Seminar 4

Formula lui Bayes

4.1 Probleme rezolvate

1. Un pacient este suspectat de o boală rară. Diagnosticul se poate pune în urma unui test clinic. Există două ipoteze mutual exclusive: pacientul are boala respectivă (H_1) sau nu (H_2). Evident $H_2 = \overline{H_1}$.

Probabilităţile apriorice ale ipotezelor sunt furnizate de Organizaţia Mondială a Sănătăţii, care afirmă (pe baza nivelului ei de cunoaştere, rezultat din datele statistice) că 2 adulţi din 1000 au această boală, adică $P(H_1) = 2/1000 = 0.002$. Prin urmare $P(H_2) = 0.998$

Notăm cu A_+ evenimentul informaţie, ce semnifică test clinic pozitiv şi cu A_- informaţia, test clinic negativ.

Experienţa de laborator arată că în 97% din cazuri în care afecţiunea respectivă există, testul iese pozitiv, deci $P(A_+|H_1) = 0.97$, iar la 95% din cazuri, în care boala nu este prezentă, testul iese negativ, adică $P(A_-|H_2) = 0.95$) şi deci $P(A_+|H_2) = 0.05$ (deoarece P_{H_2} este o funcţie de probabilitate şi deci, $P_{H_2}\complement(A) = 1 - P_{H_2}(A)$);

Dacă testul pentru pacientul în discuţie iese pozitiv care este probabilitatea $P(H_1|A_+)$, ca pacientul să aibă boala X?

Rezolvare

Conform formulei Bayes avem:

$$P(H_1|A_+) = \frac{P(H_1)P(A_+|H_1)}{P(H_1)P(A_+|H_1) + P(H_2)P(A+|H_2)} = \frac{0.002 \times 0.97}{0.002 \times 0.97 + 0.998 \times (0.05)} =$$

$= 0.0388$ Prin urmare informaţia furnizată de test schimbă probabilităţile apriorice $P(H_1) = 0.002$, $P(H_2) = 0.998$ în probabilităţile posteriorice: $P(H_1|A_+) = 0.0388$, $P(H_2|A_+) = P(\overline{H_1}|A+) = 1 - P(H_1|A_+) = 1 - 0.0388 = 0.9612$ sau în cuvinte, probabilitatea ca pacientul să fie bolnav, dat fiind că testul a ieşit pozitiv este 0.0388, ceea ce în limbaj medical se traduce spunând că din 1000 de cazuri cu test pozitiv 38 au boala respectivă. Comparativ cu informaţia dată de Organizaţia Mondială a Sănătăţii, folosirea informaţiei de test pozitiv, modifică substanţial probabilitatea prezenţei bolii (de la 2 bolnavi dintr-o 1000, la 38 dintr-o mie).

2. La un examen de tip grilă fiecare întrebare are 5 răspunsuri asociate dintre care unul singur este corect. Un student ştie răspunsul corect la 65% dintre întrebări. La cele la care nu ştie răspunsul încercuieşte unul la întâmplare.

a) Care este probabilitatea ca studentul să dea un răspuns greşit la o anumită întrebare?

b) Ştiind că acesta a răspuns corect la acea întrebare, care este probabilitatea ca răspunsul să fie obţinut prin ghicire?

Rezolvare: a) Evenimentul sigur al experimentului aleator, "studentul răspunde la o anumită întrebare", se descompune în evenimentele mutual exclusive:

H_1 - studentul ştie răspunsul corect la întrebare;

H_2 - studentul nu ştie răspunsul corect la întrebare, adică dă răspuns la întâmplare la acea întrebare.

Fie A evenimentul "studentul dă răspuns greşit la întrebare". Formula probabilităţii totale este

$$P(A) = P(H_1)P(A|H_1) + P(H_2)P(A|H_2).$$

Din datele problemei, $P(H_1) = 0.65$, iar $P(H_2) = P(\overline{H}_1) = 1 - 0.65 = 0.35$. Pe de altă parte, $P(A|H_1) = 0$, iar $P(A|H_2) = 4/5 = 0.8$. Astfel, $P(A) = 0.35 \times 0.8 = 0.28$.

b) Avem de calculat

$$P(H_2|\overline{A}) = \frac{P(H_2)P(\overline{A}|H_2)}{P(\overline{A})} = \frac{0.35 \times (1-0.8)}{1-0.28} = \frac{7}{72}.$$

□

3. Un program are două module. Primul modul conţine erori cu probabilitatea de 0.2, iar al doilea, fiind mai complex, conţine erori cu probabilitatea de 0.4, independent de erorile din primul modul. Existenţa unei erori doar în primul modul face ca programul să eşueze cu o probabilitate de 0.5, în timp ce o eroare doar în al doilea modul conduce la blocajul programului cu o probabilitate de 0.8. Existenţa unei erori în ambele module produce blocarea programului cu o probabilitate de 0.9. Dacă programul a eşuat, care este probabilitatea ca eşecul să fi fost produs de erori din ambele module (o eroare în primul modul şi o eroare în al doilea modul)?

