

Identificarea sistemelor – Laborator 3

Regresia liniară pentru aproximarea funcțiilor

Organizare

- **Dezvoltare:** Acest laborator se rezolvă independent de către fiecare student, în platforma Matlab Grader. Fiecare student a primit o invitație pentru această platformă pe adresa de email comunicată. Soluția constă din cod Matlab, dezvoltat într-o singură funcție. Puteți pretesta soluția de oricâte ori doriți, ca să verificați dacă funcționează.
- **Trimitere:** După ce sunteți satisfăcuți cu soluția, aceasta trebuie trimisă (Submit) în Matlab Grader. Nu o trimiteți decât odată; în caz de erori, aveți o a doua șansă de a trimite, dar această opțiune este doar de rezervă.
- **Verificare:** Codul va fi verificat atât de Grader automat, cât și de către profesor, inclusiv un test antiplagiat automat unde soluția dvs. este comparată cu cea a tuturor colegilor. Prezența la laborator va fi validată doar dacă aveți o soluție originală și funcțională. 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 duce la pierderea eligibilității.
- **Alte observații:** Discutarea ideilor între studenți este permisă și chiar de dorit, dar copierea sau schimbul direct de cod sunt interzise. Încălcarea acestei reguli va duce la consecințele descrise în regulile disciplinei.

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`.

Veți dezvolta o funcție cu următoarea semnătură:

```
[index, theta5, mse, yhatstar] = polyreg
```

Fiecărui student `i` se alocă de către profesor un `index` în intervalul 1-16, și acesta trebuie salvat în variabila `index` la începutul funcției. Indexul dictează care fișier de date trebuie încărcat. De exemplu, dacă aveți indexul 10, trebuie să încărcăți fișierul `lab4_10.mat`. Toate aceste fișiere de date sunt deja accesibile din codul funcției dvs, ele au fost încărcate în directorul problemei din Grader (chiar dacă nu sunt vizibile explicit).

Răspundeți următoarelor cerințe:

- Reprezentați grafic datele de identificare și/sau validare, 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:

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

Ordinea termenilor polinomiali este importantă, păstrați-i în ordinea crescătoare a puterilor lui x .

- 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.
- Salvați parametrii obținuți pentru $n = 5$ în `theta5` și folosind `Pretest`, verificați dacă sunt corecți (un “sanity check” pentru cod).
- 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. 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).
- Acordați numărul n de funcții de bază pentru performanțe cât mai bune, încercând valori de la 2 până la 20. Performanța va fi tot timpul evaluată folosind MSE pe setul diferit de validare, pentru a evita supraantrenarea.
- Pentru fiecare valoare a lui n , salvați MSE în vectorul `mse`, astfel încât `mse(1)` să fie eroarea pentru $n = 2$, `mse(2)` pentru $n = 3$ și așa mai departe.
- Creați un grafic al valorilor MSE în funcție de n , și găsiți punctul în care MSE este minimală. Salvați în `yhatstar` ieșirea pentru valoarea optimă a lui n .

