

BAYES und MARKOV Fragen zur VL

1. Wie wird das Bayes-Theorem definiert und wozu wird es benutzt?

- > Das Bayes-Theorem ist eine mathematische Formel, die in der Wahrscheinlichkeitstheorie verwendet wird, um die bedingte Wahrscheinlichkeit einer Hypothese in Anbetracht von vorhandenen Daten zu berechnen.
- > Ermittlung von gesuchten Wahrscheinlichkeiten aus gegebenen Wahrscheinlichkeiten.
- > Behandelt Zusammenhänge von Wahrscheinlichkeiten
- > Man beantwortet also die Frage: "was ist die Wahrscheinlichkeit von A, wenn B bereits eingetreten ist und einen Einfluss auf A hat".

2. 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 $P(X_i|Y_i)$

3. Worum geht es in dem Satz von Bayes? In eigenen Worten.

Es geht um die Ermittlung von Wahrscheinlichkeiten aus bereits vorhandenen Wahrscheinlichkeiten - das heißt, der Zusammenhang zwischen Wahrscheinlichkeiten soll ermittelt werden.

> 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.

Es ist besonders nützlich in Situationen, in denen Unsicherheit und Abhängigkeiten zwischen Ereignissen eine Rolle spielen.

4. In welche zwei Gruppen werden Wahrscheinlichkeiten unterteilt (bei Bayes)?

1. A-Priori-Wahrscheinlichkeiten: "A priori" bedeutet, dass die Wahrscheinlichkeit im Voraus berechnet wird, bevor neue Informationen oder Daten verfügbar sind. Dies sind anfängliche Wahrscheinlichkeiten, die Sie basierend auf Ihrem vorhandenen Wissen oder allgemeinem Verständnis einschätzen, bevor Sie spezifische Informationen oder Daten erhalten.

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

2. A Posteriori 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.

> A-Posteriori Wahrscheinlichkeit sind $p(A|B)$ und $p(B|A)$. Die bedingten Wahrscheinlichkeiten sind A posteriori-Wahrscheinlichkeiten.

5. Was ist das Ziel des Bayes Theorems? Was berechnet man mit der Formel?

Wir möchten also die bedingte Wahrscheinlichkeit $p(A|B)$ von zwei Ereignissen bestimmen, sofern einer der bedingten Wahrscheinlichkeiten $p(B|A)$ bekannt ist. Wir kennen $p(B|A)$ und wollen aber sehen, was die umgekehrte Wahrscheinlichkeit $p(A|B)$ ist.

6. Was ist das Komplement (zu A)? Wieso ist es wichtig?

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.

> Wichtig: Die Berechnung der bedingten Wahrscheinlichkeit 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.

> **Merke: Wenn ein Ereignis (B) von einem Ereignis (A) bedingt wird, dann ist es auch abhängig von seinem Komplement (Nicht A).**

7. Wie verhalten sich A und B z ihrem Komplement?

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.

8. $P(B)$ ist nicht gegeben, aber das Komplement von A ist gegeben? Wie berechne ich $P(B)$?

> Siehe Formel. Hier nur Erklärung.

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.

> **Berechnung über das Komplement.**

9. Was wird mit der Verbundregel berechnet und wie lautet die Formel?

> Formel in Formelsammlung

Über die Verbundregel kann man die Wahrscheinlichkeiten für alle A priori Wahrscheinlichkeiten berechnen, die von einer oder mehrere Wahrscheinlichkeiten abhängig sind.

10. Ist $p(A|B)$ das Komplement 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 einen nicht das andere leicht herleiten. $1 - p(B|A) \neq p(A|B)$! Aber $1 - p(B|A) = p(B| -A)$ für 2 Werte → Die jeweilige Verzweigung muss 100 % ergeben.

11. Was ist das Bayes'sche Netzwerk?

> ein graphisches Modell, das dazu dient, Beziehungen und kausale Zusammenhänge zu visualisieren

12. 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**.

