# Elementy Modelowania Matematycznego
# Wykład 2

## Wnioskowanie

Romuald Kotowski

Katedra Informatyki Stosowanej

PJWSTK 2009

# Spis treści

# Spis treści

# Spis treści

# Twierdzenie Bayesa

## Prawdopodobieństwo warunkowe

Jeśli $P(B) > 0$ to **prawdopodobieństwo warunkowe** zdarzenia $A$, pod warunkiem, że zaszło zdarzenie $B$ definiujemy jako:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{1}$$

Jeśli zdarzenia $A_1, A_2, \ldots$ tworzą podział przestrzeni $\Omega$ przez $P(A_i) > 0$ dla dowolnego $i = 1, 2, \ldots$, to dla dowolnego zdarzenia $B$ zachodzi

$$P(B) = \sum_{i=1}^{\infty} P(A_i)\, P(B|A_i) \tag{2}$$

Jest to **prawdopodobieństwo całkowite** zajścia zdarzenia $B$.

# Twierdzenie Bayesa

## Prawdopodobieństwo warunkowe - przykład

Mamy trzy urny. W pierwszej znajdują się 2 białe i 1 czarna kula, w drugiej 3 białe i 1 czarna, w trzeciej 2 białe i 2 czarne. Z losowo wybranej urny losowo wybieramy kulę. Jakie jest prawdopodobieństwo, że będzie to kula biała jeśli prawdopodobieństwo wyboru każdej z urn wynosi $1/2$?

# Twierdzenie Bayesa

## Prawdopodobieństwo warunkowe - przykład

**Rozwiązanie**

Oznaczmy przez:

- $A_i$ - wybrana kula pochodzi z $i$-tej urny,
- $B$ - wybrana kula jest biała.

Ponieważ są spełnione założenia o prawdopodobieństwie całkowitym ($P(A_i) > 0$), więc możemy je zastosować do obliczenia $P(B)$.

$$P(A_1) = P(A_2) = P(A_3) = \frac{1}{3}$$

$$P(B|A_1) = \frac{2}{3}, \quad P(B|A_2) = \frac{3}{4}, \quad P(B|A_3) = \frac{1}{2}, \tag{3}$$

$$P(B) = \frac{23}{36}$$

# Twierdzenie Bayesa

## Twierdzenie Bayesa

Jeśli zdarzenia $A_1, A_2, \ldots$ tworzą podział przestrzeni $\Omega$ i $P(A_i) > 0$, $i = 1, 2, \ldots$, dla dowolnego to dla zdarzenia $B$ takiego, że P(B)>0, to dla każdego $i$

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum\limits_{j=1}^{\infty} P(A_j)\,P(B|A_j)} \tag{4}$$

Jest to wzór na prawdopodobieństwo *á posteriori*, gdyż dotyczy prawdopodobieństwa zajścia zdarzenia $A_i$ po zajściu zdarzenia $B$. Prawdopodobieństwa $A_i$ noszą nazwę prawdopodobieństw *á priori* lub prawdopodobieństw subiektywnych.

# Twierdzenie Bayesa

## Twierdzenie Bayesa - przykład

Żarówki są produkowane w 3 fabrykach. Z fabryki pierwszej pochodzi 25% produkcji, z fabryki drugiej 35% produkcji a z trzeciej 40%.
Produkcja wadliwa wynosi odpowiednio:

- dla fabryki I 5%
- dla fabryki II 4%
- dla fabryki III 2%

Wybrana żarówka okazała się wadliwa - jakie jest prawdopodobieństwo, że pochodzi ona z fabryki pierwszej?

# Twierdzenie Bayesa

## Twierdzenie Bayesa - przykład

**Rozwiązanie**

Oznaczenia:

- $A_1$ - wybrana żarówka pochodzi z fabryki I
- $A_2$ - wybrana żarówka pochodzi z fabryki II
- $A_3$ - wybrana żarówka pochodzi z fabryki III
- $B$ - wybrana żarówka jest wadliwa

Szukamy $P(A_1|B)$.

