

Test de laborator la Inteligență Artificială, seria 36

- 2 februarie 2023 -

1. Clasificarea cifrelor folosind imagini prototip (4.5 puncte)

În cele ce urmează vom considera problema clasificării cifrelor scrise de mână (0, 1, 2, ..., 9) folosind imagini prototip. Codul Python de la care porniți încarcă o mulțime de 1000 de exemple de cifre și plotează primele astfel 100 de exemple afișând și etichetele lor.

Realizați următoarele:

- Împărțiți cele 1000 de exemple de cifre astfel: primele 700 de exemple intră în mulțimea de antrenare, următoarele 200 de exemple intră în mulțimea de validare, ultimele 100 de exemple intră în mulțimea de testare. **(0.5 puncte)**
- Calculați și plotați pentru fiecare cifră în parte (0, 1, 2, ..., 9) *exemplul prototip* al cifrei pe baza mulțimii de antrenare făcând media tuturor exemplelor de antrenare care conțin acea cifră. Astfel, pentru a calcula exemplul prototip al cifrei 0 trebuie să faceți media tuturor exemplelor cu cifra 0 din mulțimea de antrenare. Exemplele prototip obținute ar trebuie să semene cu cele din Figura 1 de mai jos. **(1.5 puncte)**

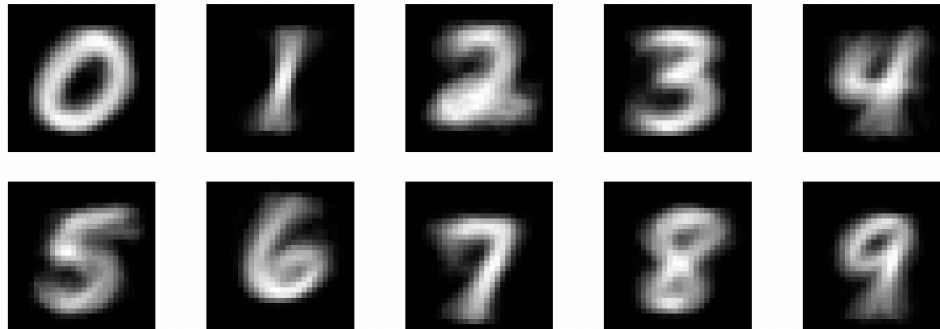


Figura 1. Exemplele prototip pentru cifrele 0-9

Dacă nu știți să rezolvați punctul b puteți considera ca exemplu prototip pentru fiecare cifră primul exemplu de antrenare conținând acea cifră. Veți folosi exemplele prototip la punctele următoare.

- Veți folosi cele 10 exemple prototip de la punctul anterior pentru construirea unui nou clasificator. Astfel, pentru un exemplu X ce urmează a fi clasificat, se calculează distanța dintre exemplul X și fiecare exemplu prototip, iar eticheta lui X este stabilită a fi eticheta exemplului prototip cel mai apropiat (în sensul distanței celei mai mici). Construiți acest clasificator și stabiliți acuratețea lui pe cele 200 de exemple de validare folosind pe rând distanța Euclidiană și distanța Manhattan. Alegeți distanța pentru care obțineți cea mai bună acuratețe pe mulțimea de validare. Afișați această acuratețe și distanța aleasă. **(1.5 puncte)**
- Folosiți clasificatorul de la punctul c) cu distanța găsită (Manhattan sau Euclidiană) și clasificați cele 100 de exemple din mulțimea de testare. Afișați matricea de confuzie și acuratețea. **(1 punct)**