13. Was sind die 2 Vorteile von Bayes Netzwerken?

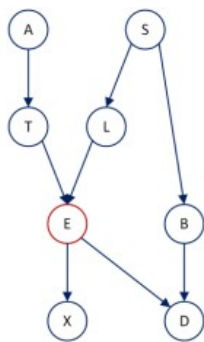
Die Vorteile von Bayes Netzwerken sind, dass Sie

1. Sehr robust sind
2. Sie bieten einen effizienten Ansatz gegen Overfitting

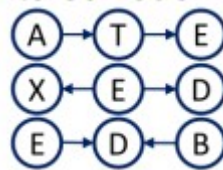
14. Wie wird Bayes Netzwerk noch genannt? Welche Art von Algorithmus ist es?

- > Bayes Belief Network.
- > Klassifikator

15. Welche 3 unterschiedliche Wege unterscheidet man beim BBN?



- chain
- fork
- collider



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: $A \rightarrow B \rightarrow C$, 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: $A \leftarrow B \rightarrow C$, 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: $A \rightarrow B \leftarrow C$, wobei B die gemeinsame Wirkung für A und C ist.

16. 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.

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

17. Wenn ich ein Netzwerk mehrmals berechne, kommt jedesmal dasselbe Ergebnis raus?

Nein. 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).

18. Was ist der Naive Bayes?

- > Ein Klassifikator.
- > Supervised

19. Wodurch unterscheidet sich Naive Bayes vom Netzwerk? Welche 2 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.

> **Attribute voneinander unabhängigen**

Die zweite Annahme ist, dass **alle Variablen gleich wichtig sind**. Das heißt, jede Variable hat denselben Einfluss auf die Zielvariable.

20. 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).

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.

21. 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.

Multinomial: Die Daten sind Multinomial 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.

Markov

22. Wofür wird das Markov Modell benutzt – Was wird damit vorhergesagt?

Das Markov-Modell, benannt nach dem russischen Mathematiker Andrey Markov, ist eine mathematische Methode zur **Vorhersage zukünftiger Zustände basierend auf gegenwärtigen Zuständen**.

> Stochastisches Modell für sich zufällig ändernde Modelle

23. Womit werden Zustandsübergänge berechnet?

> Mit Übergangswahrscheinlichkeiten und Zustandswahrscheinlichkeiten werden Zustandsübergänge berechnet.

24. Welche Annahme macht das Markov-Modell? Wovon hängen zukünftige Ereignisse ab und wovon eben nicht?

Es basiert auf der Idee, dass die zukünftige Entwicklung eines Systems oder Prozesses nur von seinem aktuellen Zustand abhängt und nicht von den vergangenen Ereignissen. Wir schauen uns also nicht die Vergangenheit an, sondern nur die Wahrscheinlichkeit, mit der aktuelle Zustände auftreten, und die Wahrscheinlichkeit, mit der einzelne Zustände in andere Zustände überführt werden.

25. Welche Merkmale und Voraussetzungen charakterisieren den Markov-Entscheidungsprozess (MDP) in Bezug auf die beobachteten Zustände, und wie unterscheidet sich ein MDP in Bezug auf sein Gedächtnis von anderen Modellierungsmethoden in der Entscheidungstheorie?

Der Markov-Entscheidungsprozess wird bei **kontrollierten Systemen mit vollständig beobachteten Zuständen** eingesetzt. Das heißt, wir kennen alle Zustände. Es ist weiterhin **Gedächtnislos**. Das heißt – wie bereits erwähnt, dass wir uns nicht die Vergangenheit anschauen, sondern nur die aktuellen Zustände und ihre Übergangswahrscheinlichkeiten.

26. Was ist die Markov-Kette und wofür wird Sie benutzt?

Eine Markov-Kette ist eine **mathematische Darstellung eines Markov-Prozesses**, bei dem die Zustände in einem **diskreten Zustandsraum modelliert werden**. Dieser Zustandsraum kann endlich sein, was bedeutet, dass es eine begrenzte Anzahl von Zuständen gibt, oder abzählbar, was bedeutet, dass es eine unendliche, aber zählbare Anzahl von Zuständen gibt.

