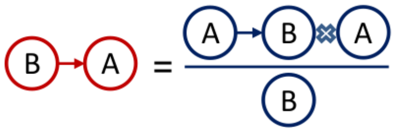
VL 5 – Bayes und Markov

Teil 1: Bayes

- behandelt die Zusammenhänge von Wahrscheinlichkeiten
- Ermittlung von gesuchten Wahrscheinlichkeiten aus gegebenen W.
- ist nach dem englischen Mathematiker Thomas Bayes benannt
- wurde postum 1763 veröffentlicht (Link)

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



- ***die Wahrscheinlichkeit von A, wenn B bereits eingetreten ist***

Wahrscheinlichkeiten (Exkurs):

- A-Priori — Anfangswahrscheinlichkeit (Schätzung anhand von allg. Vorwissen)
- A-Posteriori — Wissensstand (Ermittlung anhand von Berechnungen)

Der Satz von Bayes (auch das Bayes Theorem genannt) ist nach dem englischen Mathematiker Thomas Bayes benannt und wurde postum 1763 nach seinem Tod veröffentlicht.

Frage: Was ist eine bedingte Wahrscheinlichkeit?

Die Wahrscheinlichkeit, ein Ereignis (z.B. X=i) bei einem bestimmten anderen Ereignis (z.B. Y=i) zu beobachten. Notiert als $P(X_i|Y_i)$

Frage: Worum geht es in dem Satz von Bayes?

Es geht um die Ermittlung von Wahrscheinlichkeiten aus bereits vorhandenen Wahrscheinlichkeiten - das heißt, der Zusammenhang zwischen Wahrscheinlichkeiten soll ermittelt werden. Beim Bayes'schen Theorem geht es darum, Wahrscheinlichkeiten zu aktualisieren, basierend auf neuen Informationen oder Daten. Es ermöglicht, wie sich die Wahrscheinlichkeit eines Ereignisses ändert, wenn man bereits einige Informationen oder Annahmen hat und dann zusätzliche Daten oder Beobachtungen hinzufügt.In einfachen Worten: Das Bayes'sche Theorem hilft dabei, unsere ursprüngliche Meinung oder Schätzung über ein Ereignis zu aktualisieren, wenn wir neue Informationen erhalten. Es ist besonders nützlich in Situationen, in denen Unsicherheit und Abhängigkeiten zwischen Ereignissen eine Rolle spielen.
> Bei Bayes geht es darum zu beobachten, wie sich die Wahrscheinlichkeit eines Ereignisses ändert, wenn sich eine andere Wahrscheinlichkeit, die Einfluss auf dieses Ereignis hat, geändert hat. man aktualisiert also seine Wahrscheinlichkeit.
> Prior Beliefs get updated with new data!

Die Wahrscheinlichkeit von A, bezeichnet als P(A).
Die Wahrscheinlichkeit von B, bezeichnet als P(B).
Die bedingte Wahrscheinlichkeit von B, wenn A bereits passiert ist, bezeichnet als P(B|A).

Zum Beispiel: Wir möchten die Wahrscheinlichkeit von A ermitteln, wenn B bereits eingetreten ist. Das heißt, das Ereignis A ist abhängig von B. B muss vorher passieren, damit A passieren kann. Man setzt voraus, dass man die Wahrscheinlichkeiten von B,A und die Wahrscheinlichkeit für B, wenn A bereits passiert ist p(B|A), kennt. Dann kann man die Formel umstellen, wodurch man die Wahrscheinlichkeit für A ermitteln kann, wenn B vorher passiert.

Es ist eine Methode zur Ermittlung von Wahrscheinlichkeiten basierend auf bereits vorhandenen Wahrscheinlichkeiten und dem Zusammenhang zwischen verschiedenen Ereignissen.

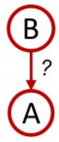
Frage: In welche zwei Gruppen werden Wahrscheinlichkeiten unterteilt?

1. **A-Priori-Wahrscheinlichkeiten**: Dies sind anfängliche Wahrscheinlichkeiten, die Sie basierend auf Ihrem vorhandenen Wissen oder allgemeinem Verständnis einschätzen, bevor Sie spezifische Informationen oder Daten erhalten.
Beispiel: Stellen Sie sich vor, Sie sind in einer Stadt, in der es normalerweise im Sommer selten regnet. Wenn Sie ohne aktuelle Wetterdaten zu überprüfen schätzen, dass die Wahrscheinlichkeit, dass es morgen regnet, sehr niedrig ist, basieren Ihre Schätzungen auf A-Priori-Wahrscheinlichkeiten, da sie auf Ihrem allgemeinen Wissen über das Klima in dieser Stadt beruhen.
"A priori" bedeutet, dass die Wahrscheinlichkeit im Voraus berechnet wird, bevor neue Informationen oder Daten verfügbar sind. Sie basiert auf dem, was bereits bekannt oder angenommen wird, bevor weitere Beobachtungen gemacht werden.
> Wahrscheinlichkeit von Ereignissen, bei denen keine Informationen über den Prädiktor berücksichtigt wurden.

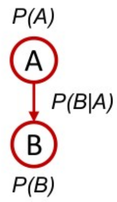
2. **A Posteori Wahrscheinlichkeit**: Die a posteriori Wahrscheinlichkeit ist die Wahrscheinlichkeit eines Ergebnisses, nachdem wir Informationen über den Prädiktor (eine Variable oder einen Faktor, der das Ergebnis beeinflusst) berücksichtigt haben. Es ist also die aktualisierte Wahrscheinlichkeit, nachdem wir bereits gewisse Informationen über den Prädiktor kennen. Wissenstandermittlung anhand von durchgeführten Berechnungen.Dies sind Wahrscheinlichkeiten, die aktualisiert werden, nachdem Sie neue Informationen oder Daten erhalten und Berechnungen durchgeführt haben.
Beispiel: Angenommen, Sie überprüfen die aktuellen Wetterdaten für Ihre Stadt und stellen fest, dass es morgen eine Gewitterwarnung gibt. Nachdem Sie diese neuen Informationen erhalten haben, passen Sie Ihre Wahrscheinlichkeit an, dass es morgen regnet, an. Ihre Aktualisierung basiert auf den neuen Daten und könnte bedeuten, dass die Wahrscheinlichkeit jetzt höher ist. Diese aktualisierte Wahrscheinlichkeit ist ein Beispiel für A-Posteriori-Wahrscheinlichkeit, da sie auf den neuen Daten und Berechnungen basiert.
"A posteriori" hingegen bedeutet, dass die Wahrscheinlichkeit nach Berücksichtigung neuer Informationen oder Daten berechnet wird. Sie beruht auf den ursprünglichen Annahmen (a priori) sowie den neuen Beobachtungen.
> Bedingte Wahrscheinlichkeiten sind a Posteori (p(A|B)).

- Ziel:
  Bestimmung der bedingten Wahrscheinlichkeit zweier Ereignisse, sofern eine der beiden bedingten Wahrscheinlichkeiten bekannt ist

- Formel:



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



- $P(A)$ ist die A-Priori Wahrscheinlichkeit von A
- $P(B)$ ist die A-Priori Wahrscheinlichkeit von B
- $P(A|B)$ ist die bedingte Wahrscheinlichkeit von A, falls B bereits eingetreten ist
- $P(B|A)$ ist die bedingte Wahrscheinlichkeit von B, falls A bereits eingetreten ist

Wir möchten also die bedingte Wahrscheinlichkeit von zwei Ereignissen bestimmen, sofern einer einer bedingten Wahrscheinlichkeiten bekannt ist. In der Abbildung rechts sehen wir, dass das Ereignis B von A abhängt p(B|A). Das heißt, wir haben die Wahrscheinlichkeit von A und von B und die Wahrscheinlichkeit von B, wenn A bereits passiert ist. Das wissen wir, aber wir wollen den umgekehrten Fall sehen – was ist die Wahrscheinlichkeit von A, wenn B bereits passiert ist? Wir drehen also die bedingten Beziehungen um.

