

Tema 9 - laborator 11

Regresie logistică Bayesiană

Ex. 1. Se dorește modelarea probabilității ca un student să promoveze un examen în funcție de numărul de ore de studiu pe săptămână. Datele sintetice arată astfel:

ore de studiu:	1	3	4	5	6	8
promovare (0/1):	0	0	1	1	1	1

Construiți un model bayesian de regresie logistică care:

- să estimeze coeficienții regresiei (interceptul și panta), probabilitatea de mai sus și frontiera de decizie;
- să ofere intervale de credibilitate (HDI) pentru frontiera de decizie.

Ex. 2. Ce factori determină admiterea la facultate în Statele Unite? În fișierul Admission.csv au fost strânse datele a 400 de cazuri de admitere la o facultate. "Admission" este un răspuns binar, cu 1 sau 0 indicând "admis", respectiv "respins". Sunt de asemenea disponibile scorul la testul GRE și rezultatul mediu din liceu, GPA (undergraduate grade point average). Fie p_i probabilitatea ca studentul cu nr. i să fie admis.

Considerăm modelul logistic

$$p_i = \text{logistic}(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}),$$

unde x_{i1} și x_{i2} sunt scorul GRE, respectiv GPA pentru studentul i .

- Folosind distribuții a priori slab informative asupra parametrilor β_0 , β_1 și β_2 , folosiți PyMC pentru a simula un eșantion suficient de mare (construi modelul) din distribuția a posteriori.
- Care este, în medie, granița de decizie pentru acest model? Reprezentați de asemenea grafic o zonă în jurul acestei grafic care să reprezinte un interval 94% HDI.
- Să presupunem că un student are un scor GRE de 550 și un GPA de 3.5. Construiți un interval de 90% HDI pentru probabilitatea ca acest student să fie admis.
- Dar dacă studentul are un scor GRE de 500 și un GPA de 3.2? (refaceți exercițiul anterior cu aceste date) Cum justificați diferența?