

KORIŠĆENJE SOFTVERSKIH PAKETA OTVORENOG KODA U TOKU NASTAVNIH PROCESA RADI REŠAVANJA TEORIJSKIH I EKSPERIMENTALNIH PROBLEMA

Zoran Ovcin¹, Ksenija Doroslovački², Biljana Carić³

^{1,2,3}Univerzitet u Novom Sadu, Fakultet tehničkih nauka, Novi Sad, Srbija

¹zovcin@uns.ac.rs, ²ksenija@uns.ac.rs, ³biljana@uns.ac.rs

Kratak sadržaj: Softverski i programski paketi se kao alati koriste u naučnim istraživanjima i nastavnim aktivnostima. Mnogi softverski paketi su dati kao otvoreni kod (Open Source). Gradivo Linearne algebre, Analize funkcija više promenljivih i Verovatnoće i slučajnih promenljivih spojeno u predmet Matematičke osnove mašinskog učenja na Master akademskim studijama se izlaže kao otvoreni kod. Koristi se literatura (MML-Book) i softverski paketi otvorenog koda (R-project). Iako kao alternativu imaju i komercijalne alate i literaturu, studenti se iz više razloga odlučuju za predloženi pristup. Rešavanje postavljenih problema ih osposobljava za savladavanje gradiva studijskog programa Veštačka inteligencija i mašinsko učenje.

Ovaj metodološki pristup izučavanja Matematičkih osnova mašinskog učenja nadgrađuje znanja i veštine osvojene na osnovnim akademskim studijama. Pristup softverskim paketima otvorenog koda je jednostavan i besplatan. Svetska zajednica otvorenog koda se širi. Studenti obučavani na ovaj način imaju omogućen lak i brz pristup toj zajednici. Postoji i velika mogućnost pristupu razvoja softverskih paketa otvorenog koda. To često otvara vrata za ulaz u privredu i razvoj komercijalnih proizvoda visoke tehnologije i veštačke inteligencije.

Rezultat je praćenje novih, aktuelnih istraživanja uz mogućnost pristupa ogromnoj zajednici otvorenog koda (Open Source Community). Višestruki su razlozi primene softverskih paketa otvorenog koda i za studente koji iskazuju interesovanje da se dalje bave naučnim radom.

Ključne reči: Softverski paketi, Otvoreni kod, Mašinsko učenje, Matematika za mašinsko učenje

1. UVOD

Cilj ovog rada su teorijska, eksperimentalna i primenjena istraživanja u oblastima koje su interesantne Katedri za matematiku na Fakultetu tehničkih nauka u Novom Sadu.

Softverski i programski paketi su nam neophodni alati u naučnim istraživanjima, pa samim tim i u nastavnim aktivnostima. Takođe, korišćenje novih programskih paketa u toku nastavnih procesa je potrebno za bolje razumevanje gradiva, kao i rešavanje teorijskih i eksperimentalnih problema.

Korišćenjem programskih paketa i postojećih metodoloških pristupa dolazimo do korisnih rezultata koji imaju primenu u skoro svim tehničkim granama pa tako i u medicini, ekologiji, farmakologiji, ekonomiji pa čak i u društvenim naukama kao što su sociološka pitanja i potrebe za određivanje koncenzusa, psihologiji i slično. Multidisciplinarnost je u današnje vreme sve češće zastupljena zbog potreba istraživanja u navedenim oblastima i njena potreba je bitna u razvoju akademskog obrazovanja i daljih istraživanja.

Primenjena matematika u tehničkim disciplinama je značajna radi pružanja celokupnog znanja i osposobljavanje za rad u strukama tehničkih disciplina. Studenti tehničkih fakulteta, inženjeri koji su stekli znanja iz raznih grana matematike, kao što su: matematička analiza, algebra, matematička logika, diskretna matematika, verovatnoća i statistika, slučajni sistemi, numerička matematika i druge matematičke discipline imaju znanja koja su potrebna za dalju realizaciju tehničkih rešenja koja se implementiraju softverskim paketima.

