Specyfikacja programu "Jednowarstwowa siec neuronowa":

Dane wejściowe (należy samemu przygotować):

1. Stworzyć zadany katalog na dane.

2. W zadanym katalogu tworzymy kilka (K=3) podkatalogów

nazwanych nazwami jezykow (np. czeski, slowacki ...)

3. W kazdym z nich umieszczamy po kilka tekstow

trenujacych sciagnietych np. z wikipedii w odpowiednich jezykach (w alfabetach łacińskich).

(Z praktyki wynika, ze wystarcza nawet mniej 10 tekstów uczących dla każdego języka, ale dlugosci chociaz ze

2 akapity)

4. W momencie uruchomienia sieć perceptronów będzie używała tych tekstów jako dane trenujące.

Opis programu:

Uzyjemy 1-warstwowej sieci neuronowej do klasyfikowania jezykow

naturalnych tekstow.

Bierzemy dokument w dowolnym jezyku (w alfabecie lacinskim) z pliku ".txt",

wyrzucamy wszystkie znaki poza literami alfabetu angielskiego (ascii) i

przerabiamy na 26-elementowy wektor proporcji liter (czyli: jaka jest

proporcja 'a', 'b', etc.)

Okazuje sie, ze taki wektor rozkladu znakow wystarcza do rozrozniania

jezyka naturalnego dokumentu tekstowego, nawet dla tak podobnych jezykow

jak np. czeski i slowacki.

Tworzymy wiec 1 warstwe K perecptronow (gdzie K to liczba jezykow, wezmy ze K=3 jezyki) i uczymy kazdego perceptrona rozpoznawania "jego" jezyk.

Uczenie perceptronów przepiega jak w poprzednim projekcie, tzn z dyskretną (0-1) funkcją aktywacji.

Mając wyuczony każdy z perceptronów, klasyfikacji do jednej z K klas dokonujemy uzywając maximum selector (zdjac dyskretna funkcje aktywacji) i normalizowac zarowno wektor wag jak i wejsc.

UWAGA: przy normalizacji można użyć miary euklidesowej.

Normalizując wektor wag nie dokładamy do niego parametru progu (theta).

Zapewnic okienko tekstowe do testowania: po nauczeniu wklejamy dowolny nowy tekst w danym jezyku i sprawdzamy, czy siec prawidlowo go klasyfikuje.

Oczywiscie w momencie pisania programu nie powinno byc wiadome ile i jakie beda jezyki.

Nie mozna uzywac zadnych bibliotek ML, wszystko ma byc zaimplementowane od zera w petlach, ifach, odleglosc tez sam ma liczyc uzywajac dzialan arytmetycznych (do operacji na liczbach mozna uzywac java.lang.Math), etc.

Można używać java.util.