Naïve Bayes Classifier

- **Also known as Posterior Probabilities**
  - ✓ Certain probabilities were altered after getting additional information
  - ✓ These new probabilities are known as *posterior probabilities*

- E.g.
  - ✓ Estimating probabilities of a team winning a tournament before start
  - ✓ During the tournament, injuries to key players / underperformance
  - ✓ Probabilities altered to cater to new situation

- NB algorithm assumes all features are independent and important, which is not always correct. Hence the term “Naïve”
- E.g: Age – Salary may be correlated (more age, more salary)

Conditional probabilities for Dependent Events

Probability of an event is dependent by the occurrence of some other event

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

where

P(A | B) = Probability of event A occurring, given B has occurred

P(A) = Prior probability of event A

P(B) = Marginal likelihood of event B

$$P(B \mid A) = \frac{P(AB)}{P(A)}$$

where

P(AB) = Joint probability of A and B

P(A) = Marginal probability of event B

Type 1

P(6) = 0.4

Type 2

P(6) = 0.7

One dice is drawn at random, rolled and it gets a **6**. What is the probability that it comes from a **Type 1** dice ?

P(6 from T1) = [P(T1 giving 6) * P(T1)] / P(6)

$$\textbf{P(Type 1 | 6)} = \frac{\textbf{P(6 | Type 1) * P(Type 1)}}{\textbf{P(6)}}$$

P(6 | Type1)	= 0.4
P(6 | Type2)	= 0.7

P(Type 1)	= 0.5
P(Type 2)	= 0.5

P(6)	= (0.4*0.5) + (0.7*0.5)
	= 0.2 + 0.35
	= 0.55

P(Type 1 | 6)	**= (0.4 * 0.5) / 0.55**
	= 0.2 / 0.55
	= 0.364

Machine 1 produces 30 bolts/hour
Machine 2 produces 20 bolts/hour.

Each bolt is marked with the machine from which it is produced.

At the end of the day, pick up all defective bolts from the total produce.

Given

- 1% of all bolts produced are defective
- Of all defective parts, 50% are from each of the 2 machines

Question

What is the probability of Machine 2 producing a defective bolt ?

Machine 1

Machine 2

P(defective | M2) =

[P(M2 | defective) * P(defective)] / P(M2)
= [0.5 * 0.01] / [0.4]
= 0.0125
= 1.25%

Total bolts=1000
M2=400

1%defect = 10

50% defects from M2 = 5

Question
What is the probability of Machine 2 producing a defective bolt ?

P(defective | M2) = 5/400
= 0.0125
= .125%

Example

10% have a disease, a test to detect the disease is 92% accurate and a false alarm rate of 5%

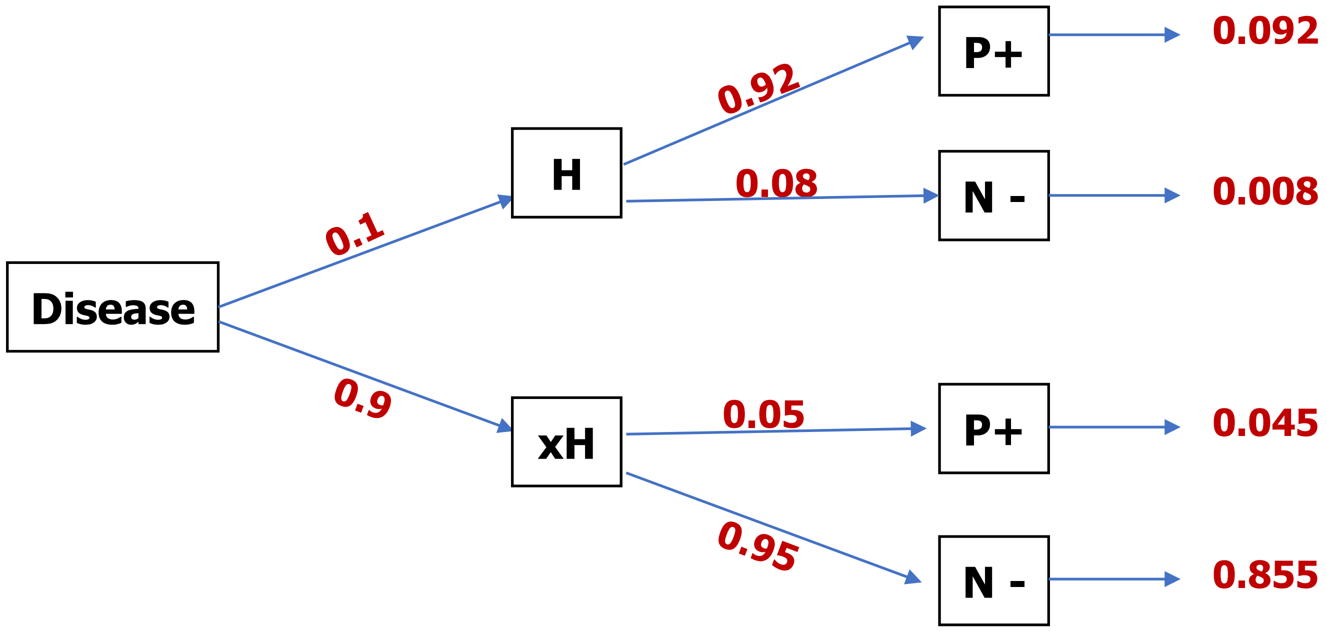
Q1) If you test positive, what is the probability you have the disease?

