

Rezolvarea unor probleme prin metode de învățare automată

Obiective



Dezvoltarea sistemelor care învață singure. Probleme de tip clasificare rezolvate cu metoda regresiei logistice. Evaluarea performanței acestor metode.

Aspecte teoretice



Metoda regresiei logistice pentru rezolvarea problemelor de clasificare.
Proiectarea sistemelor care învață singure.
Evaluarea sistemelor care învață singure. Metrici de performanță.

Termen de predare și evaluare



Laborator 10; Punctajele acordate:

- Rezolvarea problemei cu tool – 100 puncte
- Rezolvarea problemei cu cod propriu – 150 puncte + 50 puncte (daca acuratetea clasificarii > 90%)
- Rezolvarea cerințelor opționale – 200 puncte

Cerințe



Specificați, proiectați și implementați rutine de rezolvare a unei probleme de clasificare folosind metoda bazată pe regresie logică.

Exemplu live: clasificarea țesuturilor cancerigene

Se consideră informații despre cancerul de sân la femei, informații extrase din ecografii mamare (detalii la [https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+\(Diagnostic\)](https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic))) precum:

- Tipul malformației identificate (țesut *benign* sau țesut *malign*)
- Caracteristici numerice ale nucleului celulelor din aceste țesuturi:
 - o *raza* (media distanțelor între centru și punctele de pe contur)
 - o *textura* (măsurată prin deviația standard a nivelelor de gri din imaginea asociată țesutului analizat)

Folosindu-se aceste date, să se decidă dacă țesutul dintr-o nouă ecografie (pentru care se cunosc cele 2 caracteristici numerice – *raza* și *textura* –) va fi etichetat ca fiind *malign* sau *benign*.

Tema: Ce fel de floare preferi?

Se consideră problema clasificării florilor de iris în diferite specii precum: *setosa*, *versicolor* și *virginica*. Pentru fiecare floare se cunosc caracteristici precum: lungimea și lățimea sepalei, lungimea și lățimea petalei. Mai multe detalii despre acest set se pot găsi la <https://archive.ics.uci.edu/ml/datasets/Iris>.

Folosindu-se aceste informații, să se decidă din ce specie aparține o anumită floare.

Să se rezolve problema, implementându-se rutine pentru:

- clasificarea multi-clasă folosind metoda de regresie logică (cu tool)
- clasificarea multi-clasă folosind metoda de regresie logică (cu cod propriu)
- folosirea batch-urilor în procesul de antrenament și validarea încrucișată (**temă opțională**)
- investigarea diferitelor funcții de loss (**temă opțională**).

Temă opțională: Variație prag de decizie (threshold = θ)

Ce se întâmplă în cazul clasificării binare dacă se modifică pragul de decizie din 0.5 în alte valori. Cum se poate aprecia calitatea clasificatorului pentru diferite valori ale pragului?