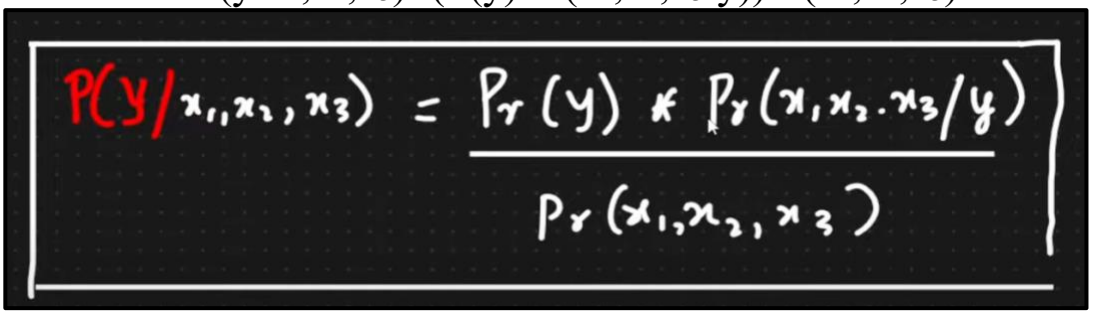
Naïve Bayes

- It is a machine learning algorithm used for **classification problems**.
- There are two type of events:
  - **Independent events**
    - **Example**: Rolling a dice or tossing a coin
    - Probability of getting a number is 1/6
    - Probability of getting a H is 1/2
    - One event doesn't impact the other event
  - **Dependent events**
    - **Example:** Bag of marbles with **3 red** and **2 green** marbles.
    - In 1st instance P(R)=3/5 (prob. of red marble)
    - **P(R and G)=P(R) * P(G/R)**
    - Above P(G/R) is probability of green given red has occurred this is conditional probability.
- Naïve Bayes work on **bayes theorem** which is as follows:
  - **P(B/A)=(P(B)*P(A/B))/P(A)**
- Let's say we have 3 independent features (x_1, x_2, x_3…) and a dependent variable (y) so we apply the
- Now how we will use the above formula:
  - $P(y/x_1,x_2,x_3)=(P(y)*P(x_1,x_2,x_3/y))/P(x_1,x_2,x_3)$



- $P(y/ x_1,x_2,x_{3...})=P(y)*P(x_1/y)* P(y)*P(x_2/y)….P(x_n/y)/(P(x_1)* P(x_2)* P(x_3)…)$
- So, in classification we calculate one probability based on **yes** and other on **no.**
- We need to predict whether a person will play tennis or not given on a given day based on the features.

- **Features**:
  - Day-D1,D2…..D14
  - **Outlook**-Sunny, Overcast and Rain
  - Temperature-Hot, Mild and Cold
  - Humidity-High and Normal
  - Wind: Weak and Strong
  - Play Tennis-Yes or No
- Let's consider a smaller table for **outlook** with **play tennis**

	Yes	No	P(Yes)	P(No)
Sunny	2	3	2/9	3/5
Overcast	4	0	4/9	0/5
Rain	3	2	3/9	2/5

- **P(Sunny/Yes)=2/9** this probability of sunny given yes
- Let's consider a smaller table for **temperature** with **play tennis**

	Yes	No	P(Yes)	P(No)
Hot	2	2	2/9	2/5
Mild	4	2	4/9	2/5
Cold	3	1	3/9	1/5

- **Play** feature

	Yes	No	P(Yes)	P(No)
Yes	9		9/14	
No		5		5/14

- Now if we have to find what is the probability that a person will tennis if the weather is sunny and hot?
  - P(yes/sunny, hot)=p(yes)*p(sunny/yes)*p(hot/yes)
    = 9/14*2/9*2/9
    = 2/63=0.031
  - P(no/sunny, hot)=p(no)*p(sunny/no)*p(hot/no)
    = 5/14*3/5*2/5
    = 3/35=0.085

- **Percentage**:
  - P(yes/sunny, hot) = 0.031/(0.031+0.085) =27%
  - P(no/sunny, hot) = 0.085/(0.031+0.085) =73%
- So, if the outlook is sunny and temperature is hot there is a 73% chance that a person will not play tennis

- As per the algorithm with respect to our output feature we take into consideration our input features and find out the percentage associated with it.