Motiv za korišćenje softverskih paketa u toku nastavnih procesa je praćenje novih tehnologija, naučnih rezultata i njihove primene u privredi. Od samog početka razumevanja i rešavanja problema treba problem rešiti ne samo analitički već i softverski. Samim tim problem, odnosno teorijsko rešenje problema će biti upotpunjeno a i softverski implementirano, što dodatno daje na značaju boljeg razumevanja, a i samoj proverbi da nakon softverskog testiranja nije došlo do greške. Provera je dvostrana - da nije došlo do greške prilikom analitičkog rešavanja kao i da nije došlo do greške prilikom pravljenja softvera.

Možemo reći da motivacija za rešavanje problema prvo analitički (teorijski), a onda i eksperimentalno je postupak koji će nam dati traženi rezultat. Potreba za veštinama korišćenja programskih paketa bitna je već na polaznim kursevima. Ovaj vid edukacije se dalje treba primenjivati i na ostalim nivoima studija tako da bi na kraju naši završeni studenti bili u svakom smislu spremni da primene algoritme na konkretnim problemima u određenim inženjerskim oblastima. Trenutno, na Fakultetu tehničkih nauka (FTN) Univerziteta u Novom Sadu (UNS) primena softverskih paketa je veća na višim godinama osnovnih akademskih studija i master akademskim studijama.

2. OTVORENI KOD (OPEN SOURCE)

Već 80-tih godina prošlog veka pojavili su se programeri koji su svoje programe davali besplatno na upotrebu. Mnogi programeri su i programski kod (Source) davali na uvid preko Interneta kao otvoreni kod (Open Source). Pojavile

su se grupe koje su u kolaboraciji pisali programe ili programske biblioteke i objavljivali ih u otvorenom kodu (na primer GNU grupa). Programi koje su pisali ostaju intelektualna svojina programera data na besplatno korišćenje. Ipak, da ne bi neko prisvojio intelektualnu svojinu, kreirane su licence koje dozvoljavaju drugim programerima upotrebu i uključivanje u svoje programe, ali sa jasnim navođenjem autora licenciranog koda i zahtevom da novonastali kod ostane pod istom licencom. Tako je nastala čuvena General Public License (GPL) i druge licence (LGPL, GPL v2, ...) koje dozvoljavaju manje ili više od GPL. Postoje licence za druge intelektualne svojine koje nisu programski kod, na primer fotografije, koje autori daju na besplatno korišćenje pod uslovom da se čuva podatak o autoru i da novonastala upotreba koja uključuje licenciranu intelektualnu svojinu ostaje besplatna ili pod istom licencom.

Softverski paketi su plod timskog rada. Razvoj Interneta 90-tih godina prošlog veka omogućio je stvaranje timova koji nisu bili fizički zajedno, čak bi bili u različitim vremenskim zonama. Pojavilo se nekoliko sistema koji su na Internet serverima omogućavali timovima rad na istom projektu sa mogućnostima kolaboracije, dokumentovanja, praćenja rada, rešavanja konflikata, ... Najpopularniji je GIT sistem. Najveći i najpopularniji server GIT sistema je GitHub ripozitori (repository = skladište podataka), osnovan 2007. godine. Od 2018. vlasništvo GitHub ripozitorija je preuzeo Microsoft. Naravno, time nije preuzeto vlasništvo nad softverskim paketima koji su uskladišteni na GitHubu. GitHub trenutno skladišti oko 200 miliona ripozitorija (skladišta softverskih projekata) za oko 80 miliona korisnika. Čak se i programski kod Microsoft Windows razvija na GitHubu.

3. R PROJEKAT

R je programski jezik koji su kreirali R Core Team i R Fondacija za statističko računarstvo prvenstveno za statistička istraživanja i grafiku. R je od 1995. godine razvijan kao Open source projekat pod GPL licencom. R se koristi za data mining, u bioinformatiči i statistici, za analizu podataka i razvoj statističkog softvera, ... R ima interfejs prema mnogim programskim jezicima i paketima.

