

Dezvoltarea sistemelor care învață singure - probleme de clasificare**Obiective**

Dezvoltarea sistemelor care învață singure. Algoritmi de învățare supervizată. Modele liniare. Specificarea, proiectarea și implementarea sistemelor care învață singure cum să rezolve probleme de clasificare.

**Aspecte teoretice**

Proiectarea și dezvoltarea sistemelor care învață singure.

Algoritmi de învățare bazați pe:

- metoda gradientului descendent
- algoritmi evolutivi

**Termen de predare**

Laborator 5

**Cerințe**

Specificați, proiectați și implementați o aplicație pentru rezolvarea uneia dintre următoarele problemele de clasificare (binară sau multiclasă) liniară ($Y = b_0 + b_1 * x_1 + b_2 * x_2 + \dots + b_n * x_n$) cu ajutorul unui sistem care învață singur să rezolve problema. Sistemul trebuie să permită alegerea unuia dintre cei 2 algoritmi de învățare: metoda **regresiei logistice** și **algoritmi evolutivi**.

Aplicația trebuie să permită:

- Încărcarea datelor problemei (probleme cu date deja definite de către programator, probleme cu date definite de utilizator)
- Alegerea și parametrizarea metodei de rezolvare a problemei
- Prezentarea rezultatelor procesului de învățare (antrenare și testare)

Aplicația trebuie să respecte specificațiile privind datele de intrare și datele de ieșire.

Se dau două fișiere de intrare și trebuie furnizat un fișier de ieșire.

Cele două fișiere de intrare conțin datele de antrenament și, respectiv, datele de testare în formatul:

- prima linie: numărul de variabile independente (n) + 1
- a doua linie: numărul de exemple (N)
- următoarele linii conțin o matrice cu $n + 1$ coloane și N linii. Fiecare linie corespunde unui exemplu format din $n + 1$ valori: n valori ale lui x (x_1, x_2, \dots, x_n) și valoarea lui y

Fisierul de ieșire va conține 2 linii:

- prima linie va conține acuratetea pe datele de test (obținută prin metoda regresiei logistice)
- a doua linie va conține acuratetea pe datele de test (obținută prin metoda algoritmilor evolutivi)

Aplicația va fi testată folosind date de dificultăți diferite (fiecare test validat având asociat un anumit punctaj).

De asemenea, pentru fiecare problemă există exemple de date de test.

Codul aferent aplicației poate fi dezvoltat 100% de către student sau poate fi bazat pe exploatarea unor biblioteci specifice.

	Code 100%					Code&tools					Min for valid
Alg	Code	Easy Test	Medium Test	Hard Test	Total	Code	Easy Test	Medium Test	Hard Test	Total	
LogRegr	200	100	160	200	660	100	20	100	160	380	220
GA	200	100	160	200	660	100	20	100	160	380	220
Total	400	200	320	400	1320	200	40	200	320	760	

Bonus

Sisteme disponibile care implementează algoritmi de învățare:

1. Weka <http://www.cs.waikato.ac.nz/ml/weka/>
2. Matlab <http://www.mathworks.com/products/neural-network/>
3. OpenCV http://docs.opencv.org/modules/ml/doc/neural_networks.html
4. Scikit-learn <http://scikit-learn.org/stable/>
5. GPLAB <http://gplab.sourceforge.net/>
6. ECJ <http://cs.gmu.edu/~eclab/projects/ecj/>

1. Să se clasifice o **cardiotocogramă fetală** ca fiind normală, suspectă sau patologică pe baza diferitelor măsurători înregistrate în decursul sarcinii. Detalii despre aceasta problema puteti gasi aici <http://archive.ics.uci.edu/ml/datasets/Cardiotocography>
2. Să se stabilească **direcția de deplasare a unui robot** (în față, viraj ușor la dreapta, viraj puternic la dreapta, viraj ușor la stânga) dotat cu senzori pe baza informațiilor colectate anterior (sub forma unor citiri a 24 de senzori de ultrasunete plasați în diferite poziții pe robot – poziția unui senzor fiind dată de unghiul de deviație - 180° (față), -165°, -150°, ..., 15°, 0° (spate), 15°, 30°, ..., 150°, 165°). Detalii despre aceasta problema puteti gasi aici <http://archive.ics.uci.edu/ml/datasets/Wall-Following+Robot+Navigation+Data>
3. Să se stabilească **tipul de defecțiune** (sunt posibile 11 defecțiuni) a unui motor electric pe baza unor caracteristici de funcționare pe care le are acest motor în diferite condiții. Detalii despre această problemă puteți găsi aici <http://archive.ics.uci.edu/ml/datasets/Dataset+for+Sensorless+Drive+Diagnosis>

