

Rezolvarea unor probleme prin metode de învățare automată



Obiective

Dezvoltarea sistemelor care învață singure. Probleme de tip regresie rezolvate cu metoda gradientului descrescător. Evaluarea performanței acestor metode.



Aspecte teoretice

Metoda gradientului descrescător pentru rezolvarea problemelor de regresie.
Proiectarea sistemelor care învață singure.
Evaluarea sistemelor care învață singure. Metrici de performanță.



Termen de predare și evaluare

Laborator 9

Punctajele acordate:

- Rezolvarea problemei cu tool – 50 puncte
- Rezolvarea problemei cu cod propriu, cazul regresiei univariate – 150 puncte
- Rezolvarea problemei cu cod propriu, cazul regresiei multi-variate – 50 puncte
- Normalizarea datelor – cod propriu 75 puncte
- Rezolvarea cerințelor opționale – maxim 300 puncte



Cerințe

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

Problemă: Ce îi poate face pe oameni fericiți?

Se consideră problema predicției gradului de fericire a populației globului folosind informații despre diferite caracteristici a bunăstării respectivei populații precum *Produsul intern brut* al țării în care locuiesc (gross domestic product – GBP), *gradul de fericire*, etc.

Folosind datele aferente anului 2017 [link](#), să se realizeze o predicție a gradului de fericire în funcție:

- doar de Produsul intern brut
- de Produsul intern brut și de gradul de libertate.

Să se rezolve problema, implementându-se rutine pentru:

- gradient descrescător stocastic (**exemplu detaliat live**)
- gradient descrescător bazat pe batch-uri – cu tool sau cu cod propriu – (**temă**)
- normalizarea datelor de antrenament și test (**temă**)

Temă opțională - rezolvarea unei probleme de regresie prin:

- implementarea unui algoritm GD non-liniar
- implementare unui algoritm de programare genetică (setul de terminale este reprezentat de feature-uri, iar setul de funcții de operatori de bază (+, -, *), putând fi extins și cu alți operatori (/ , sin, cos, exp, log, etc.))
- implementare regresie multi-target (cu mai multe output-uri) – sugestii:

- outputurile sa fie independente (de ex pe setul de date din `sklearn.datasets` pot folosi datele `psyho` din `linnerud`)
- outputurile sa fie dependente (aici s-ar putea folosi un regressor gata antrenat – gen `yolo` (<https://pjreddie.com/darknet/yolo/>) – pentru a prezice coordonatele bounding box-urilor care încadrează obiectele recunoscute în imagini; trebuie studiat cum se evaluează dacă acele BBs sunt bune sau nu; focusul este de fapt pe interpretarea outputului dat de regressor, nu pe modul în care se antrenează regressorul)