

Laborator 3

Considerăm că performanța Y a unui student la examenul de Inteligență Artificială ($Y=1$ studentul trece, $Y = 0$ studentul nu trece) depinde de numărul de ore X studiate săptămânal. Presupunem că funcția de probabilitate condiționată $\eta(x) = P(Y=1|X=x)$ este monoton crescătoare și are forma următoare:

$$\eta(x) = P(Y=1|X=x) = x/(x+c),$$

unde c este o constantă > 0 .

Realizați următoarele:

- scrieți funcția ***genereazaMultimeAntrenare.m*** care generează o mulțime de antrenare de volum n de perechi (X,Y) unde X are repartiția uniformă pe $[0, 4c]$ (c este un parametru dat, spre exemplu luați $c = 10$ pentru primele rulări) iar Y are repartiție Bernoulli de parametru $p = \eta(x)$;
- scrieți funcția ***aplicaClasificatorBayesian.m*** care aplică clasificatorul Bayesian g^* exemplelor X din mulțimea de antrenare obținând $g^*(X)$;
- scrieți funcția ***calculeazaEroareMisclasare.m*** care calculează eroarea de misclasare a clasificatorului Bayesian g^* pentru mulțimea de antrenare curentă de volum n și comparați cu eroarea teoretică ($= 0.305785$);
- scrieți funcția ***ploteazaEroareMisclasare.m*** care construiește graficul evoluției erorii de misclasare variind $n = 10, 100, 1000, 10000, 100000, 1000000$. Folosiți ca axă Ox funcția $\log(n)$ în loc de n (funcția semilogx în Matlab). Adăugați figurii un titlu și o legendă care să însoțească cu explicații rezultatele obținute.
- folosiți alte valori pentru parametrul c . Cum se modifică graficul erorii de misclasare?
- ce se va întâmpla cu eroare de misclasare dacă în loc să generăm etichetele Y ca realizările unei variabile aleatoare Bernoulli de parametru $p = \eta(x)$ le generăm cu parametrul p dat.
- reluati întregul exercițiu (a-e) pentru X având repartiție exponențială de parametru λ dat (încercați mai întâi cu $\lambda = c$).