

Identificarea sistemelor – Laborator 3

Regresia liniară pentru aproximarea funcțiilor

Organizare

- Acest laborator se rezolvă independent de către fiecare student. Doar dacă există mai mulți studenți decât calculatoare, studenții se pot grupa câte doi la un calculator.
- Soluția constă din cod Matlab. Dezvoltați acest cod într-un singur script Matlab. Codul va fi verificat și rulat de către profesor, în timpul laboratorului, și prezența la laborator va fi validată doar dacă aveți o soluție funcțională și originală. Prezențele la toate laboratoarele trebuie validate pentru eligibilitate la examen. Cum cel mult două laboratoare se pot recupera la sfârșitul semestrului, acumularea a trei sau mai multe laboratoare lipsă la orice moment de timp duce automat la pierderea finală a eligibilității.
- Discutarea ideilor între studenți este permisă și chiar de dorit, dar copierea sau schimbul direct de cod este interzis. Încălcarea acestei reguli va duce la invalidarea soluției.

Descrierea laboratorului

În acest laborator vom lucra la aproximarea funcțiilor folosind regresia liniară cu aproximatoare polinomiale, vezi secțiunea de *Regresie liniară* din Partea 3 a suportului de curs, *Baze matematice*.

Se dă un set de date de intrare-ieșire, unde ieșirea este generată de o funcție necunoscută, neliniară dar statică. Ieșirea este afectată de zgomot, pe care-l vom presupune aditiv, Gaussian, și de medie zero. Funcția are o variabilă de intrare și tot una de ieșire. Va trebui dezvoltat un model pentru această funcție. Un al doilea set de date, generat de aceeași funcție, este furnizat pentru validarea modelului dezvoltat. Cele două seturi sunt furnizate într-un fișier de date MATLAB, conținând câte o structură (tip de date MATLAB) pentru fiecare set. Structura pentru antrenarea modelului este numită `id`, iar cea pentru validare `val`. Fiecare din aceste structuri conține datele de intrare în vectorul X , și cele de ieșire corespunzătoare în vectorul Y .

Fiecărui student i se alocă de către profesor un index pentru setul de date. Apoi, studentul descarcă fișierul de date Matlab ce formează baza laboratorului de pe pagina cursului:

http://busoniu.net/teaching/sysid2019/index_ro.html

Răspundeți următoarelor cerințe:

- Reprezentați grafic datele de identificare, pentru a vă forma o idee despre forma funcției.
- Creați un aproximator polinomial de gradul $n - 1$, unde n este numărul de parametri sau funcții de bază. Valoarea lui n trebuie să fie ajustabilă în cod. De notat că există un parametru în plus pentru termenul constant din polinom, și de aceea gradul este doar $n - 1$. De exemplu, pentru $n = 4$ polinomul este de gradul 3 și expresia aproximatorului este dată de:

$$\hat{g}(x) = \theta_1 + x\theta_2 + x^2\theta_3 + x^3\theta_4$$

- Pentru orice valoare a lui n , construiți un sistem de ecuații liniare pentru regresia liniară, folosind datele de identificare. Utilizați reprezentarea matriceală explicată la curs. Rezolvați acest sistem

folosind operatorul Matlab de împărțire matriceală la stânga, “\” (sau, ca o alternativă, funcția `linsolve`). Calculați eroarea medie pătratică pe datele de identificare.

- Validați modelul obținut pe setul diferit de date de validare: calculați valorile approximate și din acestea eroarea medie pătratică. Reprezentați grafic funcția aproximată pentru setul de intrări de validare.
- Reglați numărul n de funcții de bază pentru performanțe cât mai bune, încercând valori de exemplu până la 20. Performanța va fi tot timpul evaluată folosind MSE pe setul diferit de validare, pentru a evita supraantrenarea. Creați un grafic al valorilor MSE în funcție de n , și găsiți punctul în care MSE este minimală.

Graficele pe care le obțineți vor fi similare celor exemplificate în figura de mai jos (evident, forma funcției și valoarea erorii pot fi diferite pentru setul dvs. de date).