Q2) Your friend tests negative, what is the probability the friend has the disease?

10% have a disease, a test to detect the disease is 92% accurate and a false alarm rate of 5%

Q1) If you test positive, what is the probability you have the disease?

Q2) Your friend tests negative, what is the probability the friend has the disease?



If you test positive, what is the probability you have the disease?

Test Positive = 0.045 + 0.092 ➔ 0.137
Positive and Have disease ➔ 0.092

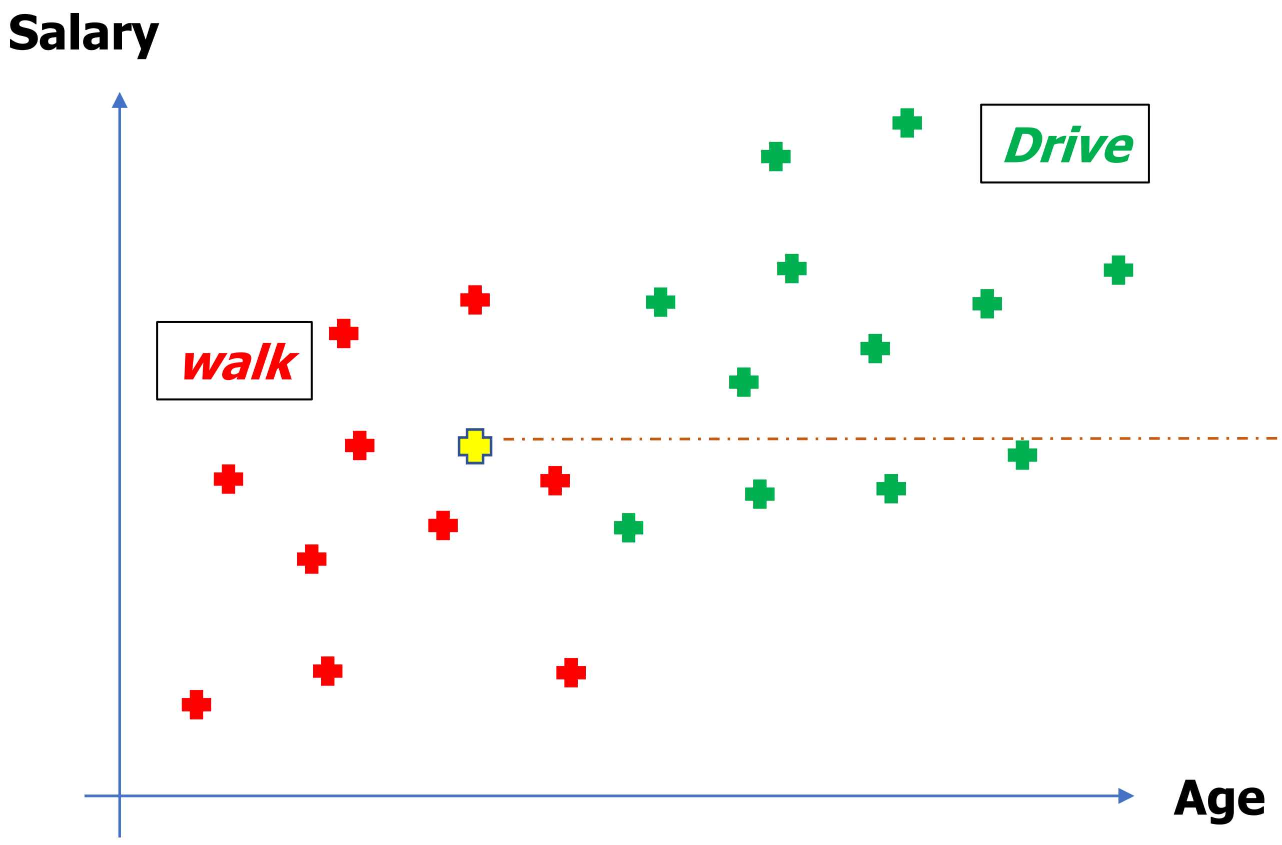
P(You have the disease | You test positive) = 0.092/0.137 ➔ 0.671 ➔ 67.1%

If your friend tests negative, what is the probability the friend has the disease?

