Wzór Bayesa

Źródło: Wilhelm Gunkel, dostępny w internecie: unsplash.com, domena publiczna.

W tym materiale poznamy twierdzenie Bayesa, będące jednym z najważniejszych twierdzeń teorii prawdopodobieństwa wykorzystywanych w zastosowaniach praktycznych.

Twierdzenie to opisuje prawdopodobieństwo zajścia danego zdarzenia na podstawie wcześniejszej wiedzy o warunkach, które mogą być związane ze zdarzeniem.

Thomas Bayes
Źródło: dostępny w internecie: commons.wikimedia.org, domena publiczna.

Na przykład wiadomo, że ryzyko nie zdania egzaminu poprawkowego wzrasta wraz z upływem czasu od pierwszego terminu egzaminu. Twierdzenie Bayesa pozwala oszacować to ryzyko bardziej precyzyjnie.

Ważnym zastosowaniem tego twierdzenia jest tzw. wnioskowanie bayesowskie, będące jedną z metod wnioskowania statystycznego. Za jego pomocą można aktualizować prawdopodobieństwo zajścia danego zdarzenia, w miarę pojawiania się nowych informacji.

Wnioskowanie bayesowskie ma zastosowanie w medycynie, filozofii, inżynierii, sporcie – w przypadku dynamicznie zmieniających się danych.

Twierdzenie Bayesa zostało tak nazwane na cześć wielebnego Thomasa Bayesa, osiemnastowiecznego angielskiego matematyka i duchownego, którego uważa się też za jednego z twórców wzoru na prawdopodobieństwo warunkowe.

Twoje cele

- Obliczysz prawdopodobieństwo zajścia danego zdarzenia, korzystając ze wzoru Bayesa.
- Opiszesz sytuację probabilistyczną, korzystając z prawdopodobieństwa warunkowego.
- Dobierzesz odpowiedni model matematyczny do rozwiązania problemu probabilistycznego z kontekstem realistycznym.

Przeczytaj

Niech $B_1, B_2, \ldots, B_n$ będzie układem wzajemnie wykluczających się zdarzeń należących do tej samej przestrzeni zdarzeń elementarnych, z których przynajmniej jedno musi zajść. Prawdopodobieństwo zajścia każdego z tych zdarzeń niech będzie dodatnie.

Niech A oznacza zdarzenie o dodatnim prawdopodobieństwie, które może zajść wyłącznie z jednym ze zdarzeń $B_1, B_2, \ldots, B_n$.

Ze wzoru na prawdopodobieństwo warunkowe wynika, że jeśli $A \subset \Omega$ i $B \subset \Omega$ oraz:

- $P(B) > 0$ to $P(A \cap B) = P(A \setminus B) \cdot P(B)$,
- $P(A) > 0$ to $P(A \cap B) = P(B \setminus A) \cdot P(A)$.

Korzystając z powyższego dla zdarzeń A i B_i, gdzie $i = 1, 2, \ldots, n$, możemy zapisać równość:

$$P(B_i \setminus A) \cdot P(A) = P(A \setminus B_i) \cdot P(B_i)$$

I przekształcić tę równość równoważnie.

$$P(B_i \setminus A) = \frac{P(A \setminus B_i) \cdot P(B_i)}{P(A)}$$

Do prawej strony równości stosujemy wzór na prawdopodobieństwo całkowite.

$$P(B_i \setminus A) = \frac{P(A \setminus B_i) \cdot P(B_i)}{P(A \setminus B_1) \cdot P(B_1) + \ldots + P(A \setminus B_n) \cdot P(B_n)}$$

Otrzymaną zależność nazywamy **wzorem Bayesa**.

Prawdopodobieństwo $P(B_i)$ nazywane jest czasem prawdopodobieństwem *a priori*, a prawdopodobieństwo $P(B_i \setminus A)$ nazywamy prawdopodobieństwem *a posteriori*, gdyż określa ono szansę zajścia zdarzenia B_i po zaobserwowaniu zajścia zdarzenia A. Wzór Bayesa stosujemy więc głównie wtedy, gdy znamy wynik doświadczenia i pytamy o jego przebieg.

Sformułujemy wzór Bayesa w postaci najczęściej używanej – ograniczymy się tylko do trzech zdarzeń A, B_1, B_2.