Frage: Was ist in diesem Beispiel eine A-Priori-Wahrscheinlichkeit und was ist A-Posteori Wahrscheinlichkeit?

A-Priori-Wahrscheinlichkeiten sind p(A) und p(B) – also die Wahrscheinlichkeiten für A und B alleine.

A-Posteori Wahrscheinlichkeit sind p(A|B) und p(B|A). Die bedingten Wahrscheinlichkeiten sind A-Posteriori-Wahrscheinlichkeiten.

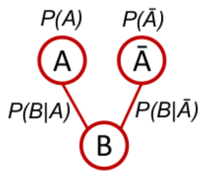
- <u>Ziel:</u>
  Berechnung über das Komplement Ā (A ist <u>nicht</u> eingetreten)

- <u>Formel:</u>

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

- daraus folgt, dass: $\boldsymbol{P(B) = P(B|A) * P(A) + P(B|\bar{A}) * P(\bar{A})}$

- die Wahrscheinlichkeit von B setzt sich zusammen aus:
  – der Wahrscheinlichkeit von B in Zusammenhang mit A
  – der Wahrscheinlichkeit von B in Zusammenhang mit Ā
  → der Berechnung über die Elternknoten (verallgemeinert)



Frage: Was ist das Komplement (zu A)?

Das Komplement eines Ereignisses A in der Bayes'schen Wahrscheinlichkeitstheorie bezieht sich darauf, dass das Ereignis A nicht eingetreten ist. Das bedeutet, das Komplement von A, oft als A' oder "nicht A" bezeichnet, umfasst alle möglichen Ergebnisse, bei denen A nicht auftritt.

Die Berechnung kann auch über das Komplement erfolgen, was bedeutet, dass die bedingte Wahrscheinlichkeit des Eintretens von B, wenn es von A abhängt, auf zwei mögliche Weisen geschehen kann: Entweder **B hängt von A ab und A ist bereits eingetreten**, oder **B hängt von A ab, doch A ist noch nicht eingetreten**. Im unteren rechten Teil sehen Sie eine graphische Darstellung des Komplements. Ich habe also die Wahrscheinlichkeit, dass B passiert, wenn A bereits passiert ist p(B|A) und die Wahrscheinlichkeit für B, wenn A nicht passiert ist p(B|Ā). Beides kann ich der Formen verwenden.
> **Merke: Wenn ein Ereignis (B) von einem Ereignis (A) bedingt wird, dann ist es auch abhängig von seinem Kompliment (Nicht A)**.

Beispiel:

Ein Beispiel für diese Beziehung zwischen einem Ereignis (B), einem Ereignis (A) und seinem Komplement (Nicht A) könnte sich auf das Wetter und das Ereignis "Regen" beziehen.
Angenommen, wir haben die folgenden Ereignisse:

A: Es ist ein Wochentag.
B: Es regnet.
Jetzt, wenn wir wissen, dass das Ereignis B (Regen) von dem Ereignis A (Wochentag) abhängt, d.h., es regnet typischerweise an Wochentagen, dann ist es auch abhängig von seinem Komplement, das "Nicht A" oder das Ereignis "Wochenende" ist.

In diesem Beispiel:

"A" repräsentiert Wochentage (Montag bis Freitag).
"Nicht A" oder "Wochenende" repräsentiert Samstag und Sonntag.

Wenn es regnet (B), dann hängt es typischerweise von Wochentagen (A) ab. Aber es hängt auch von "Nicht A" (Wochenenden) ab, da es an Wochenenden ebenfalls regnen kann.

Frage: Wann macht es zum Beispiel Sinn die Rechnung über das Kompliment zu machen? **Wieso brauchen wir überhaupt das Kompliment**?

Wie man der Abbildung entnehmen kann, sind die A priori-Wahrscheinlichkeiten p(A) und p(B), wenn sie von einander abhängen, auch von ihrem Kompliment abhängig (das sieht man in der Graphik unten rechts). Man kann also über die Formel und die Komplimente die Wahrscheinlichkeit von p(A) und p(B) ausrechnen, wenn diese nicht bekannt sind.

Frage: Wie verhalten sich A und B z ihrem Kompliment?

A und -A müssen in der **Summe 100 % ergeben**.
Wenn A dafür steht, dass ein Patient seine Medikamente eingenommen hat und er hat an 70 % aller Tage seine Medikamente eingenommen, dann muss -A (hat Medikamente nicht eingenommen) 30 % sein.

Frage: Wie rechnet man P(A|B) aus, wenn P(B) nicht gegeben ist?

Die Berechnung der Wahrscheinlichkeit von A, wenn B bereits eingetreten ist, erfolgt durch Multiplikation der bedingten Wahrscheinlichkeit von B, wenn A eingetreten ist, mit der Wahrscheinlichkeit von A. Der Zähler in dieser Berechnung bleibt im Vergleich zur vorherigen Abbildung unverändert. Was sich ändert, ist die Wahrscheinlichkeit für B (p(B)). Die Wahrscheinlichkeit für **B wird jetzt in die beiden Zweige gespalten**. Das bedeutet, die Wahrscheinlichkeit, dass B eingetreten ist, ergibt sich aus der Kombination zweier Teile: Erstens die Wahrscheinlichkeit, dass B eingetreten ist, wenn A bereits passiert ist, multipliziert mit der Wahrscheinlichkeit für A. Zweitens die Wahrscheinlichkeit, dass B eingetreten ist, wenn A nicht passiert ist, multipliziert mit der Wahrscheinlichkeit, dass A nicht passiert ist. Beide Teile werden addiert.

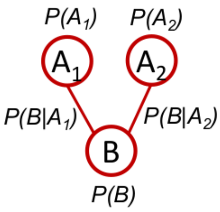
Das ist die Berechnung über das Komplement. Wenn wir mehrere Zweige oder tiefere Zweige haben ist das die Berechnung über die Elternknoten. In solchen Modellen repräsentieren "Elternknoten" die übergeordneten Ereignisse oder Faktoren, die die Wahrscheinlichkeit eines gegebenen Ereignisses beeinflussen. Die Wahrscheinlichkeit des gegebenen Ereignisses setzt sich aus den Wahrscheinlichkeiten in Bezug auf diese übergeordneten Ereignisse oder "Elternknoten" zusammen. Wenn B also von mehr abhängt als A (zum Beispiel A, C,D …), dann kann man die Formel um weitere Zweige erweitern.

- <u>Ziel:</u>
  Berechnung über die Elternknoten (Verbundregel)

- <u>Formel:</u>

$$P(B) = P_{verbund}(B) = \sum_i P(A_i) * P(B|A_i)$$

- also ist die Wahrscheinlichkeit B die Summe aus allen Wahrscheinlichkeiten von B in Zusammenhang mit den Elternknoten



$$P(B) = P(B|A_1) * P(A_1) + P(B|A_2) * P(A_2)$$

Frage: Was wird mit der Verbundregel berechnet und wie lautet die Formel?

