

# Identificarea sistemelor – Laborator 4

## Regresia liniară pentru aproximarea funcțiilor

### Organizare

- Acest laborator este parte obligatorie a cursului de Identificarea Sistemelor. Laboratorul se rezolvă independent de către fiecare student.
- Soluția constă din cod Matlab. Dezvoltați acest cod într-un singur script Matlab. Codul va fi verificat și rulat de către profesor pentru a vă lua în considerare prezența la laborator. Vom efectua această verificare pe cât posibil în timpul laboratorului, împreună cu dvs. Scrieți în orice caz codul de o manieră clară, adăugând comentarii acolo unde este necesar, pentru a-l face inteligibil și în lipsa explicațiilor verbale. La sfârșitul laboratorului trimiteți codul profesorului prin email (Zoltán Nagy la [zoltan.nagy@aut.utcluj.ro](mailto:zoltan.nagy@aut.utcluj.ro), sau Marius Costandin la [marius.costandin@aut.utcluj.ro](mailto:marius.costandin@aut.utcluj.ro)) sub forma unui fișier .m sau într-o arhivă ZIP, folosind următorul format pentru numele de fișier:  
`is_labN_indexINDEX_NUME`  
unde N este numărul laboratorului, INDEX este indexul setului de date, vezi mai jos; și NUME este numele dvs. de familie. Vă rugăm să *includeți acest nume de fișier și în subiectul emailului*.
- 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 modele liniare, vezi secțiunea de *Regresie liniară* din Partea 3 a suportului de curs, *Baze matematice*. Vom lucra cu aceleași seturi de date ca și în prima parte a proiectului.

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 două variabile de intrare și 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 următoarele câmpuri:

- O colecție de coordonate  $X$  pentru intrări, unde  $X$  este un *cell array* conținând doi vectori, fiecare vector  $X\{\text{dim}\}$  conținând o grilă de  $n$  puncte pentru dimensiunea  $\text{dim}$  a intrării.
- O colecție de ieșiri corespunzătoare  $Y$ , reprezentată printr-o matrice de dimensiunea  $n \times n$ , unde  $Y(i, j)$  este valoarea funcției  $f$  în punctul  $(X\{1\}(i), X\{2\}(j))$  (afectată de zgomot).
- Aceleași date sunt furnizate și într-un format alternativ “plat”: intrările `Xflat`, o matrice lată conținând cele  $n^2$  puncte de intrare, câte unul pe fiecare coloană, și ieșirile corespunzătoare într-un vector linie `Yflat`.

Variantele “structurate”  $X, Y$  sunt mai ușor de folosit pentru crearea graficelor, iar cele “plate” sunt mai potrivite pentru antrenarea și folosirea aproximatorului.

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

[http://busoniu.net/teaching/sysid2017/index\\_ro.html](http://busoniu.net/teaching/sysid2017/index_ro.html)

Cum seturile de date sunt comune cu prima parte a proiectului, este de preferat să folosiți același index (același fișier de date) ca și la proiect.

Pe lângă fișierele de date, website-ul furnizează și o funcție Matlab `rbfapprox` care implementează un aproximator cu funcții radiale de bază (RBFuri), și un script exemplu `lab4_template` de la care puteți porni în rezolvarea laboratorului. Acest script exemplifică în detaliu inițializarea și utilizarea aproximatorului cu RBFuri. În rezumat, se apelează `rbfapprox` pentru a crea o grilă de RBFuri echidistante, cu argumentele: limitele domeniului intrărilor funcției, și numărul de RBFuri pe fiecare dimensiune a acestui domeniu. Funcția returnează o structură conținând două funcții: `phi`, care calculează vectorul de regresori pentru orice valoare a lui  $x$ ; și `eval`, care evaluează ieșirea aproximatorului  $\hat{y}$  în orice  $x$ , dat fiind vectorul de parametri  $\theta$ .

Răspundeți următoarelor cerințe:

- Reprezentați grafic datele de antrenare (identificare), pentru a vă forma o idee despre forma funcției (folosiți `mesh`).
- 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 (sau funcția obiectiv  $V(\hat{\theta})$ , care este doar suma erorilor pătratice  $\frac{1}{2} \sum_k \varepsilon(k)^2$ ).
- Validați modelul obținut pe setul diferit de date de validare: calculați valorile approximate și din acestea eroarea medie pătratică (sau funcția obiectiv  $V(\hat{\theta})$ ). Reprezentați grafic funcția aproximată pentru setul de intrări de validare, folosind `mesh`.
- Încercați câteva valori pentru numărul de funcții de bază pe fiecare dimensiune, investigând influența acestui număr asupra calității aproximării pe datele de identificare și de validare.

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