> **Twierdzenie: Wzór Bayesa, reguła Bayesa**
>
> Niech Ω będzie zbiorem wszystkich wyników pewnego doświadczenia, a $B_1 \subset \Omega$, $B_2 \subset \Omega$ zdarzeniami o dodatnich prawdopodobieństwach, takimi że $B_1 \cup B_2 = \Omega$ oraz

$B_1 \cap B_2 = \emptyset$. Wówczas dla dowolnego zdarzenia $A \subset \Omega$ o dodatnim prawdopodobieństwie, prawdziwy jest wzór:

$$P(B_1 \setminus A) = \frac{P(B_1) \cdot P(A \setminus B_1)}{P(A \setminus B_1) \cdot P(B_1) + P(A \setminus B_2) \cdot P(B_2)}$$

Przykład 1

Niech B będzie zdarzeniem – dana osoba nosi okulary, A niech będzie zdarzeniem – dana osoba ma zielone oczy. Wtedy zdarzenie $B \setminus A$ – osoba nosząca okulary wśród osób mających zielone oczy, zdarzenie $A \setminus B$ – osoba mająca zielone oczy wśród osób noszących okulary.

Po zbadaniu pewnej grupy osób stwierdzono, że dla tej grupy osób $P(A) = 0,1$, $P(B) = 0,2$ i $P(B \setminus A) = 0,6$. Obliczymy $P(A \setminus B)$.

Korzystamy ze wzoru Bayesa.

$P(A \setminus B) = \frac{P(B \setminus A) \cdot P(A)}{P(B)}$

$P(A \setminus B) = \frac{0,6 \cdot 0,1}{0,2} = 0,3.$

Przykład 2

Matematykę lubi co piąty uczeń klasy pierwszej, co czwarty uczeń klasy drugiej i co drugi uczeń klasy trzeciej. Z grupy uczniów składającej się z 10 uczniów klasy pierwszej, 10 uczniów klasy drugiej i 10 uczniów klasy trzeciej wybrano jedną osobę.

Obliczymy prawdopodobieństwo, że wybrana osoba lubi matematykę.

Jakie jest prawdopodobieństwo, że był to uczeń klasy drugiej, jeżeli wiadomo, że wybrana osoba lubi matematykę?

Oznaczmy:
A – zdarzenie polegające na tym, że wybrany uczeń jest z klasy pierwszej,
B – zdarzenia polegające na tym, że wybrany uczeń jest z klasy drugiej,
C – zdarzenie polegające na tym, że wybrany uczeń jest klasy trzeciej,
M – zdarzenie polegające na tym, że wybrany uczeń lubi matematykę.

Zauważmy, że zdarzenia A, B, C tworzą zupełny układ zdarzeń oraz $A \cup B \cup C = \Omega$. Możemy więc zastosować twierdzenie o prawdopodobieństwie całkowitym.

$P(M) = P(A) \cdot P(M \setminus A) + P(B) \cdot P(M \setminus B) + P(C) \cdot P(M \setminus C)$

$P(M) = \frac{1}{3} \cdot \frac{1}{5} + \frac{1}{3} \cdot \frac{1}{4} + \frac{1}{3} \cdot \frac{1}{2} = \frac{19}{60}$

Ponieważ $P(M) > 0$, to możemy zastosować twierdzenie Bayesa.

$P(B \setminus M) = \frac{P(B)\cdot P(M\setminus B)}{P(A)\cdot P(M\setminus A)+P(B)\cdot P(M\setminus B)+P(C)\cdot P(M\setminus C)}$

$P(B \setminus M) = \frac{\frac{1}{3}\cdot\frac{1}{4}}{\frac{1}{3}\cdot\frac{1}{5}+\frac{1}{3}\cdot\frac{1}{4}+\frac{1}{3}\cdot\frac{1}{2}} = \frac{5}{19}$

Odpowiedź:

Prawdopodobieństwo, że wybrana osoba lubi matematykę jest równe $\frac{19}{60}$.
Prawdopodobieństwo, że wybrana osoba była uczniem klasy drugiej, jeżeli wiadomo, że wybrana osoba lubi matematykę jest równe $\frac{5}{19}$.

