


Bayes Theorem & Naïve Bayes



Agenda

- Bayes Theorem
 - Varför
 - Vad
 - Hur - Exempel
- Länkar

- Naïve Bayes
 - Bayes Theorem inom ML
 - Klassificering
 - IRL
 - Fördelar
 - Nackdelar
- Kod
 - Naïve Bayes

Termer

- Likelihood - Snarlikhet
- Probability - Sannolikhet
- Conditional probability - Sannolikhet beroende av en annan faktor.

"For lay people ... , probability (a synonym of likelihood ...) is a vague notion, related to uncertainty, propensity, plausibility, and surprise."

- Daniel Kahneman, Thinking fast and slow

"Bedömningar av snarlikhet (likelihood) och sannolikhet (probability) lyder inte under samma logiska regler. Bedömningar av snarlikhet behöver inte ta hänsyn till relativa frekvenser."

- Daniel Kahneman, Thinking fast and slow

Bayes Theorem - Varför

Statistik är viktigt inom AI.

Hjälper till att klassificera mer exakt.

Bayes Theorem ger oss möjligheten att uppdatera/**justera** sannolikheten för något i samband med mer/**ny information**.

Förlitar sig på statistisk data (om fördelningar) för att modellera olika scenarios.

T.ex. Sannolikheten att ett mail är spam, utan någon annan info. Sedan kan man lägga till vilka ord förekommer i mailet, och deras frekvens, för att **uppdatera** sannolikheten.

"Evidence should not determine beliefs, but update them."
- 3Blue1Brown

Source:

<https://machinelearningmastery.com/bayes-theorem-for-machine-learning/#:~:text=Bayes%20theorem%20provides%20a%20way.and%20the%20observed%20data%20itself.>

RE STORY...

Bayes Theorem - Vad

Inom Bayes finns det 3 delar:

- Hypotes (Hypothesis) A
- Bevis (Evidence) B
- Sannolikhet (Probability) P()

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

Och en blandning av dessa:

- Sannolikheten för hypotesen utan några bevis. Kallas för **Prior**.
 - Sannolikheten för bevisen. **Evidence**
 - Sannolikheten för bevisen om hypotesen är sant. Kallas för **Likelihood**.
 - Sannolikheten för hypotesen givet bevisen. Kallas även **Posterior**.
-
- P(A) Sannolikheten för A
 - P(B) Sannolikheten för B
 - P(B|A) Sannolikheten för B om A är sant
 - P(A|B) Sannolikheten för A om B är sant

Bayes - Exempel

Info (evidence):

1. There are less basketball players in the world than non basketball players.
2. A large percent of people who play basketball are tall.

Är det mer sannolikt att en lång person spelar basket - Ja.
Spelar alla långa personer basket? - Nej.

Information i siffror:

- One to ten people play basketball.
- Of those who play basketball, 75% are tall.
- Of the people who don't play basketball, 15% are tall.

Innan vi ens vet om en person är lång eller ej, så vet vi att sannolikheten att de spelar basket är 9% (1/11).



- Prior **odds** ratio - 1:10
- Likelihood ratio - 75:15

Bayes - Exempel

- Prior odds ratio - 1:10 basket spelare
- Likelihood ratio - 75:15 långa personer
- $P(A)$ Sannolikheten för A
- $P(B)$ Sannolikheten för B
- $P(B|A)$ Sannolikheten för B om A är sant
- $P(A|B)$ Sannolikheten för A om B är sant

Identifiera de olika delarna av formeln.



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

Bayes - Exempel

- Prior odds ratio - 1:10 basket spelare
- Likelihood ratio - 75:15 långa personer
- $P(A)$ En person spelar basket - 9% (1/11)
- $P(B)$ Sannolikheten att en person är lång
- $P(B|A)$ En person är lång och spelar basket - 75%
- $P(A|B)$ Sannolikheten att en person spelar basket om de är långa



Vad är $P(B)$?

- **Lång och spelar basket**
- **Lång och spelar ej basket**

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

Bayes - Exempel

Vad är $P(B)$?

