

Laborator 4

Considerăm problema de regresie descrisă la cursul 3 în **Slides Lecture 3b**: dată fiind o mulțime de exemple de antrenare $S = \{(x_i, u_i)\}_{i=1,n}$ găsiți polinomul optim care descrie cât mai bine corespondența dintre punctele x_i și etichetele u_i .

Realizați următoarele:

- scrieți funcția **genereazaExemple.m** care generează n exemple (x_i, u_i) unde $u_i = f(x_i) + \varepsilon_i$, cu x_i repartizat uniform pe $[0, 1]$, ε_i zgomot repartizat normal de medie 0 și deviație standard σ și f funcția definită ca un obiect *inline*.
- generați o mulțime de antrenare $S = \{(x_i, u_i)\}_{i=1,n}$ conținând $n = 10$ exemple cu $u_i = \sin(2\pi x_i) + \varepsilon_i$.
- scrieți funcția **ploteazaExemple.m** care plotează exemplele de forma (x_i, u_i) , cu x_i reprezentat pe axa Ox și u_i reprezentat pe axa Oy. Apelați funcția pentru mulțimea de antrenare S .
- scrieți funcția **gasestePolinomOptim.m** care pentru o mulțime de antrenare S și pentru un grad n găsește polinomul optim $P_{S,n}$ de grad n care minimizează suma erorilor pătratice dintre eticheta prezisă $P(x_i)$ și eticheta u_i . Pentru găsirea polinomului optim $P_{S,n}$ folosiți funcția Matlab *polyfit.m*.
- scrieți funcția **ploteazaGraficPolinom.m** care plotează graficul funcției polinomiale asociate polinomului $P_{S,n}$ pe intervalul $[0,1]$. Pentru evaluarea polinomului $P_{S,n}$ folosiți funcția Matlab *polyval.m*.
- scrieți funcția **calculeazaEroare.m** care calculează eroarea E dintre predicțiile realizate pentru punctele x_i și etichetele adevărate u_i .

Pentru $n = 0, 1, \dots, 9$ realizați:

- calculați $P_{S,n}$ și vizualizați graficul funcției polinomiale asociate lui $P_{S,n}$.
- calculați și vizualizați evoluția erorii $E_{S,n}$ pe mulțimea de antrenare în care predicțiile se realizează folosind $P_{S,n}$.
- generați o mulțime test T cu 10 exemple și calculați și vizualizați evoluția erorii $E_{T,n}$ pe mulțimea de exemple T în care predicțiile se realizează folosind $P_{S,n}$.
- împărțiți mulțimea de antrenare inițială S în două mulțimi S_1 (noua mulțime de antrenare – 7 exemple) și S_2 (mulțimea de validare – 3

exemple). Puteți folosi în acest sens funcția *randperm* pentru generarea indicilor. Antrenați modelul $P_{SI,n}$ pe S_1 (considerați gradul $n \leq 6$ întrucât aveți 7 exemple) și testați performanța modelului pe mulțimea S_2 . Alegeți din cele 7 de funcții polinomiale posibile pe cea cu cea mai mică eroare pe mulțimea de validare S_2 .