Przykład 3

W biegach przełajowych startują zawodnicy tylko z dwóch klubów. Zawodnicy z klubu *Niezwyciężeni* stanowią 30% wszystkich zawodników, a pozostali zawodnicy są z klubu *Niepokonani*. Wśród *Niezwyciężonych* jest połowa kobiet, a wśród *Niepokonanych* tylko 10% to kobiety. W sposób losowy ze wszystkich zawodników wybrano jedną osobę, okazało się, że jest to kobieta. Obliczymy prawdopodobieństwo, że jest ona zawodniczką z klubu *Niezwyciężeni*.

Oznaczmy:
A – wybrana osoba to kobieta,
B_1 – wybrana osoba jest z klubu *Niezwyciężeni*,
B_2 – wybrana osoba jest z klubu *Niepokonani*.

Na podstawie treści zadania możemy zapisać:

$P(B_1) = 0,3$

$P(B_2) = 1 - 0,3 = 0,7$

$P(A \setminus B_1) = 0,5$

$P(A \setminus B_2) = 0,1$

Korzystamy ze wzoru Bayesa.

$P(B_1 \setminus A) = \frac{P(B_1)\cdot P(A\setminus B_1)}{P(A\setminus B_1)\cdot P(B_1)+P(A\setminus B_2)\cdot P(B_2)}$

$P(B_1 \setminus A) = \frac{0{,}3\cdot 0{,}5}{0{,}5\cdot 0{,}3+0{,}1\cdot 0{,}7} = \frac{0{,}15}{0{,}22} \approx 0,68$

Odpowiedź:

Prawdopodobieństwo tego, że wylosowana kobieta jest zawodniczką z klubu *Niezwyciężeni* jest równe około $0,68$.

Przykład 4

Na pierwszym klombie rośnie 99 bratków i jedna stokrotka. Na drugim klombie rośnie 99 stokrotek i jeden bratek. Rzucamy kostką do gry. Jeśli wypadnie trójka, Anka zrywa kwiat z pierwszego klombu. W pozostałych przypadkach zrywa kwiat z drugiego klombu. Nie znamy wyniku rzutu kostką, ale wiemy, że Anka zerwała bratek. Wyznaczymy prawdopodobieństwo tego, że Anka zerwała kwiat z pierwszego klombu.

Oznaczmy:
B – zdarzenie polegające na tym, że Anka zerwała bratek,
A – zdarzenie polegające na zerwaniu kwiatu z pierwszego klombu,
C – zdarzenie polegające na zerwaniu kwiatu z drugiego klombu.

Wtedy:

$P(A) = \frac{1}{6}, P(C) = \frac{5}{6}$ – wybór klombu

$P(B \setminus A) = \frac{99}{100}, P(B \setminus C) = \frac{1}{100}$ – wybór bratka

Chcemy obliczyć $P(A \setminus B)$.

Korzystamy ze wzoru Bayesa.

$$P(A \setminus B) = \frac{P(B \setminus A) \cdot P(A)}{P(B \setminus A) \cdot P(A) + P(B \setminus C) \cdot P(C)}$$

$$P(A \setminus B) = \frac{\frac{99}{100} \cdot \frac{1}{6}}{\frac{99}{100} \cdot \frac{1}{6} + \frac{1}{100} \cdot \frac{5}{6}} = \frac{\frac{33}{200}}{\frac{104}{600}} = \frac{99}{104}$$

Odpowiedź:

Prawdopodobieństwo tego, że Anka zerwała kwiat z pierwszego klombu jest równe $\frac{99}{104}$.

Przykład 5

Test na obecność pewnego wirusa daje wynik pozytywny z prawdopodobieństwem $0,84$, a negatywny z prawdopodobieństwem $0,16$, jeśli wirus jest w organizmie.

Jeśli wirusa w organizmie nie ma, prawdopodobieństwo wyniku pozytywnego jest równe $0,04$. Zakłada się, że 1% populacji zarażonych jest tym wirusem. Obliczymy prawdopodobieństwo, że losowo wybrana osoba jest istotnie zarażona wirusem, jeżeli wiadomo, że test dał wynik pozytywny.

Oznaczmy:
A – zdarzenie polegające na tym, że test dał wynik pozytywny,
B_1 – zdarzenie polegające na tym, że wirus jest w organizmie,
B_2 – zdarzenie polegające na tym, że wirusa nie ma w organizmie,

Korzystając z treści zadania, zapisujemy odpowiednie prawdopodobieństwa.

$P(B_1) = 0,01$

$P(B_2) = 0,99$

$P(A \setminus B_1) = 0,84$

$P(A \setminus B_2) = 0,04$

Wyznaczone liczby podstawiamy do wzoru Bayesa.

