



CS324 - SKRIPTING JEZICI

Python i mašinsko učenje

Lekcija 12

PRIRUČNIK ZA STUDENTE

CS324 - SKRIPTING JEZICI

Lekcija 12

PYTHON I MAŠINSKO UČENJE

- ✓ Python i mašinsko učenje
- ✓ Poglavlje 1: Uvod u veštačku inteligenciju
- ✓ Poglavlje 2: Algoritmi mašinskog učenja
- ✓ Poglavlje 3: Linearna regresija
- ✓ Poglavlje 4: Logistička regresija
- ✓ Poglavlje 5: Paket scikit-learn
- ✓ Poglavlje 6: Pokazne vežbe #12
- ✓ Poglavlje 7: Individualne vežbe #12
- ✓ Poglavlje 8: Domaći zadatak
- ✓ Zaključak

Copyright © 2017 – UNIVERZITET METROPOLITAN, Beograd. Sva prava zadržana. Bez prethodne pismene dozvole od strane Univerziteta METROPOLITAN zabranjena je reprodukcija, transfer, distribucija ili memorisanje nekog dela ili čitavih sadržaja ovog dokumenta., kopiranjem, snimanjem, elektronskim putem, skeniranjem ili na bilo koji drugi način.

Copyright © 2017 BELGRADE METROPOLITAN UNIVERSITY. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, without the prior written permission of Belgrade Metropolitan University.

▼ Uvod

UVOD

Uvod u lekciju #12

U dvanaestoj lekciji biće reči o korišćenju Python programskog jezika u algoritmima veštačke inteligencije, konkretno za mašinsko učenje.

Python kao programski jezik se dobro pokazao kada je u pitanju manipulacija nad podacima, i upravo je jedna od primena rada sa velikom količinom podataka veštačka inteligencija, sa naglaskom na mašinsko učenje, kao oblast veštačke inteligencije koja je u velikom usponu.

Veštačka inteligencija i mašinsko učenje se uče na posebnim predmetima, ali u ovoj lekciji uvode se osnovni pojmovi mašinskog učenja kao i najjednostavniji algoritmi.

U delu lekcije koji se odnosi na programski jezik Python, biće reči o **scikit-learn** paketu za veštačko učenje i modul **linear_model**, za dva algoritma koja će se obrađivati.

Ova lekcija sadrži video materijal. Ukoliko želite da pogledate ovaj video morate da otvorite LAMS lekciju.

▼ Poglavlje 1

Uvod u veštačku inteligenciju

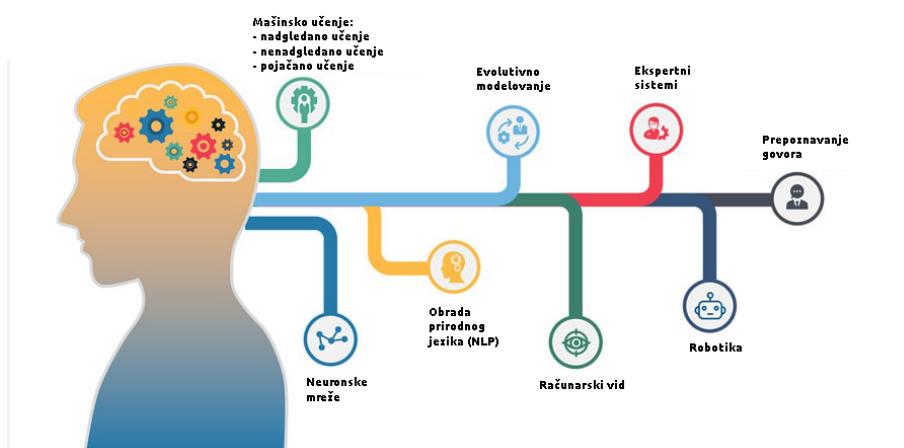
POJAM VEŠTAČKE INTELIGENCIJE

Veštačka inteligencija se koristi za konstruisanje inteligencije upotrebom hardverskih i softverskih rešenja.

