**Raport proiect ISIA – Parkinson**

Setul de date pentru aceasta problema este compusa dintr-o serie de masuratori biomedicale ale vocii a 31 de persoane, dintre care 23 au Parkinson. Fiecare coloana in setul de date reprezinta o masuratoare particulara a vocii, iar fiecare linie reprezinta una dintre cele 195 de inregistrari vocale de la oamenii care au fost testati. Scopul acestui experiment este diferentierea persoanelor sanatoase (0 in coloana status), fata de cele cu Parkinson (1 in aceasi coloana). Pentru fiecare pacient, de fiecare data cand au fost supusi experimentu-lui, s-au realizat 23 de masuratori. De exemplu, pentru pacientul “phon\_R01\_S01\_1”, diagnosticat cu Parkinson s-au masurat frecventa vocala maxima, minima si medie a pacientului, 5 masuratori ale variatiei vocii in frecventa fundamentala, 6 masuratori ala amplitudinii vocii, 2 masuratori ale indicelui din tonalitatii vocii, statusul (sanatos – 0, bolnav cu PD - 1), 2 masuratori nonliniare dinamice ale complexitatii vocii, exponentul fractal si 3 masuratori nonliniare ale variatiei frecventei fundamentale.

Problema de rezolvat a fost o problme de Perceptron Multistrat (MLP) de clasificare deoarece setului de date i se poate atribuii etichete discrete (0 persoanelor sanatoase si 1 pentru cele cu Parkinson in coloana status din “parkinson.csv”). In codul sursa, am utilizat urmtoarele biblioteci: numpy (generare matrice si vectori), sklearn( reteaua neuronal) si libraria pandas( citirea si prelucrarea datelor din setul de date “parkinson.csv”).

Text

Description automatically generated Pentru inceput, am realizat cateva modificari data setului. Am schimbat extensia fisierului din .data, in .csv si am sters coloana “nume” care nu reprezinta interes pentru problema noastra. A urmat importarea setului de date in codul sursa prin intermdiul librariei pandas. Am citit toate datele din setul de date in matricea intitulata “date\_parkinson”. A urmat convertirea acesteia in matricea propiu-zisa prin intermediul functiei to\_numpy(). Din aceasta noua matrice a urmat separarea datelor de etichete si convertirea la tipul lor de date (date – float, etichete – int ). In continuare am definit functia de calculare a acuratetii care foloseste un contor care creste de fiecare data cand etichetele de test si valorile vectorului de predictii sunt egale, urmand ca acest contor sa fie raportat la lungimea vectorului de etichte test. In umtorul pas, am impartit datele si etichetele in train si test. Am ales ca 75% din date si din etichtete sa fie folosite pentru antrenare si restul pentru testare. Consider ca aceste proportii sunt alese corect pentru acesta problema deoarece, din natura problemei care detecteaza daca cineva are PD, avem nevoie de o performanta, lucru care se obtine printr-un numar mai mare al datelor si etichetelor de antrenare. Pentru algoritmul MLP, am variat toate combinatiile posibile. Am folosit un vector de learning-rates de 0.1 si 0.01, iar pentru numarul de neuroni din stratul/straturile ascunse, am folosit dimensiunile de 22, respectiv jumatate 11 (cele 24 de atribute ale problemei minus atributul nume(nefolositor problemei) si atributul status(destinat etichetelor)). Pentru fiecare cobinatie, am obtinut 2 predictii corespunzatoare celor 2 learning-rates, carora le-am calculat si predictia medie. Pentru aceasta problema, am fost nevoit sa modific numarul maxim de iteratii al functiei MLPClassifier de la numarul default de 200 la 1200. Acest lucru a fost necesar pentru a asigura o convergenta totala (in cazul solverului “adam”, care foloseste algoritm de tip SGD, numarul de iteratii maxim reprezinta numarul epocilor, adica de cate ori algoritmul de invatare va functiona prin intregul set de antrenamet).



Text

Description automatically generatedIn urma rularii programului, am obtinut, de exemplu, urmatoarele predictii:

Ca o nota personala, consider ca Inteligenta Artificiala are importante aplicatii, mai ales in domenii ca medicina unde poate salva vietii. Problema de fata a constituit din construirea unui sistem MLP de clasificare care sa prezica daca o persoana este sau nu suferind de Parkinson. Din punct de vedere uman, detectarea acestor tipuri de afectiuni ar necesita timp, astfel ca predictia acestor sisteme inteligente potate salva acest timp, si implicit viata, iar un tratament poate fi inceput in acest timp.

Bibliografie:

<https://python-course.eu/machine-learning/neural-networks-with-scikit.php>

<https://scikit-learn.org/stable/modules/classes.html#module-sklearn.model_selection>

<https://www.techbeamers.com/read-write-file-in-python/>