$P(B_1 \setminus A) = \frac{P(A \setminus B_1) \cdot P(B_1)}{P(A \setminus B_1) \cdot P(B_1) + P(A \setminus B_2) \cdot P(B_2)}$

$P(B_1 \setminus A) = \frac{0,84 \cdot 0,01}{0,84 \cdot 0,01 + 0,99 \cdot 0,04} = 0,175$

Odpowiedź:

Prawdopodobieństwo tego, że losowo wybrana osoba jest istotnie zarażona wirusem, jeżeli test dał wynik pozytywny, jest równe $0,175$.

Słownik

wzór Bayesa

niech Ω będzie zbiorem wszystkich wyników pewnego doświadczenia, a $B_1 \subset \Omega$, $B_2 \subset \Omega$ zdarzeniami o dodatnich prawdopodobieństwach, takimi że $B_1 \cup B_2 = \Omega$ oraz $B_1 \cap B_2 = \emptyset$; wówczas dla dowolnego zdarzenia $A \subset \Omega$ o dodatnim prawdopodobieństwie, prawdziwy jest wzór:

$$P(B_1 \setminus A) = \frac{P(B_1) \cdot P(A \setminus B_1)}{P(A \setminus B_1) \cdot P(B_1) + P(A \setminus B_2) \cdot P(B_2)}$$

Animacja

Polecenie 1

Zapoznaj się z animacją, która przybliży Ci zagadnienia związane ze wzorem Bayesa. Postaraj się najpierw samodzielnie rozwiązać prezentowane tam problemy, a następnie porównaj z zamieszczonymi w animacji.

Wystąpił błąd

Film nawiązujący do treści materiału dotyczącej wzoru Bayesa.

Polecenie 2

Wśród bliźniąt prawdopodobieństwa urodzenia się dwóch chłopców i dwóch dziewczynek jest odpowiednio równe a i b. Dla bliźniąt różnopłciowych prawdopodobieństwo urodzenia się jako pierwsze dziecko jest dla obu płci jednakowe. Pani Marta będzie miała bliźnięta. Jako pierwsze dziecko urodził się chłopiec. Oblicz prawdopodobieństwo, że drugie dziecko też będzie chłopcem.

Sprawdź się

Pokaż ćwiczenia:

Ćwiczenie 1

Ćwiczenie 2

Ćwiczenie 3

Ćwiczenie 4

Ćwiczenie 5

Ćwiczenie 6

Do sklepu przywożone są pewne detale od trzech producentów. Detale mogą być w pierwszym lub drugim gatunku. W tabelce zapisano ile procent detali pochodzi od którego producenta oraz ile procent detali w pierwszym gatunku pochodzi od danego producenta.

Producent	$\%$ detali przywożonych do sklepu	$\%$ detali w pierwszym gatunku
I	52	40
II	13	65
III	35	70

Ćwiczenie 7

Dwóch łuczników strzela do nadmuchanego balonu. Balon zostaje zniszczony, jeżeli trafi go co najmniej jedna strzała. Pierwszy łucznik oddał dziewięć strzałów, a drugi dziesięć strzałów. Pierwszy łucznik trafia średnio osiem na dziesięć strzałów, a drugi siedem na dziesięć strzałów. Strzała trafiła w cel. Oblicz prawdopodobieństwo, że celny strzał oddał pierwszy strzelec.

Ćwiczenie 8

Mamy 15 monet, z których $\frac{1}{5}$ jest fałszywa. Monety fałszywe zawsze upadają reszką do góry. Monetę wybrano losowo i rzucono dziesięć razy. Za każdym razem wypadła reszka. Oblicz prawdopodobieństwo tego, że wybrano fałszywą monetę.

Dla nauczyciela

Autor: Justyna Cybulska

Przedmiot: Matematyka

Temat: Wzór Bayesa

Grupa docelowa:

III etap edukacyjny, liceum, technikum, zakres rozszerzony

Podstawa programowa:

XII. Rachunek prawdopodobieństwa i statystyka.

Zakres podstawowy. Uczeń:

1) oblicza prawdopodobieństwo w modelu klasycznym.