- Långa personer ratio - 75:15
 - Lång och spelar basket: $0.75 * 0.09 = 0.0675$
 - Lång och spelar ej basket: $0.15 * (1 - 0.09) = 0.1365$
 - Är lång: $0.0675 + 0.1365 = 0.204$
-
- Långa personer ratio - 75:15 = 5:1
 - 5 gånger mer sannolikt att en basketspelare är lång.
 - $\frac{1}{5} = 0.2$



Bayes - Exempel

Hypotes: En lång person spelar basket.

Vad är sannolikheten för att en lång person spelar basket?

- Måste vara mer än 9% eftersom 75% av basketspelare är långa.

- $P(A)$ Spelar basket 9% = 0.09
- $P(B)$ Är lång 20.4% = 0.204
- $P(B|A)$ Spelar basket och är lång 75% = 0.75

$$0.75 * 0.09 / 0.204$$

- $P(A|B)$ Är lång och spelar basket. -> 0.33



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

Bayes Theorem - Fabrik exempel

- En fabrik har två maskiner som båda tillverkar skiftnycklar.
- Maskinerna jobbar olika snabbt och skiftnycklarna skiljer sig lite åt och är markerade så att man vet vilken maskin de kommer från.
- $M1 = 30/h$
- $M2 = 20/h$
- Av alla producerade skiftnycklar är 1% defekta
- Av alla defekta är lika många kom från M1 som M2. D.v.s 50%

Vad är sannolikheten att en skiftnyckel som M2 producerat är defekt?

Bayes Theorem - Fabrik exempel

Totalt antal skiftnycklar producerade under en timme: **30 + 20 = 50**

- $P(M1) = 30/50 = 0.6$
- $P(M2) = 20/50 = 0.4$
- $P(\text{Defekt}) = 1\% = 0.01$
- $P(M1 | \text{Defekt}) = 50\% = 0.5$
- $P(M2 | \text{Defekt}) = 50\% = 0.5$

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

Vad är sannolikheten att en skiftnyckel som M2 producerat är defekt?

- $P(\text{Defekt} | M2) = ?$

Bayes Theorem - Fabrik exempel

- $P(M1) = 0.6$
- $P(M2) = 0.4$
- $P(\text{Defekt}) = 0.01$
- $P(M1 | \text{Defekt}) = 0.5$
- $P(M2 | \text{Defekt}) = 0.5$

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

Vad är sannolikheten att en skiftnyckel som M2 producerat är defekt?

- $P(\text{Defekt} | M2) = P(M2 | \text{Defekt}) * P(\text{Defekt}) / P(M2)$
- $0.5 * 0.01 / 0.4 = 0.0125$
- 1.25%

Bayes Theorem - Fabrik exempel

- $P(M1) = 0.6$
- $P(M2) = 0.4$
- $P(\text{Defekt}) = 0.01$
- $P(M1 | \text{Defekt}) = 0.5$
- $P(M2 | \text{Defekt}) = 0.5$

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

Exempel med siffror:

- 1000 Skiftnycklar
- Hur många kommer från M2?
- Hur många är defekta?
- Hur många defekta kommer från M2?

- % defekta från M2?

$$(10 * 0.5) / 400 = 1.25\%$$

Bayes Theorem - Övning

$$P(\text{Defect} | M1) = ?$$

Naïve Bayes

Bayes Theorem - Inom ML

Ett sätt att tänka på förhållandet mellan data och en model.

En modell kan ses som en hypotes om förhållandena i data, t.ex förhållandet mellan input(X) och output(y).

Varför Naive?

Antar att alla variabler är "naïva", d.v.s. Inte korrelerade (med varandra).

"The practice of applied machine learning is the testing and analysis of different hypotheses (models) on a given dataset."

- machinelearningmastery.com

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

Source: <https://machinelearningmastery.com/bayes-theorem-for-machine-learning/#:~:text=Bayes%20theorem%20provides%20a%20way.and%20the%20observed%20data%20itself.>

RE STORY...