Die Abbildung zeigt die Formel für die Berechnung über die Elternknoten. Dies ist auch als die Verbundregel bekannt. Wenn die Wahrscheinlichkeit eines Ereignisses B von zwei verschiedenen Ereignissen, A1 und A2, abhängt, können wir die Verbundregel verwenden. Die Berechnung ähnelt dem Konzept des Komplements, ist jedoch flexibel genug, um mehrere Ereignisse (von A1 bis An) zu berücksichtigen, anstatt nur zwei (Die Formel aus für das Kompliment ergibt sich aus der Verbundregel). **Über die Verbundregel kann man die Wahrscheinlichkeiten für alle A priori Wahrscheinlichkeiten berechnen, die von einer oder mehrere Wahrscheinlichkeiten abhängig sind**.

Zusammenfassung/ Formelsammlung

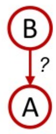
Frage: Wozu dient Bayes Theorem?

Bei Bayes geht es darum zu beobachten, wie sich die Wahrscheinlichkeit eines Ereignisses ändert, wenn sich eine andere Wahrscheinlichkeit, die Einfluss auf dieses Ereignis hat, geändert hat. man aktualisiert also seine Wahrscheinlichkeit.

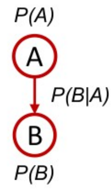
Formeln:

1. Berechnung von bedingten Wahrscheinlichkeiten:

- <u>Formel:</u>



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



- $P(A)$ ist die A-Priori Wahrscheinlichkeit von A
- $P(B)$ ist die A-Priori Wahrscheinlichkeit von B
- $P(A|B)$ ist die bedingte Wahrscheinlichkeit von A, falls B bereits eingetreten ist
- $P(B|A)$ ist die bedingte Wahrscheinlichkeit von B, falls A bereits eingetreten ist

2. Berechnung von Bedingten Wahrscheinlichkeit über das Kompliment:

- <u>Ziel:</u>
  Berechnung über das Komplement Ā (A ist <u>nicht</u> eingetreten)

- <u>Formel:</u>

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

- daraus folgt, dass: $\boldsymbol{P(B) = P(B|A) * P(A) + P(B|\bar{A}) * P(\bar{A})}$

3. Berechnung für A Priori Wahrscheinlichkeiten für beliebig viele Zweige:

$$P(B) = P_{verbund}(B) = \sum_i P(A_i) * P(B|A_i)$$

Beispiel aus der Vorlesung

Ein Patient nimmt sein Medikament an 70% aller Tage ein. An 80% dieser Tage geht es dem Patienten gut. Durchschnittlich geht es dem Patienten nur an 60% aller Tage gut.

Heute geht es dem Patienten gut. Wie hoch ist die Wahrscheinlichkeit, dass er sein Medikament eingenommen hat?

A	Der Patient hat sein Medikament eingenommen
$\bar{A}$	Der Patient hat sein Medikament <u>nicht</u> eingenommen
B	Dem Patienten geht es gut
$\bar{B}$	Dem Patienten geht es <u>nicht</u> gut

$P(A)$	Die Wahrscheinlichkeit, dass der Patient sein Medikament eingenommen hat
$P(B)$	Die Wahrscheinlichkeit, dass es dem Patienten gut geht
$P(B\|A)$	Die Wahrscheinlichkeit, dass es dem Patienten gut geht, wenn dieser sein Medikament eingenommen hat (gegeben)
$\boldsymbol{P(A\|B)}$	Die Wahrscheinlichkeit, dass der Patient sein Medikament eingenommen hat, wenn es diesem gut geht **(gesucht)**

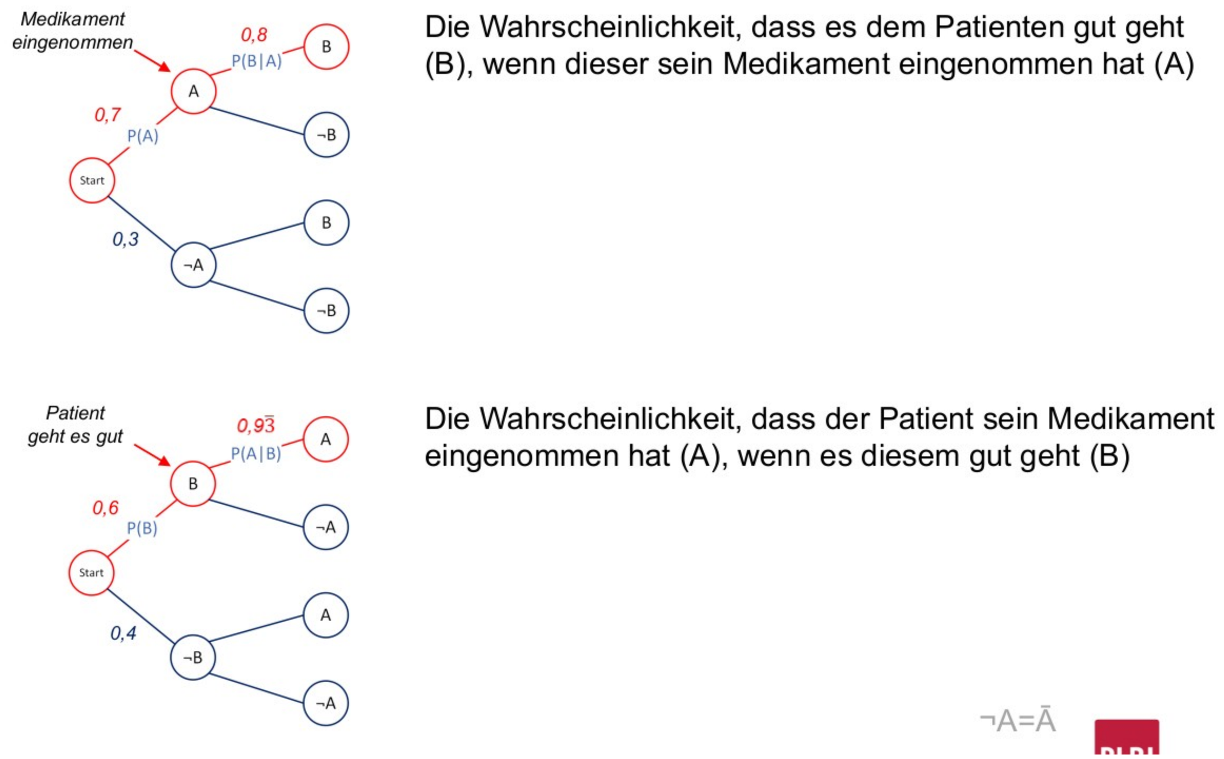
Gegeben ist folgendes Beispiel für einen Patienten, der Medikamente einnimmt. P(A|B) ist gesucht. Abbildung demonstriert sehr gut, was das Kompliment ist. Beachte: P(B|A) ist gegeben als 80%. Wir wollen es aber umstellen. Wir wissen, dass es ihm an 80% der Tage, an denen er die Medikamente eingenommen, gut geht, aber wir wollen wissen, was die Wahrscheinlichkeit ist, dass er sein Medikament eingenommen hat.