Rezolvare: Considerăm A evenimentul informaţie "programul a eşuat". Notând cu

E_1: "existenţa unei erori în primul modul",

E_2: "existenţa unei erori în al doilea modul",

o partiţie a spaţiului de selecţie va fi dată de evenimentele (ipoteze):

$$H_1 = E_1 \cap E_2,\ H_2 = \overline{E_1} \cap E_2,\ H_3 = E_1 \cap \overline{E_2},\ H_4 = \overline{E_1} \cap \overline{E_2}.$$

Evenimentul H_1 înseamnă "existenţa unei erori în primul modul şi existenţa unei erori în al doilea modul", H_2: "existenţa unei erori doar în al doilea modul", H_3: "existenţa unei erori doar în primul modul", iar H_4: "în ambele module nu există nici o eroare".

Evenimentele E_1, E_2 fiind independente, deci şi evenimentele $\overline{E_1}, E_2$, respectiv $E_1, \overline{E_2}$, respectiv $\overline{E_1}, \overline{E_2}$ sunt independente, se obţine

$$P(H_1) = P(E_1)\,P(E_2) = 0.2 \times 0.4 = 0.08,\ P(H_2) = P(\overline{E_1})\,P(E_2) = 0.8 \times 0.4 = 0.32,$$

$$P(H_3) = P(E_1)\,P(\overline{E_2}) = 0.2 \times 0.6 = 0.12,\; P(H_4) = P(\overline{E_1})\,P(\overline{E_2}) = 0.8 \times 0.6 = 0.48.$$

Aplicând formula probabilităţii totale, avem

$$\begin{aligned} P(A) &= P(H_1)P(A|H_1) + P(H_2)P(A|H_2) + P(H_3)P(A|H_3) + P(H_4)P(A|H_4) \\ &= 0.08 \times 0.9 + 0.32 \times 0.8 + 0.12 \times 0.5 + 0.48 \times 0 = 0.388. \end{aligned}$$

Se cere

$$P(H_1|A) = \frac{P(H_1)P(A|H_1)}{P(A)} = \frac{0.08 \times 0.9}{0.388} = \frac{72}{388}.$$

□

4.2 Probleme propuse

4. Fie H_1, H_2 o partiţie a spaţiului de selecţie Ω şi A un eveniment astfel încât $P(H_1) = 0.4$, $P(A|\overline{H_1}) = 0.2$ şi $P(A|\overline{H_2}) = 0.7$. Să se calculeze $P(\overline{H_1})$, $P(H_2)$, $P(A|H_1)$ şi $P(A)$.

5. Setul de date de intrare pentru un program conţine în proporţie de 70% date de tipul I şi în proporţie de 30% date de tipul II. Intrarile pot produce mesaje de atenţionare în proportie de 20% (cele de tipul I), respectiv 10% (cele de tipul II). Dacă dupa rulare este afişat un mesaj de atenţionare, care este probabilitatea ca el sa fie cauzat de datele de intrare de tipul I?

6. Intr-un sistem de comunicatie digitală, biţii transmişi sunt eronaţi din cauza zgomotului din canalul de transmitere şi astfel sunt receptaţi eronat. Notăm cu E_b, evenimentul "s-a transmis bitul b", b = 0, 1, si cu R_b, evenimentul "s-a receptat bitul b", b = 0, 1. Ştiind că $P(R_0|E_0) = 0.7, P(R_1|E_1) = 0.8$ şi că bitul 0 este transmis cu probabilitatea 0.6, să se calculeze probabilitatea să se transmita bitul 0, ştiind că s-a recepţionat bitul 1.

7. Un robot se mişcă într-un spaţiu de lucru (de exemplu, o cameră). El are un senzor care poate măsura distanţa până la orice obiect din preajmă. Pe baza informaţiei primite de la senzor, sistemul de calcul încorporat calculează probabilitatea ca uşa camerei să fie deschisă. Notăm cu H evenimentul "uşa este deschisă", care are probabilitatea 0.6. Robotul primeşte informaţia de la senzor că o anumită distanţă până în zona uşii este d, privită ca o observaţie asupra unei variabile aleatoare D. Din experienţa robotului în spaţiul de lucru, sistemul de calcul are stocată informaţia că $P(D = d|H) = 0.7$, respectiv $P(D = d|\overline{H}) = 0.1$. Să se calculeze probabilitatea ca uşa să fie deschisă ştiind că distanţa măsurată este d.

8. **Monty Hall problem**
Prezentăm acum una din cele mai cunoscute probleme din teorie probabilităţii, fiind legată de un concurs de televiziune din anii '70, Monty Hall: Într-un concurs televizat ţi se oferă

posibilitatea alegerii dintre trei uşi, in spatele unei aflându-se un automobil, iar în spatele celorlalte, capre. Tu alegi o uşă, să zicem nr.1, iar gazda emisiunii, care ştie ce se află în spatele uşilor, deschide uşa numărul 2, în spatele căreia se află o capră. Apoi, te întreabă: "vrei să alegi uşa cu numărul 3?". Este în avantajul tău să schimbi alegerea iniţială?
Pentru mai multe informaţii va invităm sa citiţi pagina Wikipedia dedicată acestui subiect.