Bayes Theorem - Inom ML

Inom klassifikation kan man tänka att:

- $P(A)$ - $P(\text{class})$
- $P(B)$ - $P(\text{data})$
- $P(B|A)$ - $P(\text{data} | \text{class})$
- $P(A|B)$ - $P(\text{class} | \text{data})$

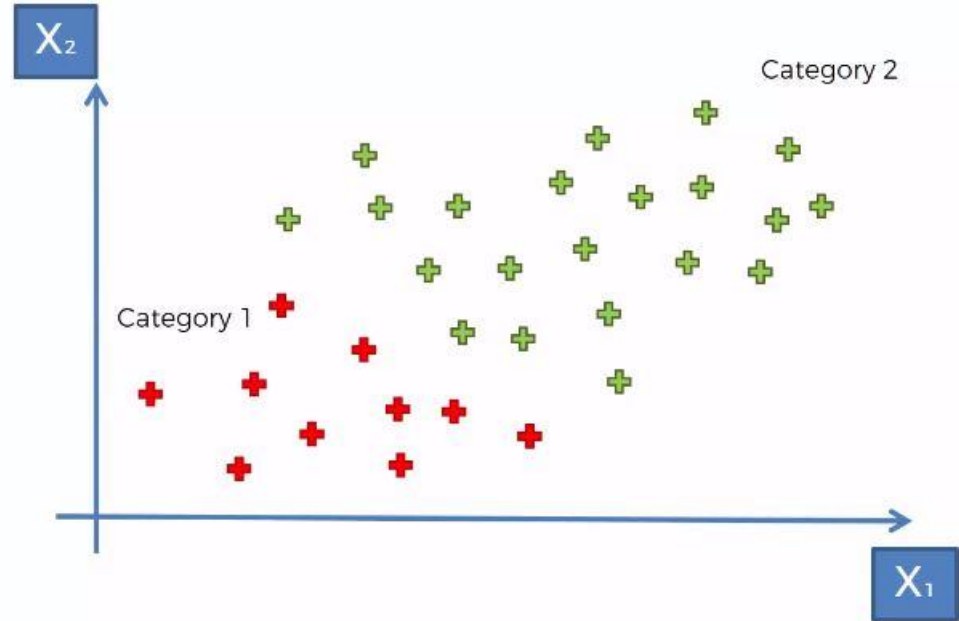
$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

Source: <https://machinelearningmastery.com/bayes-theorem-for-machine-learning/#:~:text=Bayes%20theorem%20provides%20a%20way.and%20the%20observed%20data%20itself.>

RESTOR\...

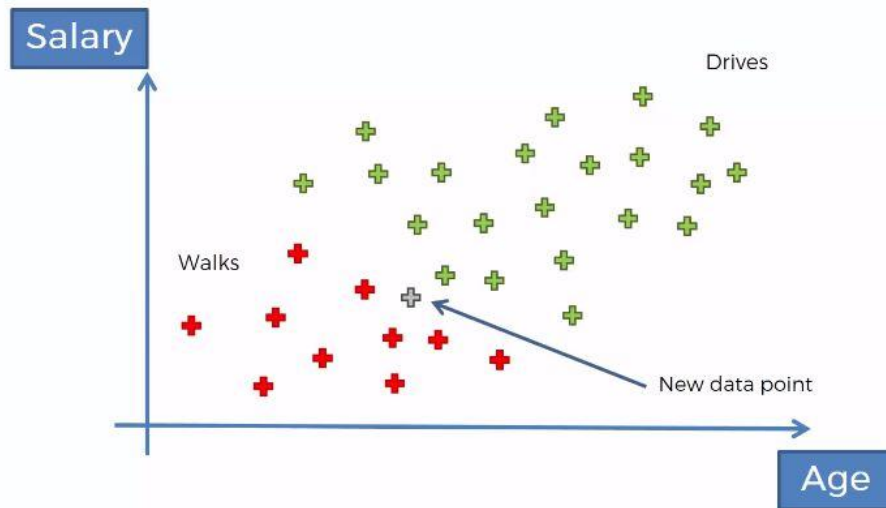
Naïve Bayes - Klassificering

- Vi har två *features*: x_1 och x_2
- För att beteckna båda: X
- Två classer: röd och grön
- Vi ska räkna ut sannolikheten för att en ny punkt tillhör en viss klass.