$P(A) = 0{,}7$	zu 70% hat der Patient sein Medikament eingenommen
$P(B) = 0{,}6$	zu 60% geht es dem Patienten (insgesamt) gut
$P(B\|A) = 0{,}8$	zu 80% geht es dem Patienten nach Medikamenteneinnahme gut
$\boldsymbol{P(A\|B)} = ?$	Die Wahrscheinlichkeit der Medikamenteneinnahme, unter der Bedingung, dass es dem Patienten gut geht **(gesucht)**

$$\boldsymbol{P(A|B)} = \frac{P(B|A) \, * \, P(A)}{P(B)} = \frac{0{,}8 \, * 0{,}7}{0{,}6} = 0{,}9\overline{3} = \mathbf{93{,}33}\%$$

Mit Hilfe des Satzes von Bayes konnte die Schlussfolgerung umgekehrt werden. Aus den gegebenen Wahrscheinlichkeiten wurde die umgekehrte Wahrscheinlichkeit abgeleitet.

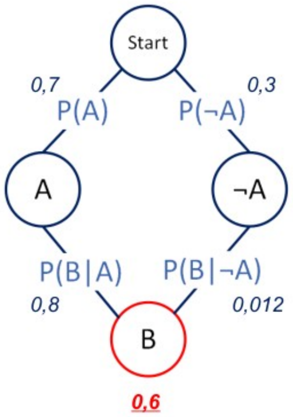
Wir können die gegebenen Daten in die erste Formel eintragen. Wir erhalten 93.33 % für P(A|B). Wir wissen, dass er mit einer Wahrscheinlichkeit von 93.33 % die Medikamente eingenommen hat an den Tagen, an denen es ihm gut ging. Das heißt, wenn es ihm gut geht, gibt es eine 93 % Wahrscheinlichkeit, dass er gerade sein Medikament benutzt. Wir konnten also dank Bayes die bedingte Wahrscheinlichkeit p(B|A) umdrehen und p(A|B) ausrechnen.



Die Abbildung zeigt das Ganze nochmal graphisch, wobei das zweite Netzwerk das darstellt, was wir hier ausgerechnet haben (p(A|B)).

Frage: Ist p(A|B) das Kompliment zu p(B|A)? Anders gefragt: Wenn p(B|A) =80 % ist, ist p(A|B) =20 %?

Nein. Wir sehen, dass man Sie einzeln ausrechnen muss. Man kann aus dem einem nicht das andere leicht herleiten. 1- **p(B|A) =/= (A|B) !**
Aber 1-p(B|A) = p(B| -A) für 2 Werte → Die jeweilige Verzweigung muss 100 % ergeben.

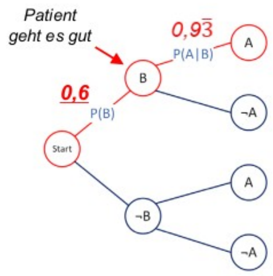


Die Berechnung von *P(B)* über den Elternknoten:

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

0,6 = 0,8 * 0,7 + x * 0,3
x = 0,012 (nach Umformung)

P(B) = 0,8 * 0,7 + 0,012 * 0,3 = 0,6



Die Berechnung von *P(A|B)* über das Komplement:

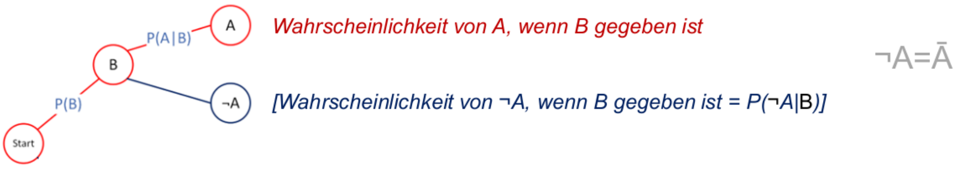
$$P(A|B) = \frac{P(B|A) * P(A)}{P(B|A) * P(A) + P(B|\bar{A}) * P(\bar{A})}$$

$$P(A|B) = \frac{0{,}8 * 0{,}7}{0{,}8 * 0{,}7 + 0{,}012 * 0{,}3} = 0{,}9\overline{3}$$

Jetzt das Ganze nochmal durch die Berechnung über den Elternknoten. Ich habe die Wahrscheinlichkeit von B und ich möchte es jetzt nach P(B|-A) umformen. Ich erhalte 0,012 und kann es in die Formel für das Komplement einsetzen und erhalte dasselbe Ergebnis.

So haben wir bewiesen, dass die Formeln funktionieren.

- Grafisches Modell zur Visualisierung von Beziehungen und Kausalzusammenhängen
- Ermöglicht bestmögliche Vorhersage (Wahrscheinlichkeit) von Folgen bestimmter Handlungen unter Nutzung aller vorhandener Informationen
- Die Berechnungsgrundlage bilden vorausgegangene Ereignisse



- Die Beziehungen stehen nicht unbedingt für ein direktes Verhältnis von Ursache und Wirkung (Stärke der Beziehung Beachten!)
- BN sind robust und bieten einen effizienten Ansatz gegen Overfitting

Im Kontext von Bayes taucht oft das sogenannte Bayes'sche Netzwerk auf. Dies ist ein graphisches Modell, das dazu dient, Beziehungen und kausale Zusammenhänge zu visualisieren. Es ermöglicht die bestmögliche Vorhersage von Ergebnissen oder Wahrscheinlichkeiten bestimmter Handlungen unter Nutzung aller verfügbaren Informationen. Das bedeutet, wir bauen das Netz auf. Wir starten in er Abbildung mit der Wahrscheinlichkeit für B. Wir laufen das Netz von B aus durch. So sehen wir dann zum Beispiel, dass die weiter unter stehenden Zweige – wie die Wahrscheinlichkeit für A und Nicht A (Komplement) – von B abhängen (Wahrscheinlichkeit von A ändert sich, je nachdem ob B passiert ist oder nicht). Wenn die Netze größer werden, kann man Sie durchgehen und sehen, welche Wahrscheinlichkeiten von welchen anderen Wahrscheinlichkeiten abhängig sind – welches Ereignis beeinflusst ein anderes Ereignis?

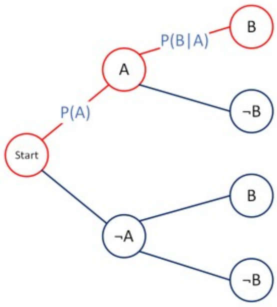
Frage: Kann ich basierend auf dem Netzwerk direkt auf Ursache-Wirkung Beziehungen schließen?

Es ist wichtig zu beachten, dass die in einem Bayes'schen Netz dargestellten Beziehungen nicht zwangsläufig eine direkte Ursache-Wirkungs-Beziehung darstellen. In solchen Netzwerken gibt es sowohl starke als auch schwache Beziehungen zwischen Ereignissen. Das bedeutet, dass nur die starken Beziehungen einen signifikanten Zusammenhang zwischen den Ereignissen anzeigen.

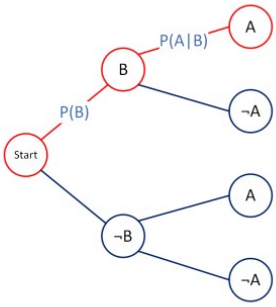
Zum Beispiel, in einem Bayes'schen Netzwerk zur medizinischen Diagnose, könnten Symptome und Krankheiten miteinander in Beziehung stehen. Ein starkes Beziehungspaar könnte Husten und Grippe sein, da Husten ein häufiges Symptom der Grippe ist. In diesem Fall zeigt die starke Beziehung einen deutlichen Zusammenhang zwischen dem Symptom Husten und der Krankheit Grippe.
Auf der anderen Seite könnten Symptome wie Kopfschmerzen und Halsschmerzen im Netzwerk schwach miteinander verbunden sein, da diese Symptome bei verschiedenen Krankheiten auftreten können und daher keinen starken Zusammenhang aufweisen.