# Twierdzenie Bayesa

## Twierdzenie Bayesa - przykład

**Rozwiązanie c.d.**

Mamy: $P(A_1) = 0.25; \quad P(A_2) = 0.35; \quad P(A_3) = 0.40;$
$P(B|A_1) = 0.05; \quad P(B|A_2) = 0.04; \quad P(B|A_3) = 0.02;$
czyli ostatecznie

$$P(A_1|B) = \frac{0.25 \cdot 0.05}{(0.25 \cdot 0.05) + (0.35 \cdot 0.04) + 0.40 \cdot 0.02} = \frac{0.0125}{0.0345} = 0.362318841$$

# Twierdzenie Bayesa

## Teoria decyzji

W teorii decyzji $A_1, A_2, \ldots$ oznaczają stany natury a $P(A_i)$ jest rozkładem *á priori* tych stanów (jeśli natura jest losowa), a $P(B|A_j)$ jest wiarygodnością wyniku $B$ dla stanów $A_j$ ($B$ oznacza, że wynik doświadczenia wynosi $B$).
Prawdopodobieństwa $P(A_i|B)$ oznaczają zatem rozkład *á posteriori* stanów natury po przeprowadzeniu doświadczenia o wyniku $B$.
Podejście (produkt) Bayesa w teorii decyzji przyjmuje, że $P(A_i)$ są zawsze znane i podsumowują naszą wiedzę i/lub intuicję o badanej rzeczywistości. Liczby te stanowią wagi w funkcji średniej użyteczności, którą maksymalizuje tzw. wariant Bayesowski.

# Twierdzenie Bayesa

## Teoria decyzji

Zauważmy, że podejście to nie zakłada losowości natury. Jeśli natomiast założymy, że natura jest losowa to tworzą one rozkład *á priori* jej stanów. Jeśli ponadto mamy możliwość przeprowadzenia doświadczenia o wyniku $B$ to rozkład ten możemy uściślić korzystając ze wzoru Bayesa.
Twierdzenie Bayesa można uogólnić na sytuację, gdy zamiast jednego zdarzenia $B$ mamy układ zdarzeń $B_1, B_2, \ldots$ również tworzący podział przestrzeni $\Omega$ taki, że $P(B_k) > 0$, $k = 1, 2, \ldots$.
Wówczas dla każdego $i$ i $k$ zachodzi:

$$P(A_i|B_k) = \frac{P(A_i)P(B_k|A_i)}{\sum\limits_{j=1}^{\infty} P(A_j)\,P(B_k|A_j)}, \quad i, k = 1, 2, \ldots \tag{5}$$

# Twierdzenie Bayesa

## Niezależność zdarzeń

Jeżeli prawdopodobieństwo zdarzenia $A$ pod warunkiem, że zaszło zdarzenie $B$ jest różne od prawdopodobieństwa bezwarunkowego zdarzenia $A$ to znaczy, że zajście zdarzenia $B$ dostarcza pewnej informacji o zdarzeniu $A$, czyli zdarzenia $A$ i $B$ są zależne.
Jeżeli natomiast $P(A|B) = P(A)$ to zdarzenia $A$ i $B$ są niezależne.
Zauważmy, że zachodzi to wówczas gdy $P(A \cup B) = P(A) \cdot P(B)$ (co wynika z wzoru). Wówczas także

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

# Twierdzenie Bayesa

## Niezależność zdarzeń

**Definicja**

Zdarzenia $A$ i $B$ są niezależne jeśli $P(A \cup B) = P(A) \cdot P(B)$ .

Poniższe zdania są prawdziwe:

- dowolne zdarzenie $A$ i zdarzenie pewne są niezależne
- dowolne zdarzenie $A$ i zdarzenie niemożliwe są niezależne
- zdarzenie pewne i niemożliwe są niezależne
- jeżeli zdarzenia $A$ i $B$ są niezależne, to $A$ i $B'$ są też niezależne

Pojęcie niezależności zdarzeń łatwo uogólnia się na większą ich liczbę.

# Naiwny klasyfikator bayesowski

Naiwny klasyfikator bayesowski jest prostym probabilistycznym klasyfikatorem. Naiwne klasyfikatory bayesowskie są oparte na założeniu o wzajemnej niezależności predyktorów (zmiennych niezależnych). Często nie mają one żadnego związku z rzeczywistością i właśnie z tego powodu nazywamy je naiwnymi. Bardziej opisowe może być określenie - „model cech niezależnych". Model prawdopodobieństwa można wyprowadzić korzystając z twierdzenia Bayesa.