Veštačka inteligencija (en. **Artificial Intelligence**, AI) je nauka koja se koristi za konstruisanje inteligencije upotrebom hardverskih i softverskih rešenja.

Veštačka inteligencija automatizuje ljudsku inteligenciju na osnovu načina na koji ljudski mozak obrađuje informacije

Elementi veštačke inteligencije



Slika 1.1 Elementi veštačke inteligencije. [Izvor: Autor]

Ova lekcija sadrži video materijal. Ukoliko želite da pogledate ovaj video morate da otvorite LAMS lekciju.

DISCIPLINE KOJE UKLJUČUJE AI

Eksperti u domenu veštačke inteligencije trebaju znati koje su discipline pridružene veštačkoj inteligenciji.

Za kreiranje dobrog ali i upotrebljivo AI rešenja, potrebno je uključiti različite discipline.

- **robotika** - za pomeranje objekata u prostoru,
- **teorija algoritama** - za konstruisanje efikasnih algoritama,

- **statistika** - za izvođenje korisnih rezultata, predviđanje budućnosti i analizu prošlosti,
- **psihologija** - za modelovanje načina funkcionisanja ljudskog mozga
- **softversko inženjerstvo** - za kreiranje rešenja koja se mogu održavati i koja mogu izdržati test vremena
- **računarska nauka** - za implementaciju softverskih rešenja u praksi
- **matematika** - za izvršavanje složenih matematičkih operacija
- **upravljanje sistema** - za kreiranje feed-forward i feedback sistema
- **teorija informacija** - za predstavljanje, kodiranje, dekodiranje i kompresovanje informacija
- **teorija grafova** - za modelovanje i optimizovanje različitih tačaka u prostoru i za predstavljanje hijerarhija
- **fizika** - za modelovanje realnog sveta
- **računarska grafika i obrada slika** - za prikaz i obradu slika i filmova

Ova lekcija sadrži video materijal. Ukoliko želite da pogledate ovaj video morate da otvorite LAMS lekciju.

PRIMER: KAKO AI REŠAVA PROBLEME IZ REALNOG SVETA?

Veštačka inteligencija automatizuje ljudsku inteligenciju na osnovu načina na koji ljudski mozak obrađuje informacije.

Veštačka inteligencija automatizuje ljudsku inteligenciju na osnovu načina na koji ljudski mozak obrađuje informacije.

Kad god rešimo problem ili komuniciramo sa ljudima, prolazimo kroz proces. Kad god ograničimo opseg problema ili interakcije, ovaj proces često može da bude modelovan i automatizovan.

AI izvršavaju računari koji izvršavaju instrukcije nižeg nivoa.

Većina AI aplikacija ima jedan primarni cilj. Kada komuniciramo sa AI aplikacijom, čini se kao da je ljudska, jer može da ograniči domen problema na primarni cilj. Prema tome, možemo da razdvojimo složene procese i simuliramo inteligenciju pomoću računarskih instrukcija nižeg nivoa.

AI može da stimuliše ljudska čula i procese razmišljanja za specijalizovane oblasti.

Treba da simuliramo ljudska čula i razmišljanje i ponekad da prevarimo AI da veruje da komuniciramo sa drugim ljudskim bićem. U posebnim slučajevima možemo čak da poboljšamo i naša čula.

Slično tome, kada komuniciramo sa chatbotom, očekujemo da nas bot razume. Očekujemo da chatbot ili čak sistem za prepoznavanje govora obezbede interfejs između računara i čoveka. Da bi ispunili naša očekivanja, računari treba da simuliraju procese ljudskog razmišljanja.

Ova lekcija sadrži video materijal. Ukoliko želite da pogledate ovaj video morate da otvorite LAMS lekciju.

Ova lekcija sadrži video materijal. Ukoliko želite da pogledate ovaj video morate da otvorite LAMS lekciju.

PRIMENA AI: SINTEZA GOVORA

Prepoznavanje govora omogućuje razumevanje ljudskog govora od strane mašine. Sinteza govora omogućuje mašini da "priča".