Bsp.

Um das zu erklären, hier ist ein Beispiel: Angenommen, Sie modellieren das Wetter mit einer Markov-Kette. In diesem Fall könnte der Zustandsraum endlich sein, da Sie das Wetter in "sonnig", "bewölkt" und "regnerisch" einteilen könnten. Oder Sie könnten den Zustandsraum als abzählbar betrachten, indem Sie das Wetter in "sonnig", "teilweise bewölkt", "stark bewölkt", "leichter Regen", "starker Regen" und so weiter unterteilen. Jeder Zustand repräsentiert eine bestimmte Art von Wetter. In einer Markov-Kette basieren die Übergänge zwischen den Zuständen auf Wahrscheinlichkeiten, die angeben, mit welcher Wahrscheinlichkeit das Wetter von einem Zustand in einen anderen wechselt. Zum Beispiel könnte es eine hohe Wahrscheinlichkeit geben, dass ein sonniger Tag zu einem bewölkten Tag wird, während die Wahrscheinlichkeit eines plötzlichen Übergangs zu einem Regentag geringer ist. Markov-Ketten werden in verschiedenen Anwendungen eingesetzt, um stochastische Prozesse zu modellieren, bei denen die zukünftige Entwicklung allein vom aktuellen Zustand abhängt.

27. Was ist das Ziel einer Markov Kette?

Das Ziel ist die Vorhersage des Eintretens eines zukünftigen Ereignisses, anhand des aktuellen Zustandes und der Übergangswahrscheinlichkeiten für den jeweiligen Zustand.

28. Was sind Markov-Ketten erster Ordnung?

Markov-Ketten erster Ordnung sind Markov-Ketten, bei denen die Übergangswahrscheinlichkeiten zwischen den Zuständen nur von ihrem direkten vorherigen Zustand abhängen. Diese Beziehung ist in der Praxis auch am stärksten.

29. Wann ist eine Markov-Kette homogen?

Eine Markov-Kette ist homogen, **wenn die Übergangswahrscheinlichkeiten zwischen den Zuständen im Laufe der Zeit konstant bleiben**. Das bedeutet, dass die Wahrscheinlichkeit, von einem Zustand in einen anderen zu wechseln, unabhängig von der Zeit ist.

Beispiel für eine nicht homogene Markov-Kette: Nehmen wir an, wir modellieren das Wachstum einer Pflanze über die Zeit. Die Pflanze kann in verschiedenen Entwicklungsstadien sein, und wir betrachten, wie sie von einem Stadium zum nächsten übergeht. Die Übergangswahrscheinlichkeiten zwischen den Stadien hängen von der Jahreszeit ab. Im Frühling hat die Pflanze eine höhere Wahrscheinlichkeit, in das nächste Stadium überzugehen, während die Übergangswahrscheinlichkeiten im Sommer, Herbst und Winter geringer sind. In diesem Fall ist die Markov-Kette nicht homogen, da die Übergangswahrscheinlichkeiten von der Zeit, d.h., der Jahreszeit, abhängen. Beispiel für eine homogene Markov-Kette: Betrachten wir eine einfache Münzwurf-Markov-Kette. Sie befinden sich in einem Zustand, der entweder "Kopf" oder "Zahl" ist, abhängig von einer Münze, die Sie werfen. Wenn Sie in "Kopf" sind, werfen Sie die Münze, und abhängig von ihrem Ergebnis wechseln Sie entweder zu "Kopf" oder "Zahl". Die Übergangswahrscheinlichkeiten sind für beide Zustände gleich: Es gibt eine 50%ige Wahrscheinlichkeit, bei "Kopf" zu bleiben, und eine 50%ige Wahrscheinlichkeit, zu "Zahl" zu wechseln. Dies ist ein Beispiel für eine homogene Markov-Kette, da die Übergangswahrscheinlichkeiten konstant und unabhängig von der Zeit oder dem Zustand sind.