# Naiwny klasyfikator bayesowski

W zależności od rodzaju dokładności modelu prawdopodobieństwa, naiwne klasyfikatory bayesowskie można „uczyć" bardzo skutecznie w trybie uczenia z nadzorem. W wielu praktycznych aplikacjach, estymacja parametru dla naiwnych modeli Bayesa używa metody **maksymalnego prawdopodobienstwa** (maximal likelihood) *á posteriori*. Inaczej mówiąc, może pracować z naiwnym modelem Bayesa bez wierzenia w twierdzenie Bayesa albo używania jakichś metod Bayesa.
Pomimo ich naiwnego projektowania i bardzo uproszczonych założeń, naiwne klasyfikatory Bayesa często pracują dużo lepiej w wielu rzeczywistych sytuacjach niż można było tego oczekiwać.

# Naiwny klasyfikator bayesowski

Model prawdopodobieństwa dla klasyfikatora jest modelem warunkowym $P(C|F_1, F_2, \ldots, F_n)$ przez zmienną zależną klasy $C$ z niewielu rezultatów albo „klas", zależnych od kilku opisujących zmiennych $F_1, \ldots, F_n$. Problem pojawia się, gdy liczba cech $n$ jest duża lub gdy cecha może przyjmować dużą liczbę wartości. Wtedy opieranie się na modelu tablic prawdopodobieństw jest niewykonalne. Dlatego też inaczej formułujemy taki model, by był bardziej przystępny.

Korzystając z twierdzenia Bayesa piszemy:

$$P(C|F_1, F_2, \ldots, F_n) = \frac{P(C)P(F_1, \ldots, F_n|C)}{P(F_1, \ldots, F_n)}$$

W praktyce interesuje nas tylko licznik ułamka, bo mianownik nie zależy od $C$ i wartości cechy $F_i$ sa dane. Mianownik jest więc stały.

# Naiwny klasyfikator bayesowski

Licznik ułamka jest równoważny łącznemu prawdopodobieństwu $P(C, F_1, \ldots, F_n)$, który można zapisać, wykorzystując prawdopodobieństwo warunkowe

$$
\begin{aligned}
& P(C, F_1, \ldots, F_n) = \\
& = P(C)P(F_1, \ldots, F_n|C) \\
& = P(C)P(F_1|C)P(F_2, \ldots, F_n|C, F_1) \\
& = P(C)P(F_1|C)P(F_2|C, F_1)P(F_3, \ldots, F_n|C, F_1, F_2) \\
& \ldots\ldots\ldots
\end{aligned}
$$

# Naiwny klasyfikator bayesowski

Włączamy teraz „naiwną'' warunkową zależność. Zakładamy, że każda cecha $F_i$ jest warunkowo niezależna od każdej innej cechy $F_j$ dla $i \neq j$, co oznacza, że

$$P(F_i|C, F_j) = P(F_i|C)$$

więc model można wyrazić jako

$$P(C, F_1, \dots, F_n) = P(C) \prod_{i=1}^{n} P(F_i|C)$$

# Naiwny klasyfikator bayesowski

Naiwny klasyfikator bayesowski ma wiele własności, które okazują się zaskakująco przydatne w praktyce, pomimo faktu, że niezależne założenia często są naruszone. Jak wszystkie probabilistyczne klasyfikatory, wykorzystujące regułą decyzyjną MAP, klasyfikacja jest tak długo poprawna, jak długo poprawna klasa jest bardziej prawdopodobna od innych (prawdopodobieństwa poszczególnych klas nie muszą być oceniane zbyt dokładnie). Inaczej mówiąc, klasyfikator jest wystarczająco mocny, by zignorować poważne niedociągnięcia naiwnego probabilistycznego modelu.

# Naiwny klasyfikator bayesowski

## Przykład: klasyfikacja dokumentu

Rozważymy klasyfikację poczty email pod względem zawartości i będziemy oceniać, czy poszczególne wiadomości są chcianą pocztą czy też spamem. Wyobraźmy sobie, że dokumenty są przypisane do pewnej liczby klas dokumentów, które mogą być modelowane jako komplety słów, gdzie (niezależne) prawdopodobieństwo, że $i$-te słowo danego dokumentu zdarza się w dokumencie klasy $C$ zapisujemy, jako

$$P(w_i|C)$$

# Naiwny klasyfikator bayesowski

## Przykład: klasyfikacja dokumentu

Zakładamy, że prawdopodobieństwo wystąpienia słowa w dokumencie jest niezależne od długości dokumentu lub też, że wszystkie dokumenty mają tę samą długość. W tym przypadku prawdopodobieństwo danego dokumentu $D$ do klasy $C$ wynosi

$$P(D|C) = \prod_i P(w_i|C)$$

# Naiwny klasyfikator bayesowski

## Przykład: klasyfikacja dokumentu

Pytanie, na jakie chcemy odpowiedzieć brzmi:

jakie jest prawdopodobieństwo, że dany dokument $D$ należy do danej klasy $C$?

Korzystając z definicji

$P(D|C) = \dfrac{P(D \cap C)}{P(C)}$ i $P(C|D) = \dfrac{P(D \cap C)}{P(D)}$

dostajemy

$$P(C|D) = \frac{P(C)}{P(D)} P(D|C) \tag{6}$$

# Naiwny klasyfikator bayesowski

## Przykład: klasyfikacja dokumentu

Przyjmijmy założenie, że są tylko dwie klasy: $S$ i $\neg S$ (w naszym przykładzie: spam i nie-spam). Mamy:

$$P(D|S) = \prod_i P(w_i|S) \quad \text{i} \quad P(D|\neg S) = \prod_i P(w_i|\neg S)$$

Korzystając ze wzoru (6), możemy powyższy rezultat zapisać jako

$$P(S|D) = \frac{P(S)}{P(D)} \prod_i P(w_i|S) \quad \text{i} \quad P(\neg S|D) = \frac{P(\neg S)}{P(D)} \prod_i P(w_i|\neg S)$$

# Naiwny klasyfikator bayesowski

## Przykład: klasyfikacja dokumentu

Dzieląc jeden przez drugi otrzymujemy:

$$\frac{P(S|D)}{P(\neg S|D)} = \frac{\frac{P(S)}{P(D)} \prod_i P(w_i|S)}{P(\neg S|D) = \frac{P(\neg S)}{P(D)} \prod_i P(w_i|\neg S)}$$

wynik ten można przedstawić w postaci

$$\frac{P(S|D)}{P(\neg S|D)} = \frac{P(S)}{P(\neg S)} \prod_i \frac{P(w_i|S)}{P(w_i|\neg S)}$$

# Naiwny klasyfikator bayesowski

## Przykład: klasyfikacja dokumentu

W ten sposób, prawdopodobieństwo stosunku $P(S|D)/P(\neg S|D)$ może być wyrażone jako stosunek prawdopodobieństw. Bieżące prawdopodobieństwo $P(S|D)$ można obliczyć jako $\log(P(S|D)/P(\neg S|D))$, korzystając z własności, że $P(S|D) + P(\neg S|D) = 1$.
Otrzymujemy więc:

$$\ln \frac{P(S|D)}{P(\neg S|D)} = \ln \frac{P(S)}{P(\neg S)} + \sum_i \ln \frac{P(w_i|S)}{P(w_i|\neg S)}$$

# Naiwny klasyfikator bayesowski

## Przykład: klasyfikacja dokumentu

Teraz możemy sklasyfikować dany dokument. Jest to spam, jeśli

$$\ln \frac{P(S|D)}{P(\neg S|D)} > 0$$

W innym wypadku dokument spamem nie jest.

# Wnioskowanie statystyczne

Wnioskowanie statystyczne sprowadza się do weryfikowania pewnych hipotez formułowanych na podstawie założonego modelu teoretycznego.Obejmuje następujące czynności:

- Sformułowanie hipotezy zerowej i hipotezy alternatywnej.
- Ustalenie poziomu istotności.
- Wybór statystyki do weryfikacji hipotezy $H_0$ i ustalenie obszaru krytycznego (wartości krytycznych).
- Obliczenie wartości statystyki w próbie.
- Sformułowanie wniosków (weryfikacja hipotezy $H_0$) przez porównanie wartości obliczonej statystyki z wartościami krytycznymi; będzie to jeden z dwóch wniosków:
  1. odrzuca się hipotezę zerową i za prawdziwą uznaje się hipotezę alternatywną,
  2. nie ma podstaw do odrzucenia $H_0$ (co nie oznacza jej przyjęcia).

# Wnioskowanie statystyczne

W rozwijanych statystycznych modelach wnioskowania bardzo często wykorzystuje się twierdzenie Bayesa, będące podstawą teorii podejmowania decyzji i metody weryfikacji hipotez statystycznych. Baza wiedzy takich systemów jest zbiorem stwierdzeń i relacji o pewnym wycinku świata rzeczywistego, której ona dotyczy.

# Klasyfikacja bayesowska

Klasyfikacja bayesowska, to metoda budowy systemu ekspertowego, w której wiedza przedstawiona jest *á priori* z warunkowymi prawdopodobieństwami i funkcjami wyróżniajacymi, a wnioskowanie polega na liczeniu następnych prawdopodobieństw oraz liczeniu wyróżniającej cechy. Mechanizm wnioskowania wykorzystujący twierdzenie Bayesa polega na obliczaniu prawdopodobieństwa każdego możliwego wyniku, gdy znany jest dany konkretny przypadek. Wadą tej metody jest fakt, że wymaga ona znajomości dokładnych wartości lub rozkładów prawdopodobieństw pojawienia się parametrów zjawiska, czyli problemu będącego przedmiotem rozważań. Innym problemem jest to, że należy dokonać pewnych nierealistycznych założeń - na przykład w klasyfikacji bayesowskiej wymagane wyniki, np. rozpoznawania, musza się wzajemnie wykluczać. Niestety w wielu przypadkach mogą występować liczne podobne wyniki (np. w diagnostyce: pacjent może mieć wiele chorób). Innym założeniem, co prawda niewymaganym przez twierdzenie Bayesa, ale wymuszonym przez praktykę, jest statystyczna niezależność cechy problemu.

# Klasyfikacja bayesowska

W bardzo wielu dziedzinach zastosowań cechy problemu nie są niezależne, więc założenie to może znacznie obniżyć skuteczność systemu. Kolejną wadą większości systemów statystycznych jest to, że przed zastosowaniem systemu trzeba posiadać wszystkie istotne informacje o danym przypadku. Może to się okazać nierealne. Jednak w uzasadnieniu wyżej wymienionych wad systemy oparte np. na statystycznej klasyfikacji obrazów bardzo dobrze nadają się tam, gdzie wiedza jest niepełna bądź niepewna. Ze względu na częstą niepewność systemy oparte na tej metodzie wydaja się być bardzo przydatne, jeśli nie niezbędne.

# Sieci Bayesa

Koncepcja sieci Bayesa wynika wprost z koncepcji prawdopodobieństwa warunkowego. Jak się okazuje w rzeczywistym świecie jest wiele sytuacji w których wystąpienie jakiegoś zdarzenia ściśle zależy od innego zdarzenia. Zastosowanie sieci Bayesa pozwala na uniknięcie obliczeń o dużej złożoności – obliczenie jednego prawdopodobieństwa a posteriori łączy się z uprzednim obliczeniem wykorzystywanych prawdopodobieństw.
Sieci Bayesa służą do przedstawiania niepewności wiedzy. Niepewność wiedzy używanej zawartej w systemach ekspertowych może mieć wiele czynników:

- niepewność ekspertów dotycząca ich wiedzy
- niepewność tkwiąca w modelowanej dziedzinie
- niepewność inżyniera próbującego przetłumaczyć wiedze
- niepewność wynikła z dokładności dostępnej wiedzy

# Sieci Bayesa

Sieci Bayesa używają teorii prawdopodobieństwa do określenia niepewności przez jawne reprezentowanie warunkowych zależności pomiędzy różnymi częściami wiedzy. Pozwala to na intuicyjną graficzną wizualizację wiedzy zawierającą wzajemne oddziaływania pomiędzy różnymi źródłami niepewności
Sieci Bayesa są stosowane w diagnostyce, w rozumowaniu przebiegającym od efektów do przyczyn i odwrotnym. W systemach ekspertowych sieci Bayesa znalazły zastosowanie w medycynie (systemy doradcze, które rozpoznają chorobę na podstawie podawanych objawów).

# Koniec?

Koniec wykładu 2