Universidade do Minho
Escola de Engenharia
Departamento de Informática

Mestrado Integrado em Engenharia Informática
Mestrado em Engenharia Informática
Aprendizagem e Extração de Conhecimento
2020/2021

Filipe Gonçalves, César Analide, Paulo Novais

Synthetic Intelligence Lab

AEC @ 2020/2021

- Paulo Novais – pjon@di.uminho.pt
- César Analide – analide@di.uminho.pt
- Filipe Gonçalves – fgoncalves@algoritmi.uminho.pt

- Departamento de Informática
  Escola de Engenharia
  Universidade do Minho
- ISLab – (Synthetic Intelligence Lab)
- Centro ALGORITMI
  Universidade do Minho

Conditional Probability

Conditional Probability

- If I have two events that depend on each other, what's the probability that both will occur?
- Notation:
  - P(A,B) is the probability of events A and B both occuring
  - P(B|A) is the probability of event B given that event A has occured

Probability of event A occured and event B occured

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

Probability of event A given B has occured

Probability of event B

Conditional Probability

- Example:
  - Two tests are given to a class
  - 60% of the students passed both tests
  - 80% of the students passed the first test
  - What percentage of students who passed the first test also passed the second one?
- Resolution:
  - A = passing the 1st test, B = passing the 2nd test $P(B|A) = \frac{P(A,B)}{P(A)} = \frac{0.6}{0.8} = 0.75$
  - 75% of students who passed the first test passed the second one

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

Probability of event A occured and event B occured

Probability of event A given B has occured

Probability of event B

Bayes' Theorem

- Through the understanding of conditional probability, the Bayes' Theorem can be defined as:

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

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

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

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

- In English: Probability of A given B = (Probability A) x (Probability of B given A)) / (Probability B)

- The key insight is that the probability of something that depends on B depends on the base probability of B and A

Bayes' Theorem

- Drug testing is a common example
- A "highly accurate" drug test can produce more false positives than true positives
- Example:
    - a Drug test can accurately identify users of the drug 99% of the time, and accurately has a negative result for 99% of non-users
    - Only 0.3% of the overall population uses this drug

