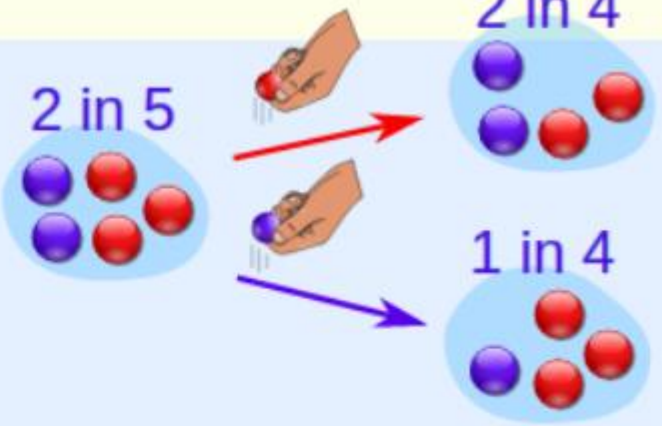
Naïve Bayes


MATCH PROBABILITY .99
PROCEED TO
VISUAL:
MALE HT 0601
WT 0212
ENDOMORPH
ANALYSIS:
SCAN LEVEL
ANALYSIS: MATCH:
389 VEHI 55378
690 SIZE 38022
600 TSPD 23022
287 HPWR 12048
105 CODE 20673
798 RNGE 29687
VISUAL:
MODEL 956
HARLEY-DAVIDSON
CSM-101 UNIT 2012
CPU - Status/ 3465
RAM - Status/ 7458
OBJECT Details
DVD/VCR DUMB HUMAN UNIT
SCAN MODE 4
ASSESSING ROOM...
GATHERING DATA...
DOOR: AJAR
THREAD HANGING FROM DOOR
GATHERING DATA...
PROCESSING...

Probability

- Probability of an event happening = Number of ways it can happen/Total number of outcomes
- Independent Events - Events are not affected by previous events.
- Dependent Events - where what happens depends on what happened before, such as taking cards from a deck makes less cards each time.

Example: Marbles in a Bag



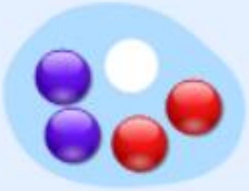
2 blue and 3 red marbles are in a bag.

What are the chances of getting a blue marble?

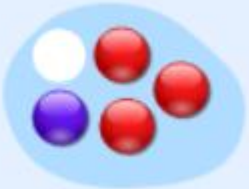
The chance is **2 in 5**

But after taking one out the chances change!

So the next time:



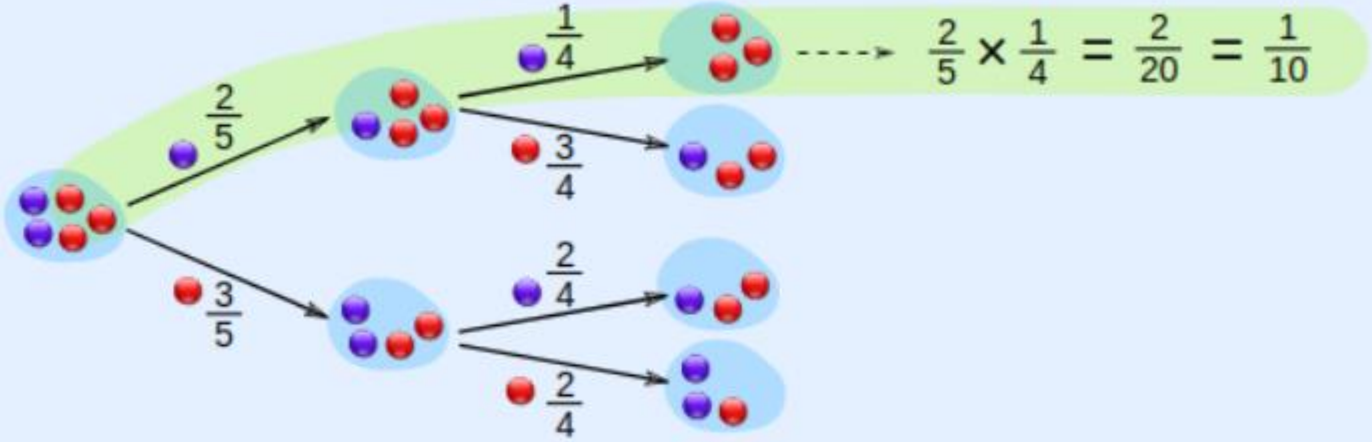
if we got a **red** marble before, then the chance of a blue marble next is **2 in 4**



if we got a **blue** marble before, then the chance of a blue marble next is **1 in 4**

Now we can answer questions like **"What are the chances of drawing 2 blue marbles?"**

Answer: it is a **2/5 chance** followed by a **1/4 chance**:



Did you see how we multiplied the chances? And got 1/10 as a result.