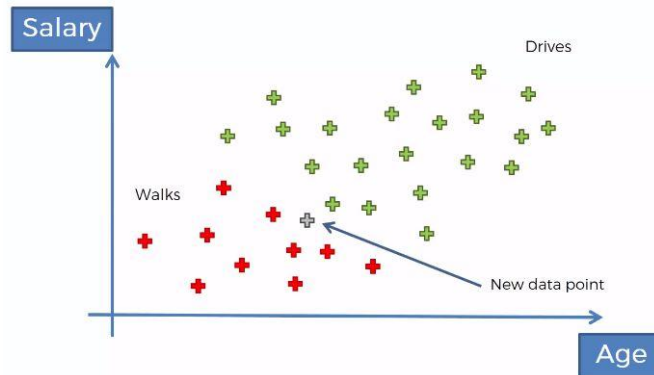
Naïve Bayes - Klassificering

- Vi har två *features*: ålder och lön
- Två classer: går eller kör till jobbet.
- Vi ska räkna ut sannolikheten för varje klass.
- Hur tar de sig till jobbet?



Naïve Bayes - Klassificering

- Använder oss av Bayes Theorem för att bestämma sannolikheten.



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

Naïve Bayes - Klassificering

- Sannolikheten för hypotesen utan några bevis. Kallas för **Prior**.
- Sannolikheten för bevisen. **Evidence**
- Sannolikheten för bevisen om hypotesen är sant. Kallas för **Likelihood**.
- Sannolikheten för hypotesen givet bevisen. Kallas även **Posterior**.

Identifiera vad varje del är.

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

Naïve Bayes - Klassificering

- Sannolikheten för hypotesen utan några bevis. Kallas för **Prior**.
- Sannolikheten för bevisen. **Evidence / Marginal Likelihood**
- Sannolikheten för bevisen om hypotesen är sant. Kallas för **Likelihood**.
- Sannolikheten för hypotesen givet bevisen. Kallas även **Posterior**.

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

#4 Posterior Probability

#3 Likelihood

#1 Prior Probability

#2 Marginal Likelihood

Source: https://en.wikipedia.org/wiki/Marginal_likelihood

Naïve Bayes - Klassificering

Räknar ut $P(\text{Walks} | X)$, sannolikheten att en person med features X går till jobbet och $P(\text{Drives} | X)$, sannolikheten att en person med features X kör till jobbet.

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

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

Naïve Bayes - Klassificering

Jämför $P(\text{Walks}|X)$ och $P(\text{Drives}|X)$ för att se vilket är mer sannolikt.

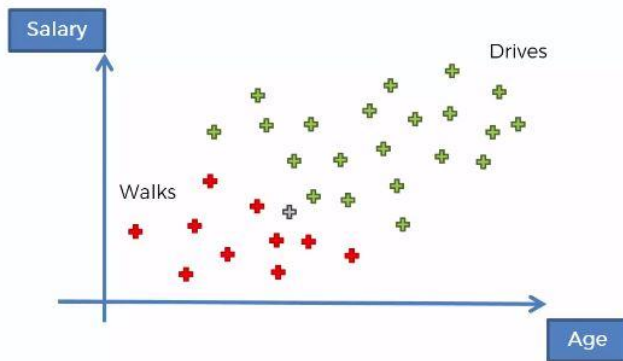
$$P(\text{Walks}|X) \text{ v.s. } P(\text{Drives}|X)$$

Naïve Bayes - Klassificering

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

Steg 1: **Prior Probability**

- Sannolikheten för att en person går, oavsett X.