Vremenom su korisnici razvijali dodatne pakete za R od kojih su većina ostali GPL. Za razvoj i distribuciju R paketa se koristi CRAN, i, sve više, GitHub ripozitori.

R ima značajnu upotrebu u mašinskom učenju i razvoju softvera veštačke inteligencije.

4. UPOTREBA U NASTAVI

Analiziraćemo postojeće metodološke pristupe u programskom jeziku i okruženju R, koji takođe koristimo u nastavi na predmetu Matematičke osnove mašinskog učenja (Mathematics for machine learning - **MML**) na Master akademskim studijama **Veštačka inteligencija i mašinsko učenje (VIMU)** na FTN UNS. Takođe navodimo da se ista rešenja mogu implementirati i u Matlabu, Pythonu, ...

U akreditaciji studijskog programa VIMU iz 2021. godine, obavezan predmet u prvom semestru je 22.EAI001 Matematičke osnove mašinskog učenja. Glavni cilj predmeta je kombinovanje znanja iz **Linearne algebre, Analize funkcija više promenljivih i Verovatnoće i slučajnih promenljivih**.

Studenti koji prate ovaj predmet su već ovladali osnovama ove tri oblasti na osnovnim akademskim studijama. Kroz obnavljanje i objedinjavanje znanja iz ove tri oblasti javlja se mogućnost za primenu u linearnim regresionim modelima i gradijentnim postupcima optimizacije. Ideja je da se na taj način naprave prvi koraci za podučavanje mašinskog učenja.

Primenjena linearna algebra koja se bavi rešavanjem sistema linearnih jednačina velikih dimenzija, teorijom matrica, karakterističnih korena, dekompozicije matrica, je aktuelno polje istraživanja i predstavlja osnovnu alatku u inženjerskim oblastima. **Optimizacija** je problem nalaženja vrednosti parametara neke funkcije (funkcija cilja), koji utiču na posmatrani proces, radi postizanja boljih rezultata. Postoje brojne matematičke metode za nalaženje vrednosti parametara za koje funkcija cilja postiže optimalnu vrednost (najmanju ili najveću). Na ovom kursu se pomoću Analize funkcija više promenljivih podučavaju gradijentni postupci.

Zavisnost funkcije cilja a i sami parametri su neretko modelirani slučajnim veličinama. Verovatnoća, Slučajne promenljive i Statistika su oblasti koje su obrađivane na osnovnim akademskim studijama. Zavisnost slučajnih veličina od parametara se često može modelirati **Linearnom regresijom**. Ova oblast se u okviru studijskog programa VIMU izučava upotrebom regresionih postupaka ugrađenih u programski paket R.

Sjedinjavanje ove tri oblasti u obavezan studijski program daje kao ishod ovladavanje veštinama i znanjima koja se mogu prepoznati i primeniti u skoro svim oblastima mašinskog učenja.

5. ZAŠTO R I GITHUB?

Gradivo predmeta MML na VIMU je velikim delom pokriveno knjigom [1.]. Izbor R-projekta za platformu na kojoj će se rešavati problemi je individualan. Studenti mogu odabrati da probleme rešavaju i u drugim softverskim paketima kao što je MATLAB ili primenom specijalizovanih biblioteka u programskom jeziku Python.

Ipak, licenca za MATLAB je skupa, a primena Pythona komplikovana.

Mnogi softverski paketi u svom razvoju u poslednjih nekoliko decenija imaju ugrađene najnovije postupke za rešavanje problema linearne algebre. Još uvek ne možemo reći da ti postupci mogu da nam daju rešenje svih specijalnih slučajeva tako da su i dalje aktuelna istraživanja iz primenjene linearne algebre pomoću kojih ćemo doći do novih rešenja, a samim tim i do poboljšanja u implementaciji softvera.

