

ELEKTROTEHNIČKI FAKULTET
UNIVERZITETA U BEOGRADU

Određivanje kvaliteta vina pomoću neuralne mreže

Projektni zadatak iz predmeta neuralne mreže

Studenti:

Dejan Golubović 3228/2016

Stefan Vranković 3379/2017

Profesor:

Doc. dr Goran Kvašček

Beograd, januar 2018

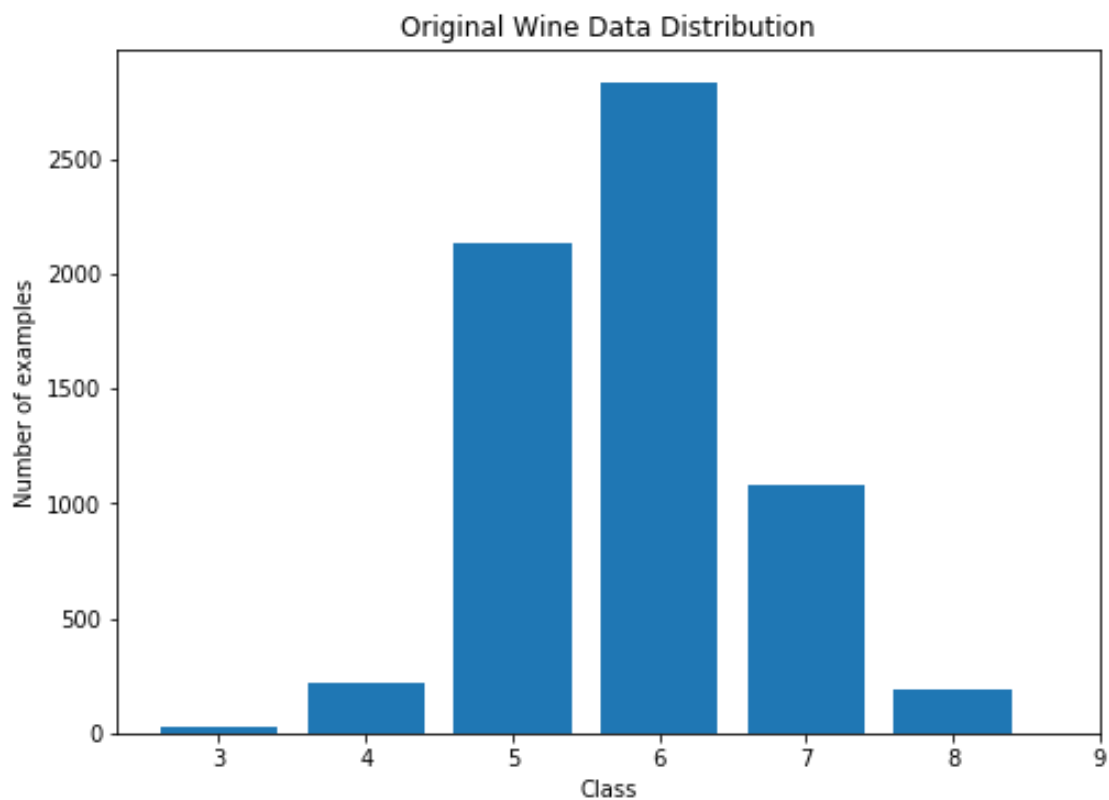
Uvod

Cilj rada

Cilj projekta je formirati i obučiti (višeslojnu) feedforward neuralnu mrežu koja će klasifikovati vina na osnovu hemijskih parametara. Vino se klasifikuje kao dobro ili loše.

Podaci

Hemijske karakteristike vina su: fiksna kiselost, promenljiva kiselost, limunska kiselina, rezidualni šećer, hloridi, slobodni sumpor dioksid, ukupni sumpor dioksid, gustina, pH vrednost, sulfati i alkohol. Rezultat je ocena od 1 do 10.



Slika 1 - Raspodela originalnih podataka

Implementacija

Korišćene tehnologije

Za rešavanje problema korišćen je programski jezik Python. Korišćene su biblioteke scikit-learn i Tensorflow. Python je izabran zbog širokih mogućnosti za rad sa mašinskim učenjem. Tensorflow i scikit-learn su biblioteke za mašinsko učenje i njihove metode su korišćene u radu. Implementacija je u Jupyter Notebook okruženju.