Simulacija ljudskog ponašanja

Ljudi imaju pet osnovnih čula: vida, sluha, dodira, mirisa i ukusa. Međutim, da bismo bolje razumeli kako da kreiramo inteligentne mašine, možemo da razdvojimo discipline na sledeći način:

- slušanje i govor,
- razumevanje jezika,
- pamćenje,
- razmišljanje,
- vid,
- pokret

Na primer, kada želimo da pomerimo robotsku ruku, treba da poznajemo složenu matematiku visokog nivoa da bismo razumeli šta se dešava.

Ova lekcija sadrži video materijal. Ukoliko želite da pogledate ovaj video morate da otvorite LAMS lekciju.

Slušanje i govor

Upotrebom sistema za prepoznavanje govora AI može da sakuplja informacije. Upotrebom sinteze govora može da pretvori interne podatke u razumljive zvukove. Tehnike prepoznavanja govora i sinteze govora prepoznaju i konstruišu zvukove koje ljudi emituju ili koje ljudi mogu da razumeju.

Zamislite da ste na putovanju u državi u kojoj se govori jezikom koji vi ne razumete. Možete da govorite u mikrofonskog telefona, očekujući da telefon „razume“ ono što vi kažete, a zatim da to prevede na drugi jezik. Isto se dešava i obratno - kada lokalni stanovnik govori, a AI prevodi zvukove na jezik koji razumete. Prepoznavanje govora i sinteza govora to omogućavaju.

Primer sinteze govora je Google Translate. Dugme sa zvučnikom ispod prevedene reči da bi prevodilac glasno izgovorio reči na jeziku koji nije engleski.

PRIMENA AI: NLP I NN

Duboko učenje predstavlja "crnu kutiju" jer prevazilazi obrazac kategorizacije.

Razumevanje jezika

Možemo da razumemo prirodni jezik ako ga obradimo. Ova oblast se naziva obrada prirodnog jezika (en. **natural language processing**, NLP). Kada je reč o obradi prirodnog jezika, želimo da naučimo jezike na osnovu statističkog učenja.

Pamćenje

Treba svetlu da predstavimo nešto što poznajemo. Za to kreiramo baze znanja i hijerarhijske reprezentacije pod nazivom ontologije. Ontologije kategorišu elemente i ideje u našem svetlu i sadrže veze između ovih kategorija.

Ova lekcija sadrži video materijal. Ukoliko želite da pogledate ovaj video morate da otvorite LAMS lekciju.

Razmišljanje

AI sistem treba da bude ekspert u određenom domenu upotrebom ekspertskog sistema, koji može da bude zasnovan na matematičkoj logici na deterministički način, kao i na nejasan, nedeterministički način.

Baza znanja ekspertskog sistema je predstavljena upotrebom različitih tehnika. Kako problem domena raste, kreiramo hijerarhijske ontologije.

Možemo da repliciramo ovu strukturu modelovanjem mreže na gradivnim blokovima mozga.

Ovi gradivni blokovi se nazivaju *neuroni*, a sama mreža se naziva **neuronska mreža** (en. **neural network**, NN).

Postoji još jedan ključni termin koji treba da povežemo sa neuronskim mrežama - **duboko učenje** (en. **deep learning**).

Ovakvo učenje se naziva duboko jer prevazilazi obrazac prepoznavanja i kategorizacije. Ono je utisnuto u neuronsku strukturu mreže. Na primer, specijalni zadatak dubokog učenja je prepoznavanje objekta pomoću računarskog vida.

PRIMENA AI: RAČUNARSKI VID I ROBOTIKA

Robotika se bavi složenim funkcijama omogućavanje pokreta mašina. Računarski vid omogućava mašini da prepozna šta se nalazi na slici ili na video sadržaju.

Vid

Treba da komuniciramo sa realnim svetom pomoću naših čula. Do sada smo govorili samo o čulu sluha, o prepoznavanju i sintezi govora. Šta se dešava ako treba da vidimo nešto? U tom

slučaju treba da kreiramo tehnike računarskog vida za učenje o okruženju. Prepoznavanje lica je korisno i većinom su ljudi u tome eksperti.