#1. **P(Walks)**

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

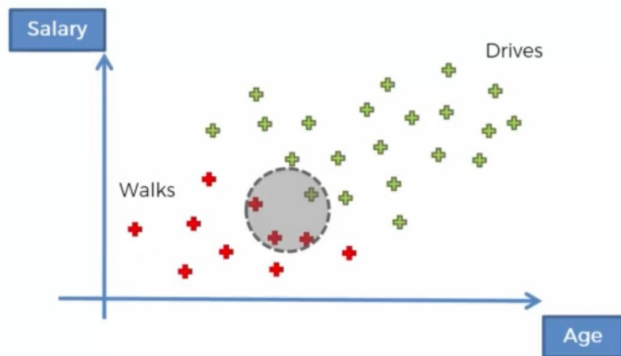
$$P(Walks) = \frac{10}{30}$$

Naïve Bayes - Klassificering

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

Steg 2: **Marginal Likelihood / Evidence**

- Sannolikheten för att en person har parametrar X.
Välj ett område (närmaste punkterna). Kan välja storleken på cirkeln (avståndet till de närmaste punkterna).



#2. $P(X)$

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

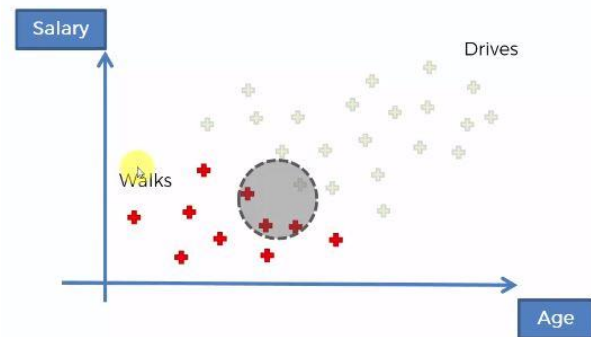
$$P(X) = \frac{4}{30}$$

Naïve Bayes - Klassificering

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

Steg 3: Likelihood

- Sannolikheten för att en person med features X går.
Av de närmaste punkterna, hur många går?.



#3. $P(X|Walks)$

$$P(X|Walks) = \frac{\text{Number of Similar Observations Among those who Walk}}{\text{Total number of Walkers}}$$
$$P(X|Walks) = \frac{3}{10}$$

Naïve Bayes - Klassificering

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

Steg 4: **Posterior Probability**

- Sannolikheten för att en person som går har parametrarna X.
- Sannolikheten att en person med parametrarna X borde klassificeras som gående. Stoppa in talen.

The diagram illustrates the calculation of the Posterior Probability, $P(Walks|X)$, using three components: Likelihood, Prior Probability, and Marginal Likelihood. Each component is represented by a blue box with a number and a label, and a green checkmark. Arrows point from these boxes to the corresponding parts of the equation.

#4 Posterior Probability

#3 Likelihood

#1 Prior Probability

#2 Marginal Likelihood

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

Naïve Bayes - Klassificering

Räknar ut $P(\text{Walks}|X)$, sannolikheten att en person med features X går till jobbet och $P(\text{Drives}|X)$, sannolikheten att en person med features X kör till jobbet.

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

Räkna ut $P(\text{Drives}|X)$

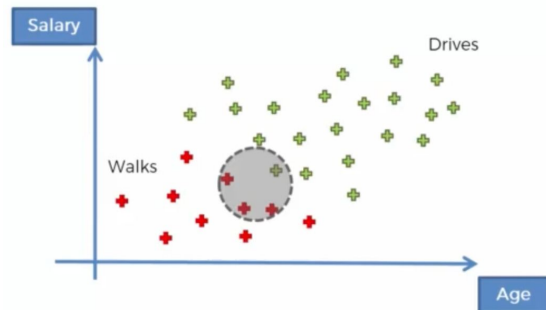
$$P(\text{Drives}|X) = \frac{P(X|\text{Drives}) * P(\text{Drives})}{P(X)}$$

Naïve Bayes - Klassificering

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

Sannolikheten för att en person som kör har parametrarna X.

- Identifiera delarna



#4 Posterior Probability

#3 Likelihood

#1 Prior Probability

#2 Marginal Likelihood

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

Naïve Bayes - Klassificering

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

Steg 4: **Posterior Probability**

- Sannolikheten för att en person som kör har parametrarna X.
- Sannolikheten att en person med parametrarna X borde klassificeras som körande.