Programski jezik R barata sa matricama kao jednim od osnovnih tipova podataka. Zato su primeri iz gradiva

Analize funkcija više promenljivih lako implementirani. Posebno primena za optimizaciju gradijentnim postupcima koji se često koriste u mašinskom učenju.

R-projekat je prvenstveno nastao radi primene u Statistici. Programi `lm` i `glm` iz R-a se koriste za modeliranje opštim i složenim linearnim modelima response promenljive preko linearnih prediktora. Takođe se koriste za opis raspodele greške modela. Ova dva programa su na neki način postali standard pri razmeni rezultata Opšte linearne regresije i Složenih linearnih modela.

Nastava na MML je održana auditorno, upotrebom klasičnih nastavnih sredstava prateći potrebna poglavlja iz osnovne literature [1.]. Primeri za vežbu iz [1.] kao i zadaci koji se pojavljuju na kolokvijumu i pismenom ispitu su rešavani primenom R programskog jezika koristeći biblioteku `matlib`.

R projekat je Open source, isto kao i biblioteka `matlib`. Njihov razvoj se nalazi na GitHub ripozitoriju, isto kao što je i knjiga [1.]. GitHub omogućava kolaboraciju pri rešavanju problema isto kao i u razvoju projekata. Skoro sigurno studenti koji završe MAS VIMU će u svom radu koristiti GitHub i kao literaturu i kao okruženje za razvoj.

Iz svega ovde navedenog, jasno je da je izbor R-projekta kao alatke i GitHub ripozitorija kao medija najlogičniji i optimalan.

6. IMPLEMENTACIJA

Za izvođenje nastave MML na VIMU otvoren je GitHub projekat i studentima je omogućen pristup. Na ripozitorij je postavljen kod u R-u kojim se rešavaju primeri iz literature.

Koristeći R Markdown (R opisni format) dokumenti se mogu potpuno reprodukovati. Interfejs kao što je R-studio se koristi za spajanje narativnog teksta i R koda da bi se proizveo elegantno formatiran izlaz. Izlaz se može izvesti i u Word, ili LaTeX format radi doterivanja.

7. REZULTATI

Kolokvijumi i ispiti su pokazali da studenti lako savladavaju predložene alatke (R-projekat i GitHub ripozitorij) i da su od svih softverskih paketa za rešavanje postavljenih problema najviše odabrali literaturu i softverske pakete otvorenog koda. Rešavanje postavljenih problema ih osposobljava za savladavanje gradiva studijskog programa MAS VIMU.

Ovaj metodološki pristup izučavanja Matematičkih osnova mašinskog učenja nadgrađuje znanja i veštine osvojene na osnovnim akademskim studijama.

Pristup softverskim paketima otvorenog koda je jednostavan i besplatan. Svetska zajednica otvorenog koda se širi. Studenti obučavani na ovaj način imaju omogućen lak i brz pristup toj zajednici. Postoji i velika mogućnost pristupu razvoju softverskih paketa otvorenog koda. To često otvara vrata za ulaz u privredu i razvoj komercijalnih proizvoda visoke tehnologije.

8. ZAKLJUČAK

Treba pratiti nova, aktuelna istraživanja i sagledati njihov dalji razvoj i tehnička rešenja sa stanovišta primenljivosti matematičkih rezultata. To je ostvarivo korišćenjem softverskih paketa otvorenog koda.

9. ZAHVALNICA

Autori se zahvaljuju na podršci u okviru projekta Departmana za opšte discipline u tehnici, Fakulteta tehničkih nauka pod naslovom: Primena informaciono-komunikacionih tehnologija u nastavi opštih disciplina (Application of information and communication technologies in the teaching of fundamental disciplines).