Računarski vid (en. **computer vision**) zavisi od obrade slike. Iako obrada slike nije direktno AI disciplina, ona je potrebna disciplina za AI.

Ova lekcija sadrži video materijal. Ukoliko želite da pogledate ovaj video morate da otvorite LAMS lekciju.

Pokret

Pokret i dodir su prirodni za nas ljude, ali su veoma složeni zadaci za računare. Pokret se vrši pomoću robotike. Robotika je zasnovana na upravljanju sistemima, gde kreiramo povratnu petlju i kontrolišemo pokret objekta na osnovu sakupljenih povratnih informacija.

Ova lekcija sadrži video materijal. Ukoliko želite da pogledate ovaj video morate da otvorite LAMS lekciju.

MAŠINSKO UČENJE

Mašinsko učenje predstavlja polje znanja koje daje računarima mogućnost da uče bez eksplicitnog programiranja.

Mašinsko učenje (en. **Machine Learning**, ML), pored obrade prirodnog jezika (en. **Natural Language Processing**, NLP), računarske vizije (en. **Computer Vision**, CV) i ekspertskih sistema (en. **Expert Systems**, ES), je pod oblast veštačke inteligencije koja mašinama (računarima) pruža mogućnost učenja bez eksplicitnog programiranja.

Definicija mašinskog učenja

Arthur Samuel (1959):

Mašinsko učenje predstavlja polje znanja koje daje računarima mogućnost da uče bez eksplicitnog programiranja.

Tom Mitchell (1998):

Za računarski program kaže se da uči da rešava određeni zadatak T, iz iskustva E, i merom performansi P, ukoliko se njegove performanse rešavanja zadatka T, mereno prema P, poboljšavaju kroz iskustvo E.

Uopšteno posmatrano, proces mašinskog učenja sadrži šest koraka:

1. Prikupljanje podataka,
2. Pre-procesiranje, „čišćenje“ i formatiranje podataka,
3. Podela podataka na skup podataka za treniranje i za testiranje,
4. Kreiranje i treniranje modela mašinskog učenja nad skupom podataka za treniranje,
5. Testiranje istreniranog modela korišćenjem skupa podataka za testiranje, prilikom čega se dobijaju različite metrike koje pokazuju koliko je model „dobar“,
6. Unapređivanje i isporuka modela.

U profesionalnom okruženju, u trećem koraku se najčešće inicijalni skup podataka deli na tri, a ne na dva podskupa: treniranje, validacija i testiranje. Naime, skup podataka za treniranje se koristi za *fitovanje*, odnosno treniranje parametara modela. Potom, skup podataka za validaciju se koristi za dobijanje metrika koje možemo iskoristiti za unapređivanje modela podešavanjem njegovih parametara. Konačno, skup podataka za testiranje se koristi za dobijanje finalnih metrika performansi modela.

▼ Poglavlje 2

Algoritmi mašinskog učenja

RAZLIČITI ALGORITMI MAŠINSKOG UČENJA

Mašinsko učenje deli se na nadgledano, nenadgledano i pojačano učenje

Algoritmi mašinskog učenja se mogu podeliti u tri osnovne grupe:

- Nadgledano učenje (en. *Supervised learning*),
- Nenadgledano učenje (en. *Unsupervised learning*),
- Pojačano učenje (en. *Reinforcement learning*)

U nastavku ovog poglavlja biće ukratko objašnjene ove tri osnovne grupe algoritama.

Ova lekcija sadrži video materijal. Ukoliko želite da pogledate ovaj video morate da otvorite LAMS lekciju.

Ova lekcija sadrži video materijal. Ukoliko želite da pogledate ovaj video morate da otvorite LAMS lekciju.

NADGLEDANO UČENJE

Kod nadgledanog učenja, algoritmu se pružaju ulazni i izlazni istorijski podaci iz kojih algoritam uči kako da od ulaznih podataka „proizvede“ izlazne podatke.

Kod nadgledanog učenja, algoritmu se pružaju ulazni i izlazni istorijski podaci iz kojih algoritam uči kako da od ulaznih podataka „proizvede“ izlazne podatke.

