

Inteligență artificială - laborator

Prelucrarea datelor

1. Încărcați fișierul de date **Iris** de la <https://archive.ics.uci.edu/ml/datasets/iris>.
2. Pentru fiecare coloană cu valori numerice calculați valoarea minimă, maximă, medie, mediană.
3. Normalizați valorile de pe fiecare coloană folosind formula $x_{norm} = \frac{x-min}{max-min}$.
4. Calculați suma ponderată a valorilor numerice de pe fiecare linie înmulțind fiecare valoare cu un coeficient pe care îl numim pondere și salvați rezultatele într-o nouă coloană a tabelului vostru. Folosiți următoarele valori pentru ponderi: [0.1, 0.8, -1.1, 0.6].
5. Scalarea unei imagini digitale presupune redimensionarea acesteia prin mărirea sau reducerea numărului de pixeli. Setul de date disponibil la <http://archive.ics.uci.edu/ml/machine-learning-databases/optdigits/optdigits-orig.tar.Z> conține imagini ale unor cifre scrise de mână reprezentate sub forma de matrice binare de dimensiune 32x32. Scrieți o metodă de scalare a unei astfel de imagini prin reducerea dimensiunii la 16x16 pixeli. Folosiți o fereastră glisantă de dimensiune 2x2 pe care o deplasați peste imaginea inițială. Dacă fereastra conține 3 sau 4 valori de 0, aceasta se transformă într-o valoare de 0 în imaginea scalată, iar altfel vom avea o valoare de 1. Aplicați această metodă unei cifre din fișier și evaluați vizual calitatea imaginii scalate. Opțional: Aplicați din nou scalarea cu masca de 2x2 pentru reducerea dimensiunii imaginii de la 16x16 la 8x8 pixeli și observați dacă cifrele se mai pot distinge.

Exemplu de citire a valorilor din setul de date Iris

Pentru citirea valorilor numerice din fișierul iris.data și încărcarea într-o listă, puteți folosi codul de mai jos:

```
import csv
myList = []

with open('c:\Users\User\Desktop\iris.data') as csvfile:
    readCSV = csv.reader(csvfile, delimiter=',')
    for row in readCSV:
        if len(row) == 5:
            myList.append([float(row[0]), float(row[1]), float(row[2]), float(row[3])])
```

Puteți folosi numpy pentru operații cu vectori. Converteți mai întâi myList în array de tip numpy și folosiți apoi metodele din acest pachet asupra coloanelor din array.

```
import numpy as np
npArray = np.array(myList)

col0_min = np.min(npArray[:,0])
col0_norm = (npArray[:,0]-min(npArray[:,0])) / (max(npArray[:,0])-min(npArray[:,0]))
```