The diagram illustrates the calculation of the Posterior Probability ($P(Drives|X)$) using the Naïve Bayes formula. It features four numbered boxes with arrows pointing to the corresponding parts of the formula:

- #1 Prior Probability** points to $\frac{20}{30}$.
- #2 Marginal Likelihood** points to $\frac{4}{30}$.
- #3 Likelihood** points to $\frac{1}{20}$.
- #4 Posterior Probability** points to the entire formula.

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

Naïve Bayes - Klassificering

Jämför de två.

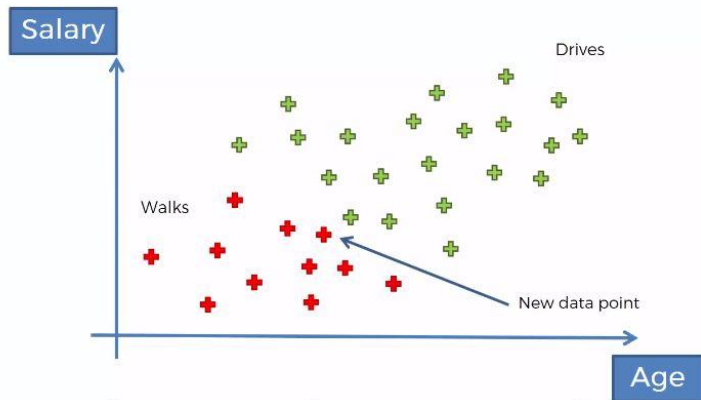
$$P(Walks|X) \text{ v.s. } P(Drives|X)$$

$$P(Drives|X) = 0.25$$

$$P(Walks|X) = 0.75$$

$$0.75 > 0.25$$

Naïve Bayes - Klassificering



Sannolikheten för att personen med parametrarna X går är 75% och endast 25% att de kör. Därmed klassificeras den nya datapunkten som gående.

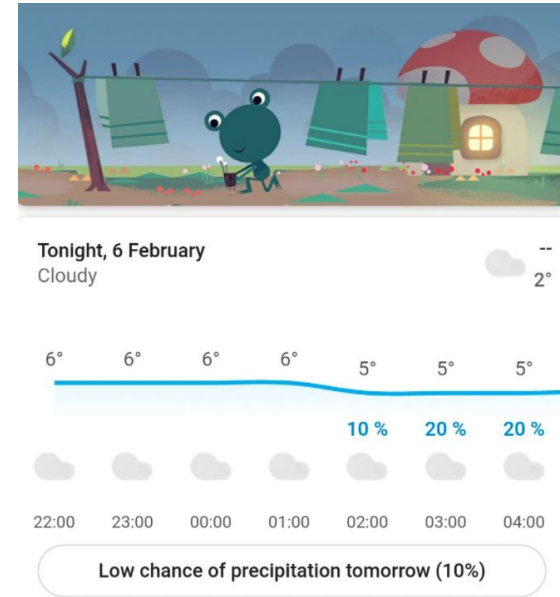
Notera att summan av utfallen är 100%.

Detta stämmer oavsett antalet klasser, summan ska alltid bli 100%!

$$0.75 > 0.25$$

$$P(Walks|X) > P(Drives|X)$$

Naïve Bayes - IRL



Klassificering av dokument och så kallad Sentiment Analysis.

Rekomendationssystem.

Kan användas för att göra förutsägelser i realtid.

Naïve Bayes - Fördelar

- Det är enkelt och snabbt att förutsäga klassen för testdata. Den fungerar också bra för prediktion av multipla klasser (>2).
- När antagandet om oberoende variabler gäller, presterar en Naive Bayes-klassificerare bättre jämfört med andra modeller som logistisk regression.
- Mindre träningsdata behövs.