Skup podataka za treniranje modela se sastoji od niza primera; svaki primer sadrži jednu osobinu ili više njih, kao i željeni izlaz. Samim tim, najčešći format skupa podataka jeste matrica sastavljena od niza vektora koji reprezentuju primere.

Kroz iterativnu optimizaciju, algoritmi nadgledanog učenja formiraju funkciju koja se može koristiti za predviđanje rezultata primera koji nisu bili korišćeni za treniranje modela.

Kako izlazi primera mogu biti kontinualne i diskretne vrednosti, tako delimo i algoritme nadgledanog mašinskog učenja u algoritme.

Ova lekcija sadrži video materijal. Ukoliko želite da pogledate ovaj video morate da otvorite LAMS lekciju.

ALGORITMI REGRESIJE

Ovi algoritmi pokušavaju da mapiraju ulazne podatke u kontinualne funkcije.

Algoritmi regresije

Ovi algoritmi koji pokušavaju da mapiraju ulazne podatke u kontinualne funkcije koje, samim tim, proizvode vrednosti kontinualnog tipa (npr. predviđanje cena nekretnine, predviđanje broja godina, i slično).

Najpoznatiji algoritmi regresije su:

- **Linearna regresija** (en. *Linear Regression*),
- **Višestruka linearna regresija** (en. *Multiple Linear Regression*),
- **Polinomna regresija** (en. *Polynomial Regression*).

Ova lekcija sadrži video materijal. Ukoliko želite da pogledate ovaj video morate da otvorite LAMS lekciju.

ALGORITMI KLASIFIKACIJE

Ovi algoritmi pokušavaju da mapiraju ulazne podatke u diskretne kategorije.

Algoritmi klasifikacije

Ovi algoritmi pokušavaju da mapiraju ulazne podatke u diskretne kategorije (npr. predviđanje da li je tumor maligni ili benigni, kojoj kategoriji pripada novinski članak, koja je vrsta životinje prikazana na slici i slično). U zavisnosti od broja mogućih kategorija izlaza, klasifikacija može biti binarna i višeklasna.

Najpoznatiji algoritmi klasifikacije su:

- **Logistička regresija** (en. *Logistic Regression*),
- **Stablo odlučivanja** (en. *Decision Tree*),
- **Slučajne šume** (en. *Random Forest*),
- **Naivni Bajes** (en. *Naive Bayes*),
- **K-najbližih suseda** (en. *K-Nearest Neighbors*),
- **Metoda potpornih vektora** (en. *Support Vector Machine*).

Ova lekcija sadrži video materijal. Ukoliko želite da pogledate ovaj video morate da otvorite LAMS lekciju.

NENADGLEDANO UČENJE

Kod nenadgledanog učenja, algoritmu se pružaju samo ulazni podaci primera koje treba obraditi.

Kod nenadgledanog učenja, algoritmu se pružaju samo ulazni podaci primera koje treba obraditi, ne i odgovarajući izlazni podaci, te najčešće ne postoji bilo kakva povratna informacija o tome kako treba da izgledaju izlazni podaci.

Algoritam ima za cilj otkrivanje šablona i grupisanje tih neoznačenih podataka, kako bi otkrio neku vrstu strukture istih.

Grupisanje

Grupisanje (en. **Clustering**), metoda za grupisanje u skupu podataka na osnovu nezavisnih varijabli (npr. analiza osoba na društvenim mrežama, analiza potrošača, grupisanje gena koji su slični ili povezani različitim varijablama). Najpoznatiji algoritam iz ove klase je algoritam **K-srednje vrednosti** (en. **K-Means**).

Ova lekcija sadrži video materijal. Ukoliko želite da pogledate ovaj video morate da otvorite LAMS lekciju.

POJAČANO UČENJE

Pojačano učenje je model mašinskog učenja koji se bavi načinom na koji softverski agenti treba da preduzmu akcije u okruženju u kome se nalaze kako bi maksimizovali nagradu.