Literatura:

- [1] Deisenroth, Marc P., Faisal, A. Aldo and Ong, Cheng Soon. *Mathematics for Machine Learning*. Cambridge University Press, <https://github.com/mml-book/mml-book.github.io> 2020.
- [2] R Core Team, *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.r-project.org>. 2022.
- [3] Yihui Xie and J.J. Allaire and Garrett Grolemond. *R Markdown: The Definitive Guide*. Chapman and Hall/CRC. ISBN 9781138359338. URL <https://bookdown.org/yihui/rmarkdown>. (2018)

10. DODATAK

Ovde vidimo upotrebu R Markdown koristeći alat knitr unutar Rstudio. Na ovaj način dobijamo prikaz komandi sa narativnim opisom i rezultat njihovog izvršavanja. Celo poglavlje DODATAK je nastalo pomoću primene paketa knitr na ulaznom fajlu primer.Rmd sa malo doterivanja. Rešen je jedan zadatak sa vežbi na MML, MAS VIMU.

Da bi se koristio R Markdown i knitr, treba:

- Instalirati R
- Instalirati RStudio

- Instalirati *knitr* paket: `install.packages("knitr")`
- Da bi izvršili ovaj kod u R Studiju idite na File -> New -> R Markdown
- Instalirajte *matlib* paket: `install.packages("matlib")` i pakete koje *matlib* povlači
- Sačuvajte sa `.Rmd` ekstenzijom
- Kliknite na Knit to Word i potom doterajte Word dokument

Data je matrica A_Φ linearne transformacije $\Phi: \mathbb{R}^{4 \times 1} \rightarrow \mathbb{R}^{4 \times 1}$: $A_\Phi = \begin{bmatrix} 1 & 0 & -2 & 1 \\ 2 & -1 & 0 & 1 \\ 3 & -2 & 2 & 1 \\ 2 & 0 & -4 & 2 \end{bmatrix}$.

1. Naći $\dim \operatorname{Im} \Phi$ i $\dim \operatorname{Ker} \Phi$. Primenjujući elementarne (Gausove) transformacije na matricu A_Φ dobijamo

```
library('matlib')
a1=c(1, 2, 3, 2)
a2=c(0, -1, -2, 0)
a3=c(-2, 0, 2, -4)
a4=c(1, 1, 1, 2)
a=cbind(a1,a2,a3,a4)
gaussianElimination(a)

      a1 a2 a3 a4
[1,] 1 0 -2 1
[2,] 0 1 -4 1
[3,] 0 0 0 0
[4,] 0 0 0 0
```

odakle vidimo da je $\operatorname{rank} A_\Phi = \dim \operatorname{Im} \Phi = 2$. Onda je $\dim \operatorname{Ker} \Phi = 4 - 2 = 2$.

2. Naći bazu prostora $\operatorname{Im} \Phi$.

Iz rezultata primene Gausove eliminacije vidimo da su prve dve kolone matrice A_Φ linearno nezavisne.

$$\text{Zato je } \operatorname{Im} \Phi = \operatorname{span} \left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ -2 \\ 0 \end{bmatrix} \right\}$$

3. Naći ortogonalnu bazu prostora $\operatorname{Im} \Phi$.

```
u1=c(1, 2, 3, 2)
u2=c(0, -1, -2, 0)
b=cbind(u1,u2)
bgs=GramSchmidt(b, normalize = FALSE, verbose = TRUE)
```

Initial matrix: u1 u2 [1,] 1 0 [2,] 2 -1 [3,] 3 -2 [4,] 2 0

Column 1: z1 = x1 u1 u2 [1,] 1 0 [2,] 2 0 [3,] 3 0 [4,] 2 0

Column 2: z2 = x2 - Proj(x2, z1) u1 u2 [1,] 1 0.4444444 [2,] 2 -0.1111111 [3,] 3 -0.6666667 [4,] 2 0.8888889

Da bismo dobili cele brojeve

```
bgs %*% diag(c(1,9))
```

```
      [,1] [,2]
[1,] 1 4
[2,] 2 -1
[3,] 3 -6
[4,] 2 8
```

$$\text{Vidimo da je } \operatorname{Im} \Phi = \operatorname{span} \left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \\ 2 \end{bmatrix}, \begin{bmatrix} 4 \\ -1 \\ -6 \\ 8 \end{bmatrix} \right\}$$