Naïve Bayes - Nackdelar

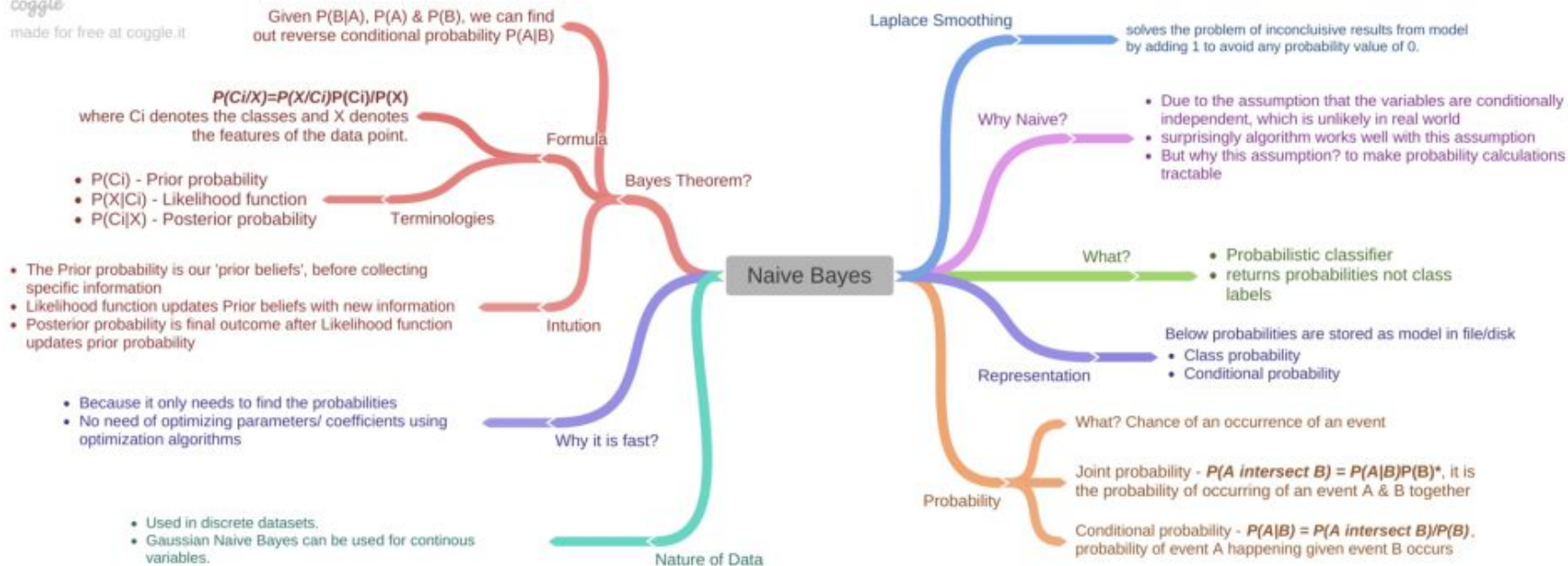
- Om en helt **ny kategorivariabel** observeras (i testdata), kommer modellen att ge den noll i sannolikhet: $P(\text{NyVar}) = 0$ och kommer inte att kunna göra en förutsägelse. Detta kallas ofta **nollfrekvens**. För att lösa detta kan vi använda en utjämningsteknik, t.ex. **Laplace-uppskattning**. Enkelt sagt lägger den till åtminstone ett exempel i varje kategori så att inget kan bli noll. Läs bifogat exempel.
- En begränsning av Naïve Bayes är antagandet att prediktorer (variabler) är oberoende. I det verkliga livet är det nästan omöjligt att vi får en uppsättning prediktorer som är helt oberoende.

Source: <https://towardsdatascience.com/all-about-naive-bayes-8e13cef044cf>

Naïve Bayes

coggle

made for free at coggle.it



Sammanfattning

- Bayes Theorem

- Varför
- Vad
- Hur - Exempel

*"... look at the cases where the evidence is true and then consider the **proportion** of those cases where the hypothesis is also true."*

- [3Blue1Brown](#)

- Naïve Bayes

- Bayes Theorem inom ML
- Klassificering
- IRL
- Fördelar
- Nackdelar

Länkar

- [What is a hypothesis in ML?](#)
- [Bayes simple](#)
- [Bayes - extra info](#)
- [Bayes - medical paradox](#)
- [Probability Distributions](#)
- [Naive Bayes - real examples](#)
- [Naive Bayes - simply](#)
- [Intro to Bayes for ML](#)