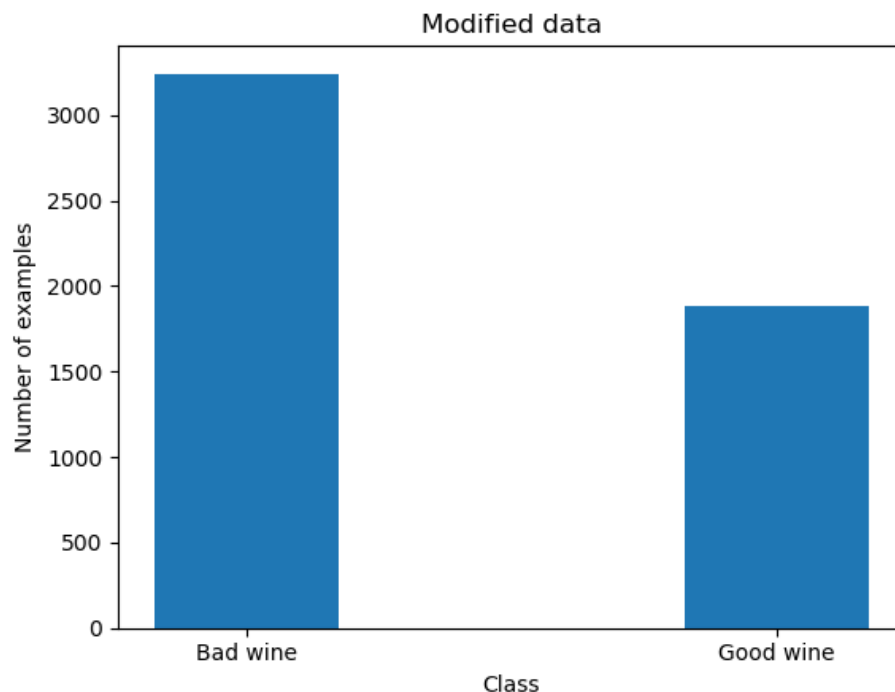
Način rada

Rad ima 3 glavna dela:

- 1) Učitavanje podataka
- 2) Implementacija mreže
- 3) Obučavanje i testiranje mreže

Učitavanje podataka

Podaci se prvo pročitaju iz fajla. Zatim se podele na matrice X i Y. Nakon toga se podaci dele na skupove za obučavanje i za testiranje. To se radi tako što se iz svake klase uzme 80% podataka za obučavanje, a 20% za testiranje. Zatim se podaci podele na dobra i loša vina, ocene 0-5 loša, 6-10 dobra. Na kraju se radi normalizacija X ulaznog trening i test skupa. Za normalizaciju se koristi scikit-learn biblioteka.



Slika 2 - Nova raspodela podataka

Implementacija mreže

Mreža se implementira tako da može da podrži promenljiv broj skrivenih slojeva i neurona. Ulazni sloj ima 11 neurona, prema ulaznim podacima. Izlazni sloj ima dva neurona zbog kodiranja izlaza. Koristi se gradijentno opadajući algoritam za obučavanje. Dve funkcije, `train()` za obučavanje i `test()` za testiranje i evaluaciju. Kao aktivaciona funkcija korišćen je `relu` za skrivene slojeve. Izlazni sloj koristi `softmax`. Mreža se implementira u Pythonu tako što se formira Tensorflow graf. Onda se taj graf izvršava. Kodiranje izlaza urađeno je tako da za svaku klasu postoji jedan izlazni neuron, ukupno dva. Svaki neuron ima vrednost između 0 i 1. Klasifikacija se vrši tako što se izabere maksimum.

Obučavanje i testiranje mreže

Nebalansirane klase dovode do problema gde je mreža pristrasnija brojnijoj klasi. To čak može dovesti do jako dobre tačnosti, međutim ta tačnost se ne može uzeti kao merodavna. Jedna klasa se klasifikuje odlično a druga loše. Stoga su korišćene druge metrike prilikom evaluacije.