Pojačano učenje je model mašinskog učenja koji se bavi načinom na koji softverski agenti treba da preduzmu akcije u okruženju u kome se nalaze kako bi maksimizovali nagradu.

Pojačano učenje „vuče korene“ iz psihologije ponašanja, te se često kao primer u toku objašnjavanja pojačanog učenja uzima dresiranje pasa. Naime, kad psu kažemo šta treba da uradi, verovatno nas neće odmah razumeti, ali metodom nagrađivanja i kažnjavanja pas polako uči šta je uradio dobro odnosno loše. Primenom iste metode možemo istrenirati programe da izvršavaju različite zadatke, igraju igrice i slično.

Najpoznatiji algoritmi pojačanog učenja su **Markovljev proces donošenja odluke** (en. **Markov Decision Process**) i **Kju-učenje** (en. **Q-Learning**).

Ova lekcija sadrži video materijal. Ukoliko želite da pogledate ovaj video morate da otvorite LAMS lekciju.

OPTIMIZACIJA ALGORITAMA

Jednom istrenirani algoritam mašinskog učenja može se optimizovati korišćenjem različitih tehnika.

Jednom istrenirani algoritam mašinskog učenja može se optimizovati korišćenjem različitih tehnika.

Smanjivanje dimenzionalnosti (en. **Dimensionality Reduction**), metoda za smanjivanje broja nezavisnih varijabli prilikom treniranja u tzv. glavne komponente koje sažeto prenose slične informacije. Svrha smanjivanja dimenzionalnosti se ogleda u ubrzavanju ostalih algoritama mašinskog učenja, jer je tako potreban manji broj računanja; kao i pronalaženje grupe najpouzdanijih karakteristika odnosno nezavisnih varijabli u podacima.

Najpoznatiji algoritam za smanjivanje dimenzionalnosti jeste tzv. **analiza osnovnih komponenti** (en. **Principal Component Analysis**, PCA), koja koristi operacije linearne algebre nad matricama podataka da izračuna projekciju originalnih podataka na isti (ili poželjno manji) broj dimenzija.

Ova lekcija sadrži video materijal. Ukoliko želite da pogledate ovaj video morate da otvorite LAMS lekciju.

▼ Poglavlje 3

Linearna regresija

POJAM LINEARNE REGRESIJE

Linearna regresija hipotezu h mapira x na linearnu funkciju.

Linearna regresija predstavlja osnovni algoritam nadgledanog učenja.

Kod linearne regresije, na osnovu već postojećih ulaznih i izlaznih podataka, algoritam treba da predvidi izlaz za nove podatke koje su u opštem slučaju realne vrednosti.

Kao kod svih algoritama nadgledanog učenja, postoji skup podataka (en. **dataset**) koji se naziva **trening skup** (en. **training set**)

Unutar trening skupa postoji ukupno m trening primera, a svaki od primera sadrži x ulaznih promenljiva, i y izlaznih promenljiva.

Jedan trening primer predstavlja skup:

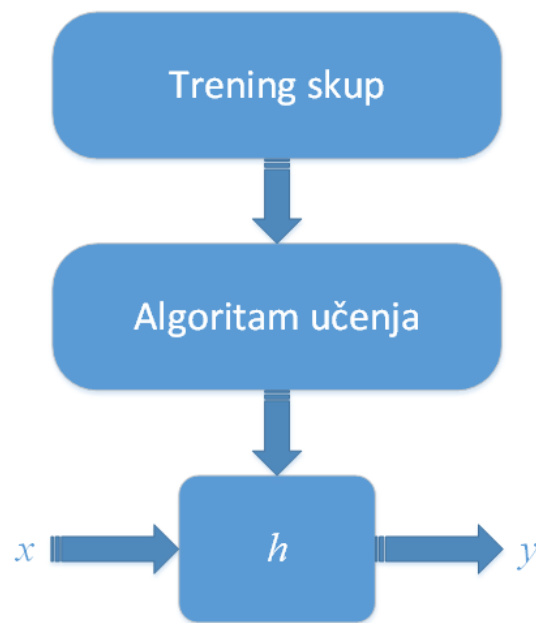
$$(x, y)$$

Trening primer na i -toj poziciji (i -ti indeks):

$$(x^{(i)}, y^{(i)})$$

Trening skup ubacujemo u algoritam mašinskog učenja. Zadatak algoritma učenja jeste da vrati funkciju h (hipotezu), koja predstavlja funkciju koja mapira ulazne promenljive x na **estimiranu (predviđenu) vrednost y** .

