Autor: Maksymilian Romańczuk GD41582

Data: 04.11.19

Przedmiot: Sztuczna Inteligencja

Numer zadania: 1

Cel: Klasyfikacja płci za pomocą sieci neuronowych, zadanie polega na sprawdzeniu, jaką płeć miała osoba będąca w domu starców w Palo Atlot w Kalifornii w latach 1964-1975.

Opis:

Metoda oceniania: hold-out, lambda 2/3

Zbiór danych: boot, channing

channing: 5 kolumny, 462 wierszy

[,1] sex factor płeć osoby

[,2] entry numeric wiek osoby gdy przychodziła do domu starców(w miesiącach)

[,3] exit numeric wiek osoby gdy odchodziła z domu starców(w miesiącach)

[,4] time czas spędzony w domu starców(w miesiącach)

[,5] cens numeric wartości 1 gdy osoba zmarła w ośrodku, 0 gdy osoba odeszła z domu starców bądź wciąż w nim żyje

Kod z komentarzem:

//instalacja pakietu amore i biblioteki boot

install.packeges("AMORE")

library(AMORE)

library(boot)

//generujemy dane uczące sieć

data("channing")

channing

table(channing$cens)

l.danych=nrow(channing)

set.seed(8)

//2/3 losowych indexów do trenowania i pozostałe 1/3 do testowania

idxTren<-sample(1:l.danych,2\*l.danych/3)

idxTest<-setdiff(1:l.danych,idxTren)

//definiujemy funkcję zmieniającą zmienną zawierającą nazwę etykiety

//na liczbę zmiennych binarnych, ile jest wartości etykiet

target<-function(x)

{

n<-length(x)

wartosci<-levels(x)

l<-length(wartosci)

T<-matrix(0,nrow=n,ncol=l)

for(i in 1:l)

T[,i]<-(x==wartosci[i])

colnames(T)<-wartosci

return(T)

}

//zastosowanie funkcji dla danych określających etykiety

wZadane<-target(channing$sex)

wZadane

set.seed(8)

//tworzymy strukturę sieci

siec<-newff(n.neurons=c(6,6,2),

learning.rate.global=0.03,

momentum.global=0.5,

hidden.layer="sigmoid",

output.layer="purelin",

method="ADAPTgdwm",

error.criterium="LMS")

//trenujemy sieć

wynik<-train(siec,

channing[idxTren,c(2:5)],

wZadane[idxTren,],

error.criterium="LMS",

report=TRUE,

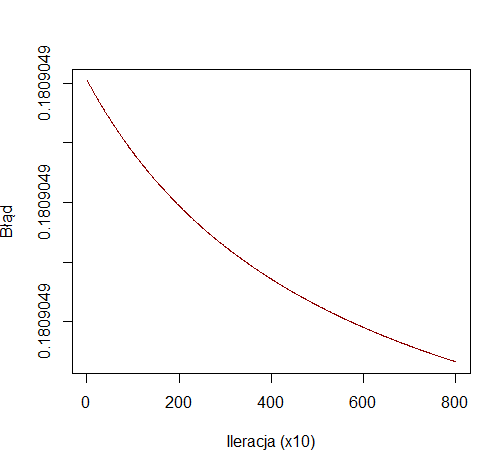
show.step=10,

n.shows=800)

//wyświetlam wartość błędów

plot(wynik$Merror,type="l",xlab="Ileracja (x10)",

ylab="Błąd", col="darkred")



//stosujemy wytrenowaną sieć do danych testowych

y<-sim(wynik$net,channing[idxTest, c(2:5)])

y

//definiuję funkcję oceny klasyfikacji (zamieniam liczby na etykietę)

test.klasyf<-function(zad,wy)

{

zadane<-max.col(zad)

rozpoznane<-max.col(wy)

print(table(zadane,rozpoznane))

}

wynik<-test.klasyf(wZadane[idxTest,],y)

//określam dokłądność klasyfikacji

cat("Dokładność klasyfikacji:",

sum(diag(wynik))/sum(wynik)\*100, "%\n")

Dokładność klasyfikacji: 80%

Podsumowanie: Klasyfikacja płci w domu starców na poziomie 80% oznacza, że sieć jest w stanie określić płeć osoby przy użyciu danych z bazy. Może nie jest to wyniki, 90+ ale jest to wynik wciąż pozwalający na dobrą klasyfikację płci.