	P	N
P	TP	FN
N	FP	TN

Tabela 1 - Konfuzionna matrica

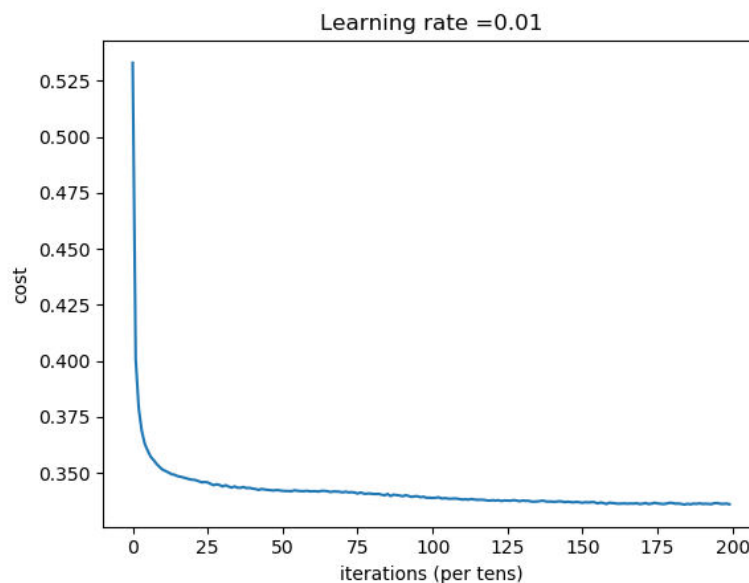
True Positive Rate - procenat tačnih predikcija kod pozitivne klase.

True Negative Rate - procenat tačnih predikcija kod negativne klase.

U radu se teži balansiranim performansama u ove dve metrike, kao i ukupnim balansiranim performansama na trening i test skupovima.

Algoritam za obučavanje je minibatch gradijentno opadajući. To znači da se u svakoj epohi uzima određen broj podataka za treniranje. Uzeto je da se u svakoj epohi uzima jednak broj podataka iz obe klase. Time se neutrališe pristrasnost mreže brojnijoj klasi.

Mreža koja je dala najbolje performanse sadrži tri skrivena sloja sa po 20 neurona. Kao zaštita od preobučavanja koristi se regularizacija. Drugi način bio je smanjenje kompleksnosti mreže. Regularizacija je dala bolje rezultate.

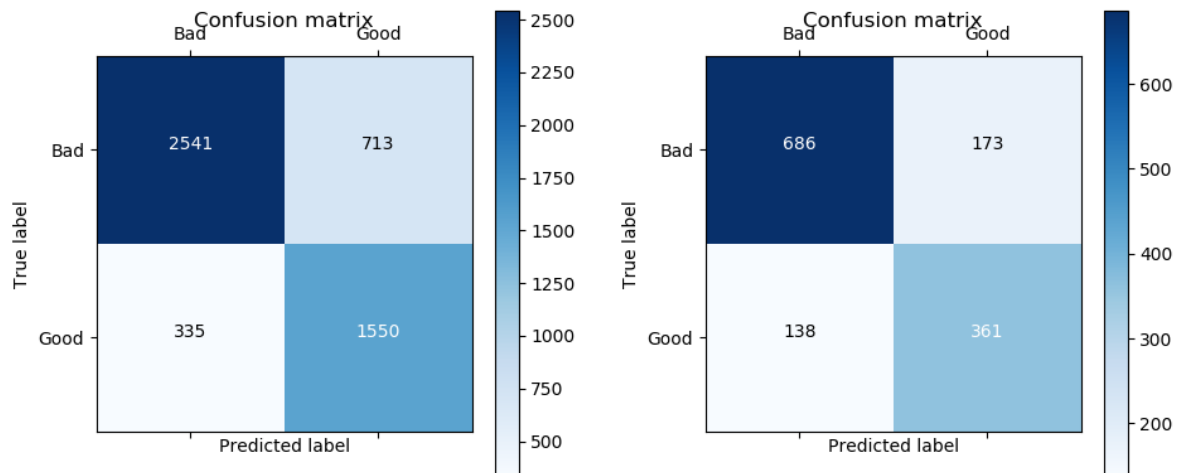


Slika 3 - Funkcija greške

Rezultati

	Trening	Test
Tačnost	79.6	77.1
TPR (Loše vino)	78.1	79.9
TNR (Dobro vino)	82.2	72.3

Tabela 2 - rezultati



Slika 4 - Konfuzione matrice na trening i test skupu

Zaključak

Dobijeni rezultati imaju dobru tačnost. Uočava se balansiranoost rezultata u metrikama od interesa. Takođe rezultati su slični na trening i test skupovima, nema preobučavanja. Stoga se smatra da je glavni cilj rada postignut. Mreža uspešno klasifikuje vina.

I dalje ima prostora za napredak. Usled modifikovanog minibatch algoritma, došlo je do određenog preobučavanja na klasi dobrih vina. Takođe se može istražiti da li se može postići bolja ukupna tačnost.

Sa ovim podacima može se dalje istraživati na temu uticaja karakteristika na krajnji rezultat. Isto se može istraživati pronalazak neuobičajenih vrednosti.