Funkcija h mapira x na y .



Slika 3.1 Proces nadgledanog učenja. [Izvor: Autor]

Za **linearnu regresiju**, funkcija h mapira x na linearnu funkciju:

$$h_{\theta}(x) = \theta_0 + \theta_1 x$$

Veličine θ_0 i θ_1 predstavljaju parametre hipoteze.

Kada \mathbf{x} ima samo jednu dimenziju, onda je u pitanju *linearna regresija sa jednom promenljivom* ili univarijabilna linearna regresija (en. *univariate linear regression*)

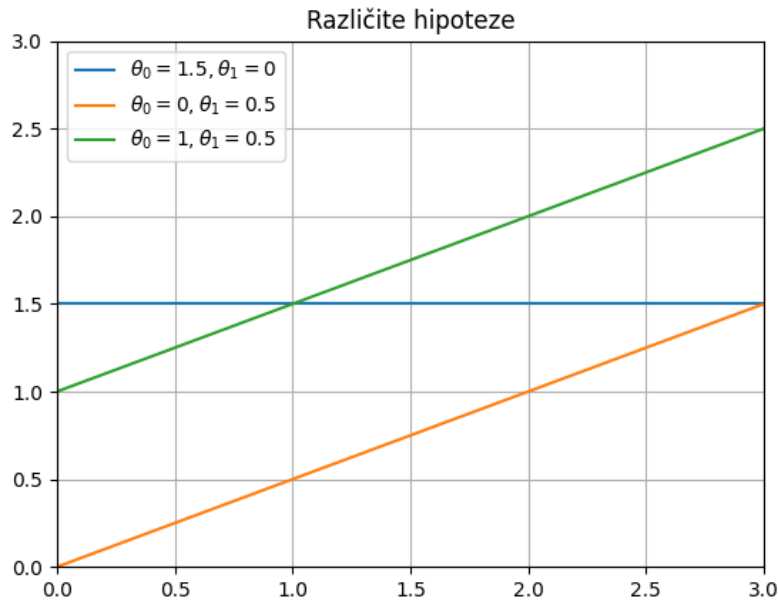
FUNKCIJA GREŠKE

Funkcija greške predstavlja kvadriranu razliku estimiranih i pravih vrednosti izlaznih promenljivih.

Data je hipoteza $\mathbf{h}(\mathbf{x})$:

$$h_{\theta}(x) = \theta_0 + \theta_1 x$$

Sa različitim vrednostima parametara θ_0 i θ_1 hipoteza će različito izgledati.



Slika 3.2 Različite vrednosti hipoteze za različite parametre θ . [Izvor: Autor]

U linearnoj regresiji cilj jeste pronaći vrednosti za parametre θ_0 i θ_1 koji će najbolje opisati trening set kroz linearnu funkciju. Ovo znači da estimirane vrednosti $h_{\theta}(\mathbf{x})$ treba da budu što bliže pravim vrednostima \mathbf{y} , za sve vrednosti trening skupa (\mathbf{x}, \mathbf{y}) .

Matematički gledano, treba minimizovati razliku između estimirane vrednosti i prave vrednosti, u odnosu na parametre θ .

$$\begin{aligned} & \underset{\theta_0, \theta_1}{\text{minimize}} \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 \\ & \underset{\theta_0, \theta_1}{\text{minimize}} \frac{1}{2m} \sum_{i=1}^m (\theta_0 + \theta_1 x^{(i)} - y^{(i)})^2 \end{aligned}$$

Definiše se funkcija greške (en. cost function) $J(\theta_0, \theta_1)$ koja predstavlja ovu razliku, i nju treba minimizