Curs 2/2: Formula lui Bayes

Conf.dr. Maria Jivulescu

Departamentul de Matematică
UPT

Cuprins

- Formula lui Bayes
- Formula probabilității totale
- Aplicații

Recapitulare probabilitate condiționată

Notații

- dat evenimentul E_1, avem $P(E_1)$;
- date două evenimente E_1, E_2, avem $P(E_1 \cap E_2) = P(E_1, E_2)$ probabilitatea ca evenimentele să se realizeze simultan (joint probability)
- **probabilitate condiționată**: $P(E_2|E_1) := \frac{P(E_1 \cap E_2)}{P(E_1)}$
- regula produsului: $P(E_1 \cap E_2) = P(E_1)P(E_2|E_1)$
- simetria probabilității ev. intersecție: $P(E_1 \cap E_2) = P(E_2 \cap E_1)$
- avem: $P(E_2|E_1) = \frac{P(E_1 \cap E_2)}{P(E_1)} = \frac{P(E_1|E_2)P(E_2)}{P(E_1)}$
- **prob cond. nu este simetrica**: $P(E_1|E_2) \neq P(E_2|E_1)$
- avem: $P(E_2|\bar{E}_1) = 1 - P(\bar{E}_2|\bar{E}_1)$
- ev. independente: $P(E_2|E_1) = P(E_2) \Leftrightarrow P(E_1 \cap E_2) = P(E_1)P(E_2)$

Introducere in problematică

Avem următorul joc: se dau 3 monede, cu regulile că

- una dintre ele este falsă, adică aruncată cade pe ”cap” cu probabilitate $2/3$
- nu se precizează care din monede este falsă.

Se aruncă cele 3 monede si avem următorul rezultat:

la prima și a doua monedă se obține cap, iar la a treia se obține pajură.

Care este probabilitatea ca prima monedă aruncată să fie cea falsă?

Observații:

- Fie H_i : moneda i aruncată este cea falsă; Avem $P(H_i) = \frac{1}{3}$.
- A : pentru cele trei monede se obține: cap, cap, pajură
- Avem $P(A|H1) = P(A|H_2) = \frac{2}{3} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{6}$ și $P(A|H_3) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{12}$
- Dorim să calculăm $P(H_1|A) =?$

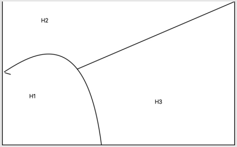
Din formula lui Bayes, vom obține $P(H_1|A) = \frac{2}{5}$, deci rezultatul aruncării celor trei monede crește probabilitatea ca prima monedă să fie cea falsă de la $1/3$ la $2/5$.

Ingredientele formulei lui Bayes

Notații

- $(\Omega, \mathcal{K}, P)$ un câmp de probabilitate
- Fie evenimentele $H_1, H_2, \ldots H_n \in \mathcal{K}$, ce formează o partiție a evenimentului sigur Ω:
  - $\cup_{i=1}^{n} H_i = \Omega$
  - $H_i \cap H_j = \emptyset$, $i \neq j$
- Evenimentele H_i, $i \in \{1, n\}$ se numesc **ipoteze**.
  Ipotezele sunt acceptate cu o anumită probabilitate, $P(H_i)$.
- Avem:

$$1 = P(\Omega) = P(H_1) + P(H_2) + \cdots + P(H_n)$$



Ingredientele formulei lui Bayes

- Dacă apare o nouă informație A, atunci nivelul de veridicitate (verosimilitate) al acestei informații, se reprezintă prin probabilitățile condiționate $P(A|H_k)$, $k \in \{1, n\}$.
- $P(A|H_k)$ este nivelul de verosimilitate al informației A în condițiile acceptării ipotezei H_k.
- $P(H_k)$- probabilităților ipotezelor se mai numesc **probabilități apriorice**

Cum se calculează probabilitatea $P(A)$, adică nivelul de veridicitate al informației A?

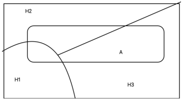


Figure: Evenimentele implicate în formula lui Prob Totale

Formula probabilității totale

Formula probabilității totale

Dacă $A \in \mathcal{K}$ este un eveniment informație, atunci gradul/nivelul de veridicitate al acestei informații este:

$$P(A) = \sum_{i=1}^{n} P(H_i)P(A|H_i)$$

Demonstrație:

- $H_1, H_2, \ldots, H_n$ fiind mutual exclusive două câte două, deci și $(A \cap H_i), (A \cap H_j)$, $i \neq j$, mutual exclusive două câte două;
- scriem $A = (A \cap H_1) \cup (A \cap H_2) \cup \cdots \cup (A \cap H_n)$;
- $P(A) = \sum_{i=1}^{n} P(A \cap H_i)$;
- Din formula probabilității condiționate avem: $P(A \cap H_i) = P(H_i)P(A|H_i)$,

Formula lui Bayes rectifică, actualizează probabilitățile ipotezelor, pe baza informației A, și anume:

$$P(H_k|A) = \frac{P(H_k)P(A|H_k)}{\sum_{i=1}^{n} P(H_i)P(A|H_i)} \quad k \in \{1, n\}$$

Demonstrație Exprimăm $P(H_k \cap A)$ în două moduri:$P(H_k \cap A) = P(H_k)P(A|H_k) = P(A)P(H_k|A)$.
Din ultima egalitate avem:

$$P(H_k|A) = \frac{P(H_k)P(A|H_k)}{P(A)} = \frac{P(H_k)P(A|H_k)}{\sum_{i=1}^{n} P(H_i)P(A|H_i)}, \quad k \in \{1, n\}$$

$P(H_k|A)-$se numesc probabilități posterioare

Aplicații

Formula lui Bayes rectifică, actualizează probabilitățile ipotezelor, pe baza unei noi informații și are multiple aplicații: [1]

- Analiza testelor de diagnostic medical
- Modelarea ipotezelor in Machine Learning (model probabilist pentru descrierea relației dintre date și ipoteze;
- Clasificari=model predictiv de a eticheta eșantioane de date (Naive Bayes Classifier);
- Clasificatorul Bayes optimal (model probabilist care face o predicție pentru un nou exemplu.)
- algoritmi probabiliști[2]

[1] A Gentle Introduction to Bayes Theorem for Machine Learning

[2] Probability and Computing, M. Mitzenmacher, Eli Upfal

Aplicație

O companie medicală testează un nou medicament.
Rata fals negativă este mică: dacă ai boala, probabilitatea ca testul să fie negativ este 0.001. **Rata fals pozitivă** este tot mică: dacă nu ai boala, probabilitatea ca testul să fie pozitiv este 0.005. Se presupune că 2% din populație are boala. Dacă o persoana este aleasă aleator uniform și este testată pozitiv, care este probabilitatea ca acestă persoana să aibă boala?

Rezolvare:

- Ev. ipoteza: H_1 persoana are boala, $P(H_1) = 0.02$
- Ev ipoteza: $H_2 = \bar{H}_1$: persoana nu are boala, $P(H_2) = 1 - P(H_1)$
- Eveniment informatie: A_+ : test pozitiv, $A_- = \bar{A}_+$: test negativ
  $P(A_-|H_1) = 0.001$, $P(A_+/H_1) = 0.999$, $P(A_+|H_2) = 0.005$
- $P(A_-|H_2) = 0.995$
- Trebuie calculat $P(H_1|A_+) =?$.
- $P(A_+) = P(H_1)P(A_+|H_1) + P(H_2)P(A_+|H_2)$
- fr. lui Bayes $P(H_1|A_+) = \frac{P(H_1)P(A_+|H_1)}{P(A_+)} = \frac{0.02\times0.999}{0.02\times0.999+0.98\times0.005} \approx 0.8$

Concluzii cu care să rămânem

- Ipotezele H_i fac o partiție a lui Ω: $(\sum_i P(H_i) = 1)$
- Evenimentul A -eveniment informație
- Prin formula probabilității totale determinăm $P(A)$
- Formula lui Bayes recalculeaza $P(H_i|A)$

La cursul următor vom discuta
Capitolul II: Variabile aleatoare discrete!