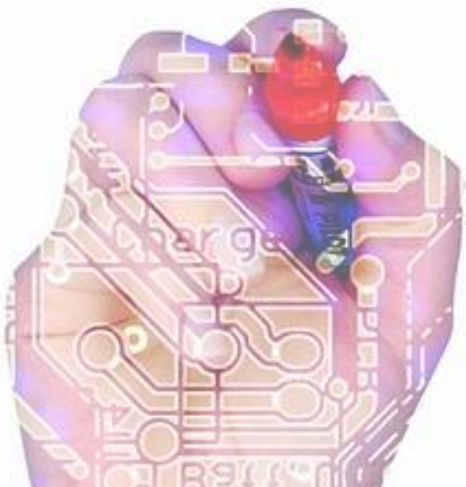
Bayes' Theorem

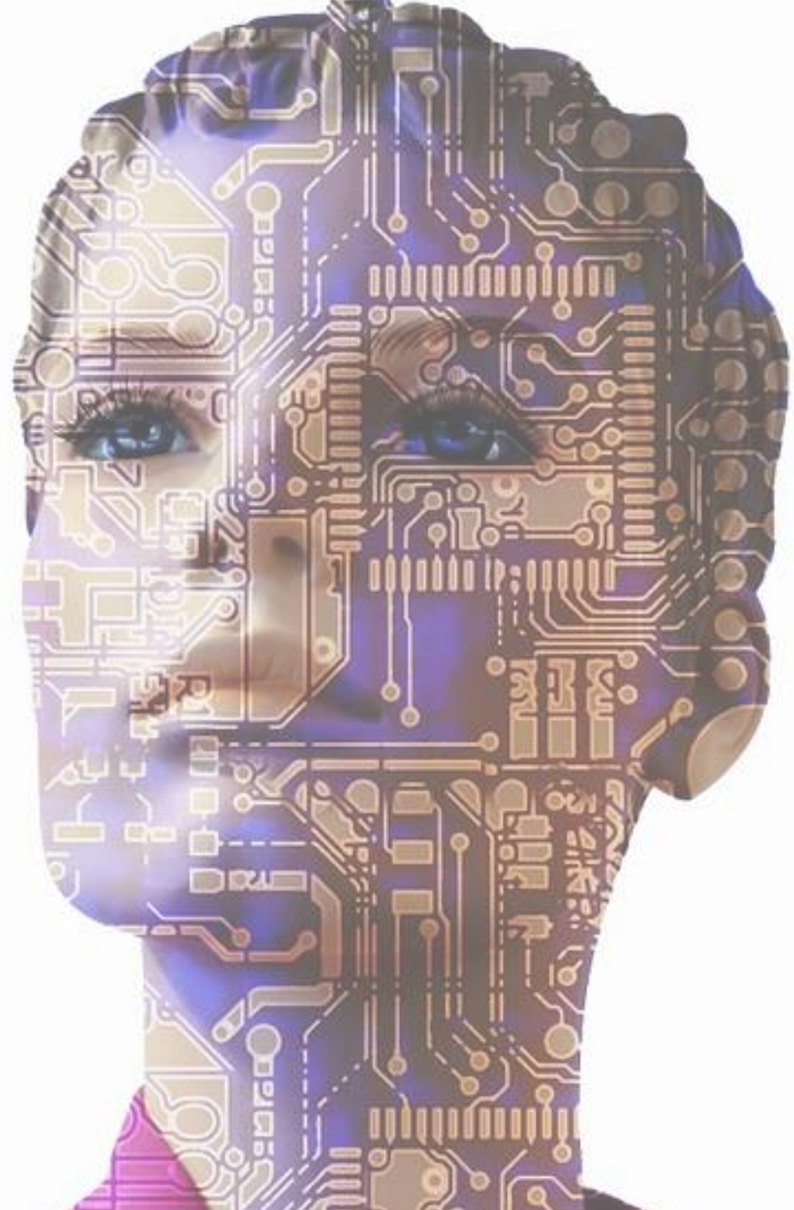
- Event A = Is a user of the drug / Event B = tested positively for the drug
- P(B) = 0.99 x 0.003 + 0.01 x 0.997 = 1.3%
  - Prob. of testing positive if you do use + prob. of testing positive if you don't

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} = \frac{0.003 * 0.99}{0.013} = 22.8\%$$

- The odds of someone being an actual user of the drug given that they tested positive is only 22.8%
- Even though P(B|A) is high (99%), it doesn't mean P(A|B) is high!