30. Was sind Markov-Ketten n-ter Ordnung?

Eine Markov-Kette höherer Ordnung (n-te Ordnung) bezieht sich auf eine Markov-Kette, bei der die Übergangswahrscheinlichkeiten nicht nur von einem vorherigen Zustand abhängen, sondern von den letzten n Zuständen.

31. Was sind die bedingten Wahrscheinlichkeiten bei Markov? Was zeigen Sie?

Wir haben hier also auch bedingte Wahrscheinlichkeiten und zwar die bedingten Wahrscheinlichkeiten, dass aus einem Zustand in der Gegenwart derselbe oder ein anderer Zustand in der Zukunft eintritt.

$P(A|A) = P(A_v|A_{v-1})$: Das bedeutet, dass dies die Wahrscheinlichkeit ist, dass aus dem gegenwärtigen Zustand A auch in der Zukunft der Zustand A wird. Dies impliziert, dass der Zustand A im Zeitpunkt v von seinem vorherigen Zustand A im Zeitpunkt v-1 abhängt.

$P(A|B) = P(A_v|B_{v-1})$: Das bedeutet, dass dies die Wahrscheinlichkeit ist, dass aus dem gegenwärtigen Zustand B der Zustand A wird. Dies impliziert, dass der Übergang von B zu A im Zeitpunkt v stattfindet, wobei der vorherige Zustand B im Zeitpunkt v-1 berücksichtigt wird.

32. Warum beträgt die Gesamtwahrscheinlichkeit aller abgehenden Ereignisse eines Zustands in einem Markov-Prozess immer 1 und wie hängt dies mit den möglichen zukünftigen Zustandsübergängen zusammen?

Die Gesamtwahrscheinlichkeit aller möglichen Ereignisse, die von einem Zustand ausgehen, beträgt immer 1. Dies ergibt sich aus der Logik. Wenn ein Zustand vorliegt und es zwei mögliche Wege gibt, wie sich dieser Zustand in der Zukunft entwickeln kann, dann ergibt die Summe der Wahrscheinlichkeiten beider Wege natürlich 100 %. In einem Markov-Modell repräsentieren die "abgehenden Pfeile" die Wahrscheinlichkeiten für Zustandsübergänge von einem gegebenen Zustand zu anderen möglichen Zuständen. Diese Pfeile zeigen an, wohin der Prozess von einem bestimmten Zustand aus gehen kann. Die "ankommenden Pfeile" oder "einkommenden Pfeile" hingegen sind die Wahrscheinlichkeiten, wie ein bestimmter Zustand erreicht wurde. Diese Pfeile zeigen an, von welchen vorherigen Zuständen der aktuelle Zustand erreicht werden kann.

33. Sind die ankommenden Pfeile in der Summe auch 1? Begründe.

Die Summe der ankommenden Pfeile zu einem bestimmten Zustand ist normalerweise nicht auf 1 festgelegt, weil ein Zustand von verschiedenen vorherigen Zuständen erreicht werden kann, und die Wahrscheinlichkeiten für diese Übergänge können unterschiedlich sein.

34. Wann nutzt man Markov-Ketten und wann Hidden Markov?

Antwort: Markov-Ketten nutzt man, wenn alle Zustände bekannt sind. Hidden Markov nutzt man, wenn nicht alle Zustände bekannt sind.

Verborgene Zustände: Im Gegensatz dazu sind bei HMMs die Zustände nicht direkt beobachtbar. Stattdessen werden Beobachtungen generiert, die von den versteckten Zuständen abhängen, und Sie müssen die versteckten Zustände aufgrund der Beobachtungen schätzen oder vorhersagen. Anwendungsbereiche: HMMs werden in Situationen eingesetzt, in denen Sie den wahren Zustand des Systems nicht direkt beobachten können, sondern nur Beobachtungen erhalten. Typische Anwendungsfälle sind Spracherkennung

35. Was muss ich für HMM wissen?

Voraussetzung: Muss die Übergangswahrscheinlichkeiten kennen.

> Transition Matrix