Frage: Was sind die 2 Vorteile von Bayes Netzwerken?

Die Vorteile von Bayes Netzwerken sind, dass Sie
1. Sehr robust sind
2. Sie bieten einen effizienten Ansatz gegen Overfitting.



Die Wahrscheinlichkeit von B, wenn A gegeben ist



Die Wahrscheinlichkeit von A, wenn B gegeben ist

¬A=Ā

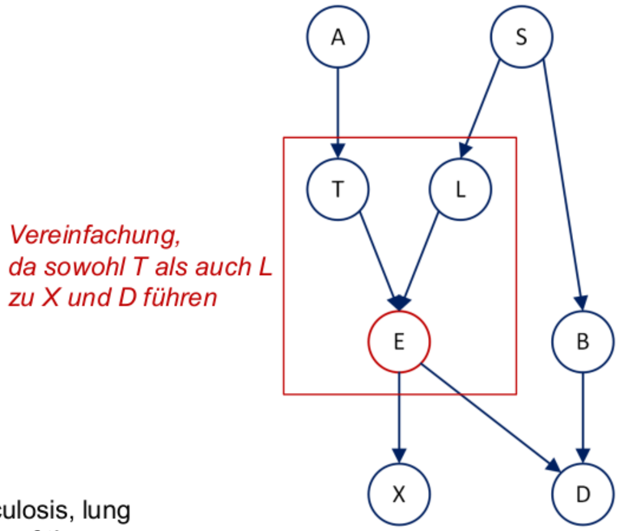
In der Abbildung sehen Sie ein gesamtes Netz. Das soll vereinfacht darstellen, wie ein Netz aussehen würde für die Ereignisse A und B. Als erstes sehen Sie die Wahrscheinlichkeit von B, wenn A gegeben ist. Zuerst muss A gegeben sein (p(A) als erste Edge und A als erster Note) und wenn A gegeben ist, führt es zu B. Wobei man auch sieht, dass es nicht nur zu B oder A, sondern auch zu ihrem Kompliment führen kann. Aus A kann also B oder -B resultieren. Deshalb kann ich auch das Kompliment für die Berechnung der bedingten Wahrscheinlichkeiten miteinbeziehen.

ASIA-Datensatz

Der Datensatz wurde künstlich erzeugt.

<u>Er enthält die folgenden Variablen:</u>

- D: dysnopea (yes/no)
- T: tuberculosis (yes/no)
- L: lung cancer (yes/no)
- B: bronchitis (yes/no)
- A: visit to Asia (yes/no)
- S: smoking (yes/no)
- X: chest X-ray (yes/no)
- *E: T ODER L (yes/no)*



<u>Motivation:</u>

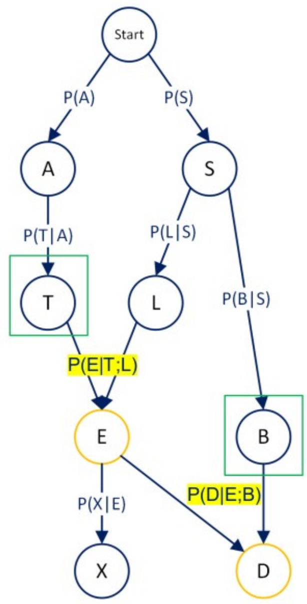
"Shortness-of-breath (dyspnoea) may be due to tuberculosis, lung cancer or bronchitis, or none of them, or more than one of them. A recent visit to Asia increases the chances of tuberculosis, while smoking is known to be a risk factor for both lung cancer and bronchitis. The results of a single chest X-ray do not discriminate between lung cancer and tuberculosis, as neither does the presence or absence of dyspnoea."

Einer der ersten Veröffentlichungen über das Bayes-Netz kommt aus dem Jahr 1988. Damals wurde ein künstlich erschaffener Datensatz verwendet: der ASIA Datensatz. Dieser Datensatz enthält Informationen über verschiedene medizinische Variablen und Ereignisse, darunter Dyspnoea (Atembeschwerden), Tuberkulose, Lungenkrebs, Bronchitis, Besuch in Asien, Rauchen, Röntgenbild der Brust und die Unterscheidung zwischen Tuberkulose und Lungenkrebs/Bronchitis.

D (dyspnoea), binary 1/0 corresponding to "yes" and "no"
T (tuberculosis), binary 1/0 corresponding to "yes" and "no"
L (lung cancer), binary 1/0 corresponding to "yes" and "no"
B (bronchitis), binary 1/0 corresponding to "yes" and "no"
A (visit to Asia), binary 1/0 corresponding to "yes" and "no"
S (smoking), binary 1/0 corresponding to "yes" and "no"
X (chest X-ray), binary 1/0 corresponding to "yes" and "no"
E (tuberculosis versus lung cancer/bronchitis), binary 1/0 corresponding to "yes" and "no"

D ist hier die Response/ Abhängige Variable!

In der Abbildung sehen wir das Netzwerk für den Datensatz. Tuberculosis (T) hat hier eine Beziehung zu einem vorherigen Aufenthalt in Asien (A). Lungenkrebs (L) hat eine direkte Beziehung zum Rauchen (S). Bronchitis (B) hat eine Beziehung zum Rauchen (S). Atemnot (D) hat eine Beziehung zum Bronchitis (B). E ist eine Vereinfachung für Lungenkrebs und Tuberkolose. Das heißt, wenn einer von beiden wahr ist, dann ist auch E wahr. Das wurde vereinfacht, weil ansonsten sowohl T als auch L jeweils zu X und D führen würde (unübersichtlich). Wir führen mit den Datensatz eine Beispielberechung durch. Vorher hatten wir nur 2 Variablen A und B. Hier verdeutlichen wir es anhand vieler Variablen, die alle In Beziehung stehen.



$P(A) \approx 0.01$; $P(\neg A) \approx 0.99$
$P(S) \approx 0.50$; $P(\neg S) \approx 0.50$
$P(T|A) \approx 0.05$; $P(T|\neg A) \approx 0.01$
$P(L|S) \approx 0.10$
$P(E|T;L) \approx 1.00$
$P(B|S) \approx 0.60$; $P(B|\neg S) \approx 0.30$
$P(X|E) \approx 0.98$
$P(D|E;B) \approx 0.90$

Diese Werte können aus dem Bayes-Netz (oder direkt aus dem Datensatz) abgelesen bzw. ermittelt werden

Mit welcher Wahrscheinlichkeit erkrankt ein Patient jedoch gleichzeitig an Tuberkulose und Bronchitis?

Verbundregel:

$$P(T,B) = P_{verbund}(T) * P_{verbund}(B)$$
$$= \left(\sum_A P(A) * P(T|A)\right) * \left(\sum_S P(S) * P(B|S)\right)$$
$$= (0.01 * 0.05 + 0.99 * 0.01) * (0.50 * 0.60 + 0.50 * 0.30) = 0.00468$$

Berechnung über die Elternknoten!