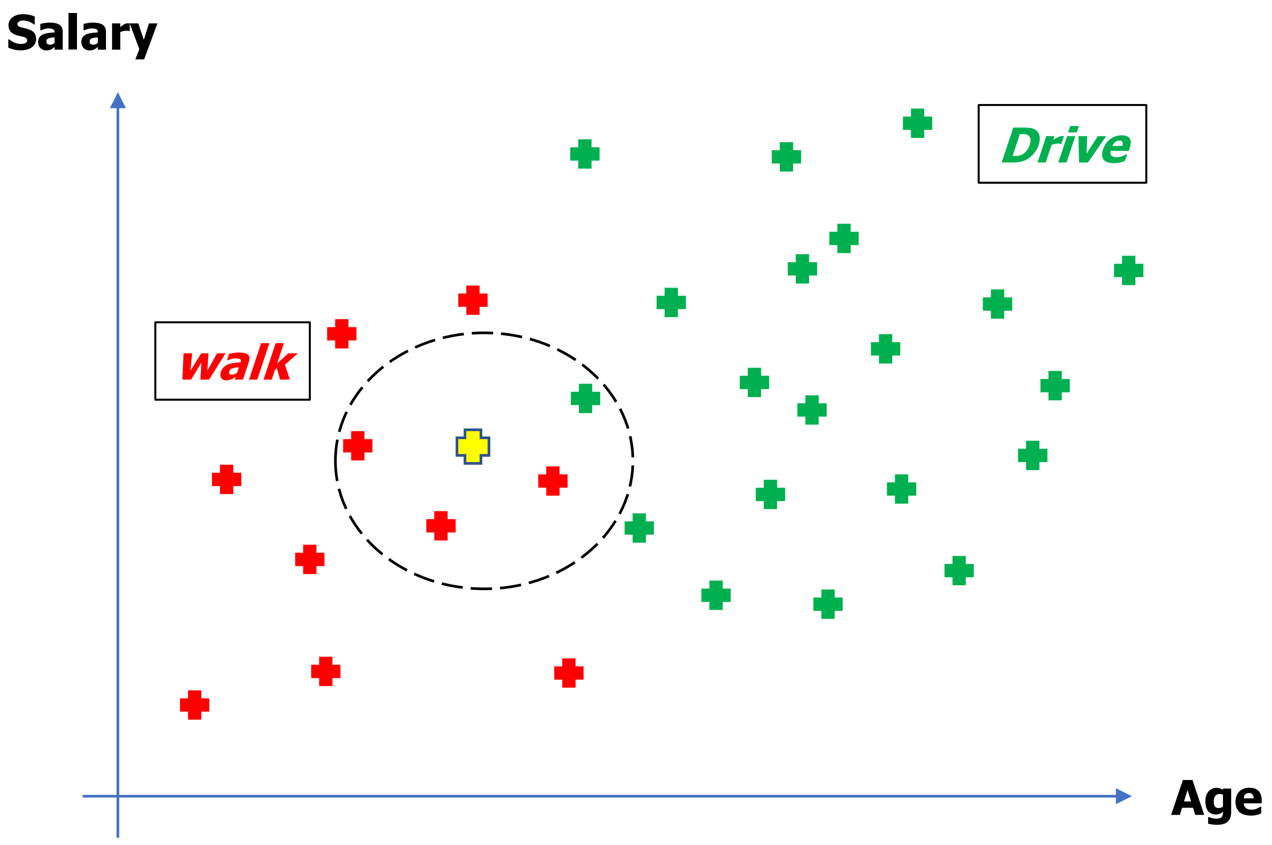
Test Negative = 0.008 + 0.855 ➔ 0.863
Negative and Have disease ➔ 0.008

P(Friend has the disease| Friend tests negative) = 0.008/0.863 ➔ 0.0092 ➔ 0.9%

1. P(Walk | Age) = [P(Age | Walk) * P(Walk)] / P(Age)
2. P(Drive | Age) = [P(Age | Drive) * P(Drive)] / P(Age)
3. Compare P(1) vs P(2) to classify



Predict whether this person will **Walk** or **Drive** ?



1. **P(Walks) = 10/30**

2. **P(Age) = 4/30**
   Probability of a random variable added falls in the circle

3. **P(Age | Walk) = 3/10**
   Probability of someone who walks exhibits the feature 'Age' (from the circle)

P(Walk | Age) = [(3/10)*(10/30)] / (4/30)
= 0.75
= 75%

P(Drive | Age) = 1 – 0.75
= 0.25
= 25%

Since P(Walk|Age) > P(Drive|Age),
will be classified as Walk