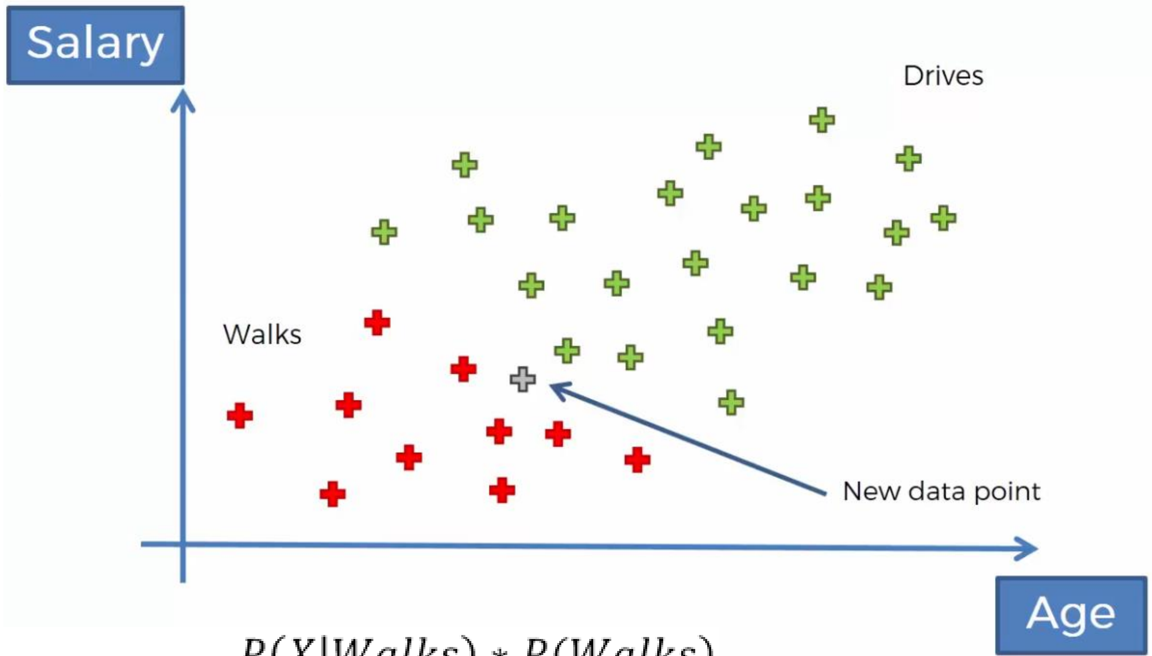
Naïve Bayes





Bayes' Theorem

- Bayes' Theorem can also be applied as a machine learning classifier
- Example 1: Based on Age and Salary features, classify if a person drives or walks daily to work
- Formula:
  - P(A) = P(Walks)
  - P(B) = P(X), where X = features set
  - P(A|B) = P(Walks|X)
  - P(B|A) = P(X|Walks)

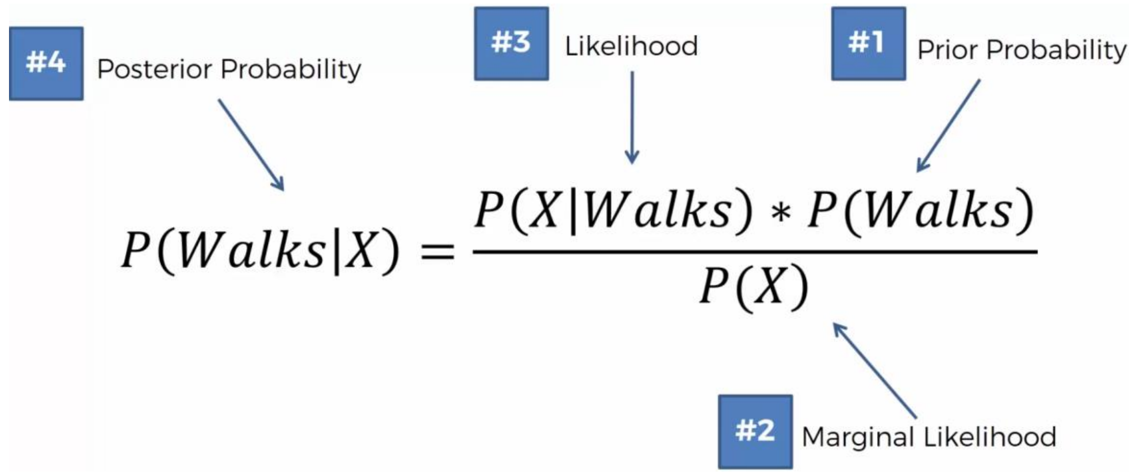


$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} \Rightarrow P(Walks|X) = \frac{P(X|Walks) * P(Walks)}{P(X)}$$

Bayes' Theorem – How to calculate?

Step 1 – Calculate P(Walks|X):

1. Calculate Prior Probability
2. Calculate Marginal Likelihood
3. Calculate Likelihood
4. Based on calculated values [1-3], we end up calculate Posterior Probability



Bayes' Theorem – How to calculate?

Step 2 – Calculate P(Drive|X):

1. Calculate Prior Probability (same as in Step 1)
2. Calculate Marginal Likelihood
3. Calculate Likelihood
4. Based on calculated values [1-3], we end up calculate Posterior Probability

#4 Posterior Probability

#3 Likelihood

#1 Prior Probability

$$P(Drives|X) = \frac{P(X|Drives) * P(Drives)}{P(X)}$$

#2 Marginal Likelihood

Bayes' Theorem – How to calculate?

Step 3:

1. Compare & select the highest probability of both

   $P(Walks|X)\ v.s.\ P(Drives|X)$

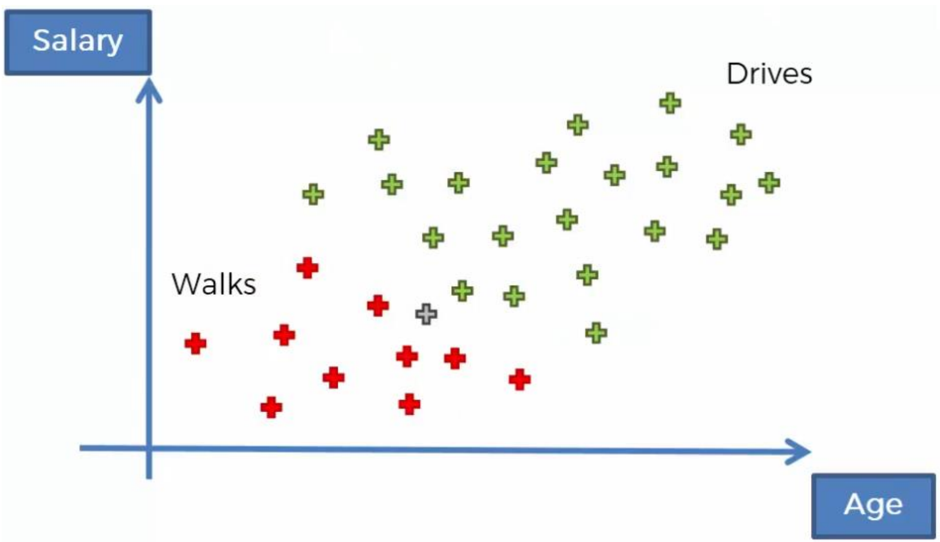
2. For classification purposes only (Class A vs Class B):

$$\frac{P(X|Walks) * P(Walks)}{\cancel{P(X)}} \quad v.s. \quad \frac{P(X|Drives) * P(Drives)}{\cancel{P(X)}}$$

3. For classifier performance evaluation, P(X) must be always calculated

Bayes' Theorem – How to calculate (Step 1)?

1. P (Walks) = Probability that a person walks to work (without knowing any feature)



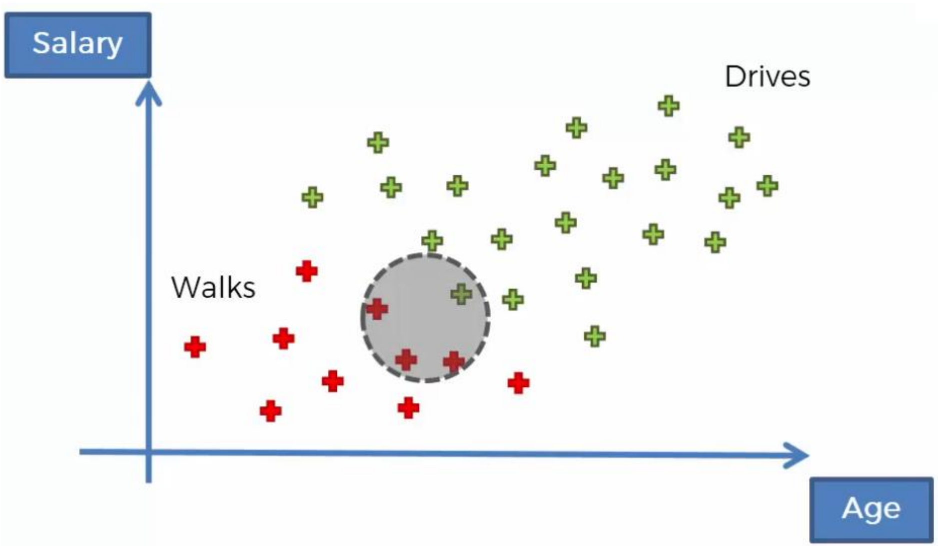
$$P(Walks) = \frac{Number\ of\ Walkers}{Total\ Observations}$$

$$P(Walks) = \frac{10}{30}$$

Bayes' Theorem – How to calculate (Step 1)?

2. P (X) = probability of similar features exist in the dataset

- Select a radius size and place a circle over the new data point to be classified
- Calculate percentage of #Similar_Observations (inside the circle) vs #TotalObservations



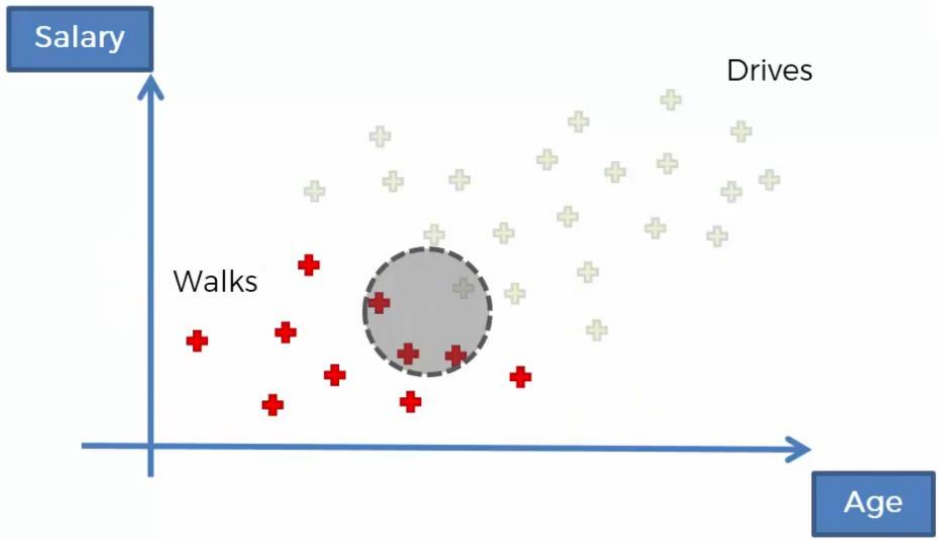
$$P(X) = \frac{Number\ of\ Similar\ Observations}{Total\ Observations}$$

$$P(X) = \frac{4}{30}$$

Bayes' Theorem – How to calculate (Step 1)?

3. P (X|Walks) = probability of similar features exist in the dataset given that person walks

- Filter the walkers observations in the dataset
- Select a radius size and place a circle over the new data point to be classified
- Calculate percentage of #Similar_Walkers (inside the circle) vs #Total_Walkers



$$P(X|Walks) = \frac{\begin{array}{c}Number\ of\ Similar\\ Observations\\ Among\ those\ who\ Walk\end{array}}{Total\ number\ of\ Walkers}$$

$$P(X|Walks) = \frac{3}{10}$$

Bayes' Theorem – How to calculate (Step 1)?

4. P (Walks|X) = probability of being a walker given the set of features

- Through the previous values, simply calculate P(Walks|X) given the Bayes' Theorem formula

#4 Posterior Probability

#3 Likelihood

#1 Prior Probability

$$P(Walks|X) = \frac{\frac{3}{10} * \frac{10}{30}}{\frac{4}{30}} = 0.75$$

#2 Marginal Likelihood

Bayes' Theorem – How to calculate (Step 2)?