Wir fangen oben im Netz an und wir haben oben zwei Zweige: Asienaufenthalt und Rauchen. In der Abbildung stehen die Werte. Man sieht, dass 99 % der Leute nicht in Asien waren (-A) und 1 % war es. Rauchen und Nicht-Raucher haben ein 50/50 Verhältnis. Die Wahrscheinlichkeit, dass Tuberkulose auftrat, wenn die Person in Asien war (P(T|A)), beträgt 5 % und 1 %, wenn Sie nicht in Asien war. Hier muss aber bedacht werden, dass es sich um einen sehr unausgewogenen Datensatz handelt, da 99 % der Personen nicht in Asien waren. Wie schon vorher erwähnt: Nicht alle Pfeile sind gleich. **Es gibt stärkere und schwächere Beziehungen**!
Die Wahrscheinlichkeit Lungenkrebs zu haben, wenn man Raucher ist (p(L|S)), liegt bei 10 %. Die Wahrscheinlichkeit eine Bronchitis zu haben, wenn man raucht (p(B|S)), liegt bei 60 %. Nicht Raucher haben eine Wahrscheinlichkeit von 30 %. Der Datensatz war ausgewogen in Bezug auf Raucher und Nicht-Raucher (50/50). Deshalb ist die Beziehung zwischen Bronchitis und Rauchen stärker zu bewerten.
Aus Lungenkrebs oder Tuberkulose und Bronchitis ergibt sich in 90 % der Fälle eine Atemnot (p(D|E,B)).

Mit welcher Wahrscheinlichkeit erkrankt ein Patient jedoch gleichzeitig an Tuberkulose und Bronchitis?
Wir können hierfür die Verbundregel nutzen. Tuberkolosis hängt vom Asienaufenthalt ab und Bronchitis vom Rauchen. Wir setzen beides in die Regel ein. In der Formel haben wir die Wahrscheinlichkeit für Tuberkolosis, wenn Asienaufenthalt stattfand oder nicht, und Bronchitis, wenn man Raucher oder Nicht-Raucher ist.

$$P(T,B) = (P(A)* P(T|A)+ P(-A)* P(T|-A)) * (P(S)* P(B|S)+ P(-S)* P(B|-S))$$

$$=(0.01 * 0.05 + 0.99 * 0.01) * (0.50 * 0.60 + 0.50 * 0.30) = 0.00468$$

```python
# import packages
import pandas as pd
from sklearn import preprocessing
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB

# import data
dataset = pd.read_csv("asia.csv")

# preprocess data (converting string labels into numbers)
le = preprocessing.LabelEncoder()
dataset[['A', 'S', 'T', 'L', 'B', 'E', 'X', 'D']] = \
    dataset[['A', 'S', 'T', 'L', 'B', 'E', 'X', 'D']].apply(le.fit_transform)

# split into training and test dataset
samples = dataset.drop('D', axis=1)
target = dataset.D

X_train, X_test, y_train, y_test \
    = train_test_split(samples, target, test_size = 0.2, train_size = 0.8)

# train classifier
bn_clf = GaussianNB()
bn_clf.fit(X_train, y_train)

# test classifier
y_pred = bn_clf.predict(X_test)

# performance measure
print(bn_clf.score(X_test, y_test)*100)
```

Vorgehen:

1. Datensatz einlesen
2. Kategorien (yes / no) in Zahlenwerte (1 / 0) formatieren
3. Datensatz in Trainingsdatensatz und Testdatensatz aufteilen
4. Einen Bayes-Klassifikator (hier: Naive Bayes) auswählen
5. Den Klassifikator trainieren
6. Den Testdatensatz auf den trainierten Klassifikator anwenden
7. Die Genauigkeit berechnen

8. Das Ergebnis bewerten!

Die Abbildung zeigt, wie man Bayes Netz und den ASIA-Datensatz in Python implementiert.
le.fit_transform wandelt die Yes und No Werte in numerische Werte (1 und 0) um.

Naive-Bayes Klassifikator

- Naive Annahme:
  - Alle Attribute sind gleich wichtig
  - Alle Attribute sind voneinander unabhängig

- Die naive Annahme stimmt in der Realität nie!
- Die naive Annahme funktioniert jedoch ziemlich gut

- Vorgehen:
  - Es werden alle Attribute verwendet
  - Häufigkeitstabelle aus den Daten erstellen
  - Wahrscheinlichkeit für die Zugehörigkeit zur Klasse errechnen
  - $P(A|B) = \frac{P(B_1|A)*P(B_2|A)*\cdots*P(B_n|A)*P(A)}{P(B)} = \frac{\prod_i P(B_i|A)*P(A)}{P(B)}$
  - Fehlende Werte werden bei der Berechnung weggelassen

Frage: Ist der Naive Bayes Klassifikator (und andere Bayes basierte Klassifikatoren) unsupervidiert oder supervidiert?

Der naive Bayes-Klassifikator (und andere) ist ein Supervised Learning-Algorithmus. Man muss für das Modellieren eine Response-Variable/ Labels mit übergeben.

Jetzt schauen wir uns die einzelnen Bayes-Klassifikatoren an. Es gibt den Naive Bayes-Klassifikator. Weiterhin gibt es das Bayes-Netz und es gibt einen Zwischenfall.

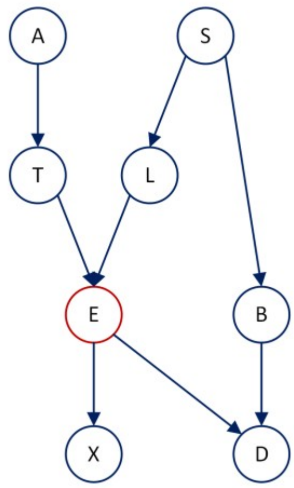
Frage: Wodurch unterscheidet sich Naive Bayes vom Netzwerk? **Welche Annahmen werden gemacht**?

Beim Netzwerk hatten wir Punkte, die alle eine Abhängigkeit zueinander aufwiesen. Bei dem Naive Bayes nehmen wir eine Vereinfachung vor. Wir gehen davon aus, dass alle Attribute/Variablen/Ereignisse nur von der Zielvariable abhängig sind. Ich habe also nur Verbindungen der einzelnen Punkte (unabhängigen Variablen) zur Zielvariable (Response). Die Variablen sind nicht mehr untereinander verbunden.Sie sind alle voneinander unabhängig und Sie haben nur eine Beziehung zur Responsevariable. Es gibt kein Netzwerk mehr (Deshalb ist es wohl die „Naive“ Variante ^^).
Zusammenfassung: **Es besteht nur eine Beziehung zwischen Response und unabhängigen Variablen**.

Die zweite Annahme ist, dass **alle Variablen gleich wichtig** sind. Das heißt, jede Variable hat denselben Einfluss auf die Zielvariable. Das stimmt in der Realität nie, aber die Annahme funktioniert erstaunlich gut.

Frage: Wie geht der Naive Bayes vor? **Was wird berechnet**?

Es verwendet alle Variablen/ Attribute. Der Algorithmus berechnet die **Wahrscheinlichkeit, dass eine gegebene Beobachtung zu einer bestimmten Klasse gehört**. Dabei werden die Werte der Attribute genutzt, um diese Zuordnungswahrscheinlichkeit zu ermitteln. Dazu werden die einzelnen Wahrscheinlichkeiten berechnet. Das heißt, ich berechne jede einzelne Wahrscheinlichkeit von jedem einzelnen Ereignis für das Zielereignis (Responsevariable).



