

Universitatea “Al. I. Cuza” din Iași
Facultatea de informatică
Anul III, 2023–2024

Numele studentului: _____
Grupa studentului: _____

Examen PMP (29.1.2024)
Test 2
 timp de lucru: 50 min.

Răspunsul de la teste va trebui încărcat, până la finalul testului, pe “repozitorul” personal de pe GitHub, într-un director numit “Examen”. Pe lângă codul în Python, vor trebui încărcate și screenshoturi ale execuțiilor, precum și eventualele grafice.

Subiectul 1. Preziceți valoarea locuințelor ocupate de proprietari (în mii de USD) în zona Boston pe baza diferiților factori, cum ar fi numărul mediu de camere, rata criminalității și suprafața comercială din oraș. Fișierul `BostonHousing.csv` conține un set de date cu observații, din care ne interesează doar cele corespunzătoare valorii locuințelor (*medv*), numărul mediu de camere (*rm*), rata criminalității (*crim*) și proporția suprafeței comerciale non-retail (*indus*).

- Încărcați setul de date într-un Pandas DataFrame.
- Definiți modelul în PyMC folosind variabilele independente (*rm*, *crim*, *indus*) pentru a prezice variabila dependentă (*medv*).
- Obțineți estimări de 95% pentru HDI ale parametrilor. Care credeți că este variabila care influențează cel mai mult rezultatul?
- Simulați extrageri din distribuția predictivă posterioară și utilizați aceste extrageri simulate pentru a găsi un interval de predicție de 50% HDI pentru valoarea locuințelor.

Subiectul 2. Să presupunem că aruncați (la infinit) cu banul și că ați obținut o stemă abia la a 5^a aruncare. Doriți să determinați distribuția a posteriori pentru probabilitatea θ de a obține stemă la o aruncare, corespunzătoare acestei observații.

- Modificați codul de la pag. 7 din cursul 11 (ce implementează metoda *grid-computing*) astfel încât în loc de numărul de steme la aruncarea unei monede să modelați prima apariție a unei steme.
- Care este θ care maximizează probabilitatea a posteriori (estimativ)?

Indicație: Probabilitatea de apariție a unei steme abia la a n^a aruncare este modelată de distribuția Geometric(θ) (obținută, de exemplu cu `stats.geom.pmf(n , θ)` în SciPy).

Tabel punctaj:

subiect:	1.a.	1.b.	1.c.	1.d.	2.a.	2.b.	oficiu	Total
punctaj max.:	2	7	4	6	7	3	1	30
punctaj obținut:								