1. Repeat the previous process to calculate P(Drives|X)

#4 Posterior Probability

#3 Likelihood

#1 Prior Probability

$$P(Drives|X) = \frac{\frac{1}{20} * \frac{20}{30}}{\frac{4}{30}} = 0.25$$

#2 Marginal Likelihood

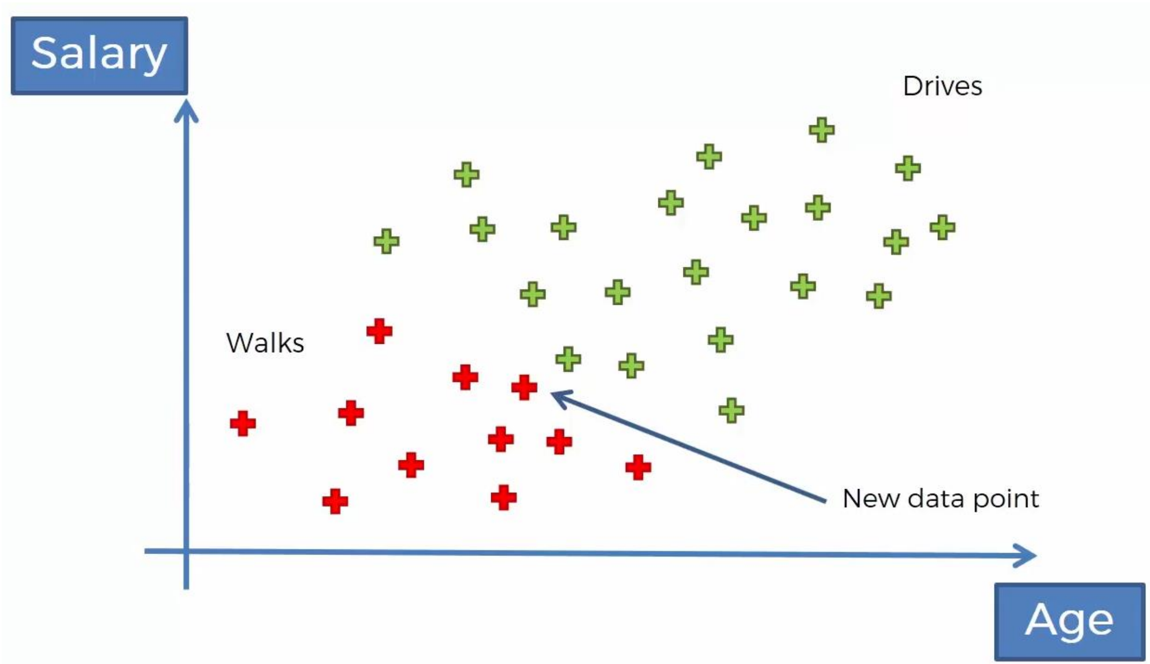
Bayes' Theorem – How to calculate (Step 3)?

1. Compare & select highest probability

$$P(Walks|X)\ v.s.\ P(Drives|X)$$

→ 0.75 vs 0.25

→ Classification = Walker



Bayes' Theorem

- Example 2: how would we express the probability of an email being spam if it contains the word "free"?

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} \Rightarrow P(Spam \mid Free) = \frac{P(Spam)P(Free \mid Spam)}{P(Free)}$$

- The numerator is the probability of a message being spam and containing the word "free"
  - This is subtly different from what we're looking for
- The denominator is the overall probability of an email containing the word "free"
  - Equivalent to: ***P(Free|Spam) x P(Spam) + P(Free|Not Spam) x P(Not Spam)***
  - This ratio is the percentage of emails with the word "free" that are spam

What about all the other words?

- We can construct P(Spam|Word) for every (meaningful) word we encounter during training
- Then multiply these together when analyzing a new email to get the probability of it being spam
- Assumes the presence of different words are independente of each other – one reason this is called "Naïve Bayes"

Universidade do Minho
Escola de Engenharia
Departamento de Informática

Mestrado Integrado em Engenharia Informática
Mestrado em Engenharia Informática
Aprendizagem e Extração de Conhecimento
2020/2021

Filipe Gonçalves, César Analide, Paulo Novais