Beispiel:
Die Zielvariable ist D [Klassen: yes / no]
Alle anderen Variablen werden als unabhängig voneinander betrachtet

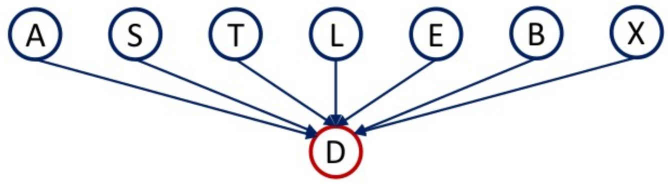
Die Formel wäre (für die Klasse D):

$$P(yes|D) = \frac{P(A|yes) * P(S|yes) * P(T|yes) * P(L|yes) * P(E|yes) * P(B|yes) * P(X|yes) * P(yes)}{P(D)}$$

$$P(no|D) = \frac{P(A|no) * P(S|no) * P(T|no) * P(L|no) * P(E|no) * P(B|no) * P(X|no) * P(no)}{P(D)}$$

Durch Normierung mittels *P(yes|D) + P(no|D) = 1* fällt *P(D)* weg.

Zuordnung zur Klasse D_{yes}, wenn $P(yes|D) > P(no|D)$ und umgekehrt.

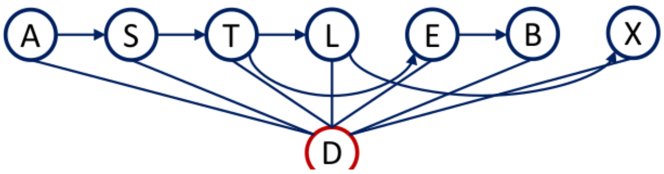


In der Abbildung soll der Unterschied zwischen dem Netz und der vereinfachten Annahme von Naive Bayes dargestellt werden. Links sieht man das Netz und man sieht, dass die unabhängigen Variablen noch Beziehungen untereinander haben und unten sieht man die Annahme des Naive Bayes graphisch dargestellt. Man sieht, dass nur noch Beziehungen zur Response-variable D gibt.Ich möchte jetzt für jede Variable ausrechnen, wie hoch die Wahrscheinlichkeit ist, dass Sie zur Klasse „Atemnot“ oder „keine Atemnot“ gehört. Das mache ich unabhängig voneinander. Wie hoch ist zum Beispiel die Wahrscheinlichkeit von A (Asienaufenthalt) für D. Ich kriege dann für jede Variable einen Prozentwert und ich multipliziere alle Prozentwerte von jeder Variable zusammen.
Ich rechne bei Naive Bayes die Wahrscheinlichkeiten für die einzelnen Klassen getrennt. D=Yes wird einzeln ausgerechnet und D=No wird auch einzeln betrachtet. Dann schaue ich mir an, was häufiger vorkommt. Ich schaue mir jetzt an, was die Mehrheit pro Zeile/ Observation gewinnt. Wenn ein Patient in der Mehrheit für D yes = 60 % und für no = 40 % bekommt, dann packe ich Patienten 1 in die Yes Gruppe und sage voraus, dass er Atemnot hat. Das mache ich für alle Observationen /Patienten. **P(D) würde aus der Formel oben rausfallen, weil es in der Summe =1 ist. Ich behalte nur noch die Werte im Zähler und das ist die große Vereinfachung der Formel bei Naive Bayes**.

Es existieren folgende Varianten:

- **Gaussian** Naive Bayes (Annahme der Normalverteilung)
- **Multinomial** Naive Bayes (Annahme der Multinomialen Verteilung)
- **Bernoulli** Naive Bayes (Annahme der binären Variablen)
- **Complement** Naive Bayes (bei ungleichen Verteilungen)
- **Categorical** Naive Bayes (bei kategorischen Daten)

- **Tree-augmented** Naive Bayes (TAG)
  – Semi-Naive Bayes
  – Jedes Attribut hängt von der Klasse ab
  – Jedes Attribut hängt von einem anderen Attribut ab



Frage: Welche Varianten des Naive Bayes gibt es und welche Annahmen machen Sie? 6 Arten

Um den Naive Bayes auszurechnen, gibt es verschiedene Varianten.

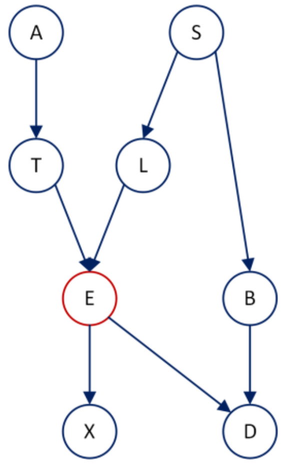
Gaussian macht die Annahme, dass die Werte Normalverteilt sind.
Multinominal: Die Daten sind Multinominal verteilt.
Bernoulli: Daten sind binär verteilt
Complement: Daten sind ungleich verteilt
Categorical: Daten sind kategorisch (Mann/Frau z.B.)

Tree-augmented Naive Bayes: Das ist der Semi-Ansatz (zwischen Netzwerk und Naive). Hier hängen die Variablen nicht nur vom Label/ response ab, sondern auch noch von einer anderen Variable – ABER: Es **hängt NUR von EINER anderen Variable ab**. Nicht mehr. Dadurch ist es genauer, aber es hat den Nachteil, dass ein Attribut eine sehr starke Beziehung zu mehreren Variablen haben kann. Dann würde die Variable die stärkste Beziehung auswählen, auch wenn die anderen Variablen fast so starke Beziehungen aufweisen. Alternativ kann ich auch eine Beziehung erzwingen, die vielleicht gar nicht so stark war.

Achten Sie also gut auf die Daten auf.

Der Bayes – Klassifikator (BBN)

- BBN = Bayesian Belief Network
- berechnet die Wahrscheinlichkeiten auf der Grundlage des gegebenen Netzwerkes
- dabei wird unterschieden zwischen:
  - chain $A \rightarrow T \rightarrow E$
  - fork $X \leftarrow E \rightarrow D$
  - collider $E \rightarrow D \leftarrow B$



- Statische Bayes Netze
  - betrachten Sachlagen ohne zeitliche Einflüsse
  - betrachten die Sachlage zu einem Zeitpunkt
- Dynamische Bayes Netze
  - erweitern die statischen Netze um den Faktor Zeit
  - jedes Attribut kann zu unterschiedlichen Zeitpunkten unterschiedliche Werte annehmen

Schauen wir uns das Netzwerk nochmal an. Das Netz wird auch das Bayesian Belief Network genannt. Es berechnet die Wahrscheinlichkeiten auf Grundlage des gegebenen Netzes, indem die Beziehungen aller Variablen zueinander betrachtet werden (Die Elternknoten spielen eine wichtige Rolle).

Frage: Welche 3 unterschiedliche Wege unterscheidet man?

Chain: Eine Chain, auch als "serielle Abhängigkeit" bezeichnet, tritt auf, wenn drei Variablen in einer Reihe miteinander verbunden sind.
In einer Chain gibt es eine Kausalkette von der ersten zur letzten Variable, wobei die mittlere Variable eine vermittelnde Rolle spielt.
Das bedeutet, dass die erste Variable die Ursache für die mittlere Variable ist, und die mittlere Variable ist wiederum die Ursache für die letzte Variable.
Mathematisch ausgedrückt: A --> B --> C, wobei A die Ursache von B und B die Ursache von C ist.