Zakres rozszerzony. Uczeń spełnia wymagania określone dla zakresu podstawowego, a ponadto:

1) oblicza prawdopodobieństwo warunkowe i stosuje wzór Bayesa, stosuje twierdzenie o prawdopodobieństwie całkowitym.

Kształtowane kompetencje kluczowe:

- kompetencje w zakresie rozumienia i tworzenia informacji
- kompetencje matematyczne oraz kompetencje w zakresie nauk przyrodniczych, technologii i inżynierii
- kompetencje cyfrowe
- kompetencje osobiste, społeczne i w zakresie umiejętności uczenia się

Cele operacyjne:

Uczeń:

- oblicza prawdopodobieństwo zajścia danego zdarzenia, korzystając ze wzoru Bayesa
- opisuje sytuację probabilistyczną, korzystając z prawdopodobieństwa warunkowego
- dobiera odpowiedni model matematyczny do rozwiązania problemu probabilistycznego z kontekstem realistycznym

Strategie nauczania:

- konstruktywizm

Metody i techniki nauczania:

- ocena punktowa ważona
- obieg kart

Formy pracy:

- praca w grupach
- praca w parach
- praca całego zespołu klasowego

Środki dydaktyczne:

- komputery z dostępem do Internetu w takiej liczbie, żeby każdy uczeń miał do dyspozycji komputer

Przebieg lekcji

Faza wstępna:

1. Uczniowie wspólnie metodą oceny punktowej ważonej przypominają wszystkie wiadomości i umiejętności dotyczące prawdopodobieństwa, jakie do tej pory uzyskali. Jeden z uczniów, na podstawie wypowiedzi pozostałych, tworzy graficzny model zależności między pozyskanymi informacjami. Wynikiem pracy może być konkluzja, które definicje i twierdzenia dotyczące prawdopodobieństwa trzeba znać koniecznie (najlepiej na pamięć), a które są ich pochodnymi.
2. Nauczyciel podaje temat i cele zajęć, uczniowie ustalają kryteria sukcesu.

Faza realizacyjna:

1. Uczniowie pracują w parach. Zapoznają się z animacją. Najpierw próbują samodzielnie rozwiązać podane tam przykłady, a następnie porównują z przedstawionym rozwiązaniami. W podobny sposób pracują, analizując przykłady przedstawione w sekcji „Przeczytaj”.
2. Teraz pary uczniów łączą się w grupy 4 osobowe i pracują metodą obiegu kart, rozwiązując zadania z sekcji „Sprawdź się”. Grupa rozpoczynająca rozwiązywanie danego zadania zapisuje początek rozwiązania i podaje następnej grupie kartkę z zapisem. Ta z kolei dopisuje następną część i podaje kartkę dalej, itp. Ostatnia grupa, która kończy rozwiązanie zadania sprawdza interaktywnie poprawność uzyskanego wyniku. Jeśli wynik jest błędny, kartka wędruje do początkowej grupy, która weryfikuje swoje rozwiązanie. Grupa podaje kartkę następnej grupie, itd. Jeśli nadal odpowiedź nie jest poprawna, grupa może poprosić o pomoc nauczyciela.
3. Końcowym elementem tej części zajęć może być dyskusja – czy łatwo jest rozwiązywać zadania, śledząc tok rozumowania innej grupy osób i czy wyodrębnienie na początku

lekcji najważniejszych wzorów i umiejętności pomogło w doborze odpowiednich strategii rozwiązywania zadań.

Faza podsumowująca:

1. Wskazany przez nauczyciela uczeń przedstawia krótko najważniejsze elementy zajęć, poznane wiadomości, ukształtowane umiejętności.
2. Nauczyciel omawia przebieg zajęć, wskazuje mocne i słabe strony pracy uczniów, ocenia pracę grup i par.

Praca domowa:

Zadaniem uczniów jest poszukanie w dostępnych źródłach, informacji na temat zastosowania wzoru Bayesa w innych dziedzinach wiedzy i przygotowania krótkiej prezentacji na ten temat. Od tych prezentacji rozpocznie się następna lekcja.

Materiały pomocnicze:

Klasyczna definicja prawdopodobieństwa (treść rozszerzona)

Wskazówki metodyczne:

Z animacją każdy uczeń może zapoznać się w domu i na jej podstawie przygotować jedno zadanie, które na lekcji da do rozwiązania koleżance lub koledze z ławki. Animację można wykorzystać na zajęciach z prawdopodobieństwa całkowitego.