The chances of drawing 2 blue marbles is 1/10

Conditional Probability

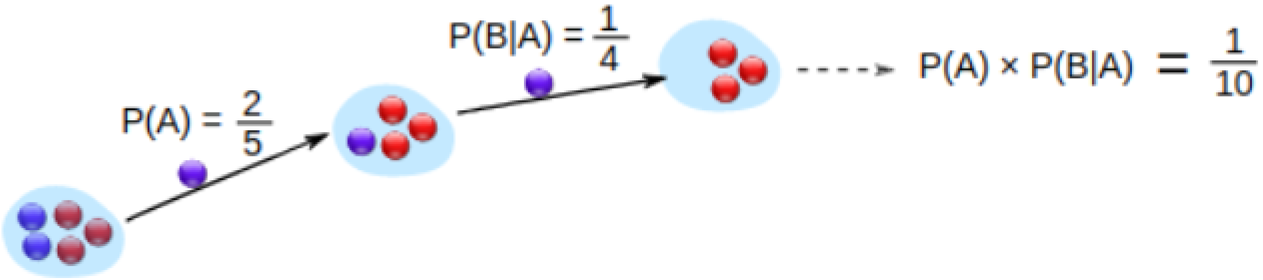
- P(A) means "Probability Of Event A"
- In our marbles example Event A is "get a Blue Marble first" with a probability of 2/5:
- P(A) = 2/5
- And Event B is "get a Blue Marble second" ... but for that we have 2 choices:
    - If we got a Blue Marble first the chance is now 1/4
    - If we got a Red Marble first the chance is now 2/4
- So we have to say which one we want, and use the symbol "|" to mean "given":
- P(B|A) means "Event B given Event A"
- In other words, event A has already happened, now what is the chance of event B?

P(B|A) is also called the "Conditional Probability" of B given A.

And in our case:

$$P(B|A) = 1/4$$

So the probability of getting **2 blue marbles** is:



Naïve Bayes

Its Naive(innocent) because it assumes that all the features are independent of each other. Which is almost never possible.

- Easy to understand.
- All features are independent.
- All impact results equally.
- Need small amount of data to train the model.
- Fast – up to 100X faster.
- It is highly scalable.
- It can make probabilistic predictions.
- It's simple & out-performs many sophisticated methods.
- Stable to data changes.

Bayes Theorem

- It describes the probability of an event, based on prior knowledge of conditions that might be related to the event

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

Bayes Theorem Explained

Suppose:

- Fact_1 = 200 cars/day
- Fact_2 = 300 cars/day
- Out of all Cars produced: 2% are faulty/having issue Out of these faulty cars 50% came from each Factory.

- Question: What is the probability that a car manufactured by Fact_1 is faulty? P(Faulty | Fact_1)?

Bayes Theorem Explained

Solution:

- Car Manufactured by Factory 1: P(Fact_1) = 200/200+300=0.4
- Car Manufactured by Factory 2: P(Fact_2) = 300/200+300=0.6

- 2% of cars are Faulty = P(Faulty) = 0.02

- Probability of a Faulty Car coming out of Factory 1: P(Fact_1 | Faulty) = 0.5
- Probability of a Faulty Car coming out of Factory 2: P(Fact_2 | Faulty) = 0.5

- P(Faulty | Fact_1) = $P(Fact\_1|Faulty)*P(Faulty)/P(Fact\_1)$=0.5*0.02/0.4 = 2.5%

Bayes Theorem Explained

Example:

- Total 500 Cars
- Fact_1 = 200
- Fact_2 = 300
- Faulty = 10
- 50% came from Fact_1 = 5
- % of Faulty Cars came from Fact_1 = 5/200 = 2.5%

Types of Naïve Bayes

- Gaussian Naive Bayes - Feature columns are normal distribution
- Multinomial Naive bayes - Feature columns are counters
- Bernouli's Naive bayes - Feature columns are boolean

Compare

Bernoulli Naive Bayes :

- assumes features are binary (e.g: 0 or 1)
- 0: word does not occur in the document
- 1: word occurs in the document

Multinomial Naive Bayes :

- used for discrete data (E.g: rolling dice, movie rating from 1 to 10, etc)
- In text learning we have the count of each word to predict the class or label.

Gaussian Naive Bayes :

- used for normal distribution which means all features are continuous

Bernouli vs Multinomial

In case of email classifier

Bernoulli :

- Assume spam mail has email handle in subject
- Build a feature where 0 means it's not present and 1 if it is there
- Binomial distribution

Multinomial:

- In addition to above condition, more dollar sign means spam more likely
- Same kind of word e.g: CASH or LOTTERY
- Label these words by their count
- Multinomial distribution