Fork: Eine Fork, auch als "Divergenz" bezeichnet, tritt auf, wenn zwei Variablen von einer dritten Variable beeinflusst werden.
In einer Fork gibt es eine gemeinsame Ursache, die sich auf beide der abzweigenden Variablen auswirkt.
Mathematisch ausgedrückt: A <-- B --> C, wobei B die gemeinsame Ursache für A und C ist.

Collider: Ein Collider, auch als "Konvergenz" bezeichnet, tritt auf, wenn zwei Variablen auf eine dritte Variable einwirken.
In einem Collider gibt es eine gemeinsame Wirkung, die von beiden der einmündenden Variablen beeinflusst wird.
Mathematisch ausgedrückt: A --> B <-- C, wobei B die gemeinsame Wirkung für A und C ist.

Siehe Abbildung, um es besser zu verstehen.

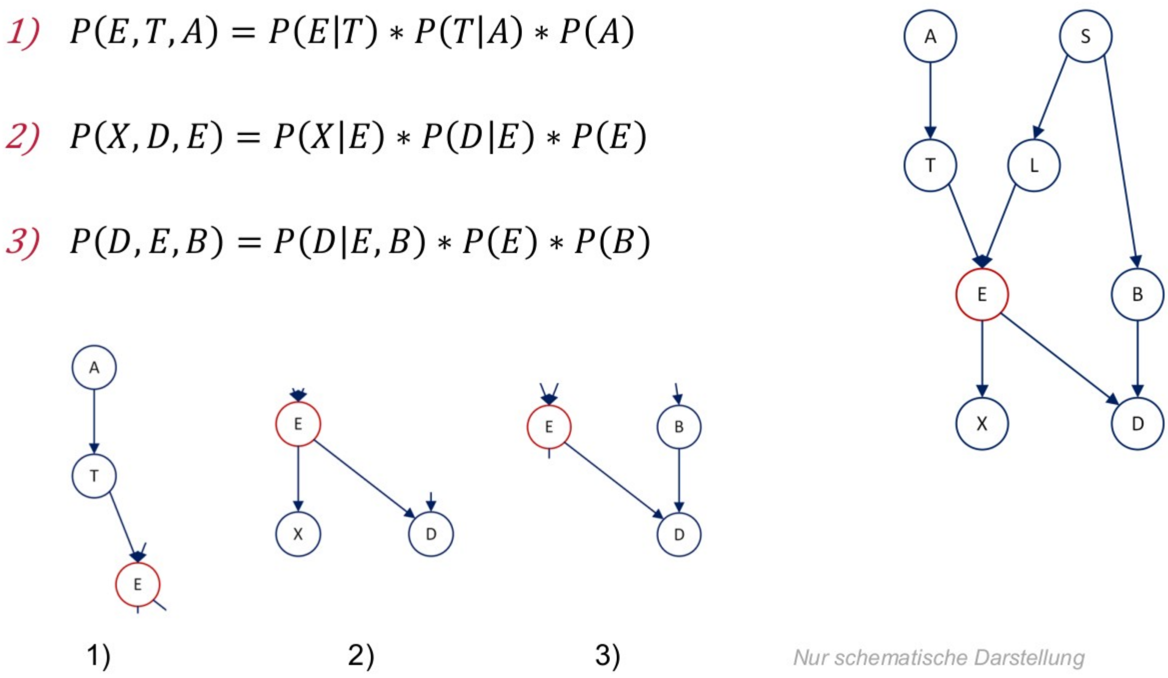
Frage: Was ist der Unterschied zwischen Statischen Bayes Netzwerk und dem dynamischen Bayes Netzwerk?

Statisches Bayes'sches Netzwerk:

Ein statisches Bayes'sches Netzwerk ist ein Netzwerk, das Ereignisse oder Variablen in einem bestimmten Zustand zu einem einzigen Zeitpunkt modelliert.
Es berücksichtigt nicht die Veränderungen oder Entwicklungen im Laufe der Zeit.
Die Wahrscheinlichkeiten und Abhängigkeiten zwischen den Variablen bleiben konstant und unverändert.
Statische Bayes'sche Netzwerke eignen sich gut zur Modellierung von Situationen, in denen die Ereignisse in einem Moment in der Zeit untersucht werden, ohne Berücksichtigung von zeitlicher Abfolge oder Veränderung.

Dynamisches Bayes'sches Netzwerk:

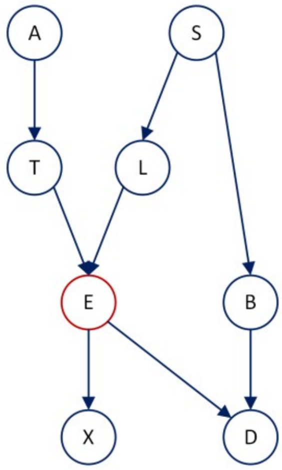
Ein dynamisches Bayes'sches Netzwerk ist ein Netzwerk, das die Veränderungen von Ereignissen oder Variablen über die Zeit modelliert.
Es berücksichtigt die zeitliche Abfolge und die Entwicklung von Zuständen oder Ereignissen.
Die Wahrscheinlichkeiten und Abhängigkeiten können sich im Laufe der Zeit ändern, und das Netzwerk kann die Entwicklung von Zuständen oder Variablen verfolgen.
Dynamische Bayes'sche Netzwerke eignen sich gut zur Modellierung von sich über die Zeit ändernden Situationen, wie z. B. in der Finanzmodellierung, der medizinischen Verlaufsanalyse, der Prognose oder der Robotik.



Hier sind drei Beispiele gegeben. Wenn ich zum Beispiel die Wahrscheinlichkeit für P(E,T,A) wissen will, dann muss ich die Wahrscheinlichkeit von E unter der Prämisse, dass T eingetreten ist, multiplizieren mit T, nachdem A eingetreten ist, und dann muss ich es nochmal mit der Wahrscheinlichkeit für A multiplizieren.

Allgemeines Vorgehen:

1. bei der Klassenvariablen beginnen
2. alle Pfade bis nach oben verfolgen
3. das gesamte Netzwerk verfolgen



Beispiel:

- Die Wahrscheinlichkeit D hängt von E und B ab
- E hängt bedingt von T und L ab
- T hängt bedingt von A ab
- L hängt bedingt von S ab
- B hängt bedingt von S ab
- (X hängt bedingt von E ab)

Was ist das allgemeine Vorgehen? Wir beginnen bei der Klassenvariable (hier D). Wir verfolgen alle Pfade bis ganz nach oben, bis wir das gesamte Netzwerk verfolgt haben.

Wenn ich ein Netzwerk mehrfach berechnen lasse, dann wird nicht jedes mal dasselbe rauskommen. Stattdessen werden die starken Beziehungen immer da sein, während die Variablen mit den schwachen Beziehungen variieren werden (einige werden sogar gar nicht auftauchen. X hat hier zum Beispiel keine Beziehung zu D. X könnte raus). ZUSATZ: **BAYES ARBEITET MIT HÄUFIGKEITSTABELLEN**.

- bnlearn: https://erdogant.github.io/bnlearn/pages/html/index.html
- → Use Case Medical domain
- Achtung: der mitgelieferte Datensatz weicht inhaltlich ab
- deshalb **asia.csv** aus dem ILIAS-Kurs verwenden

- pybbn: https://py-bbn.readthedocs.io/
- → API Documentation
- zum Vergleich hier ebenfalls **asia.csv** verwenden

- online sind weitere packages zu finden…

Bayes Netz ist nicht in Sklean vorhanden. Stattdessen müssen Sie die 2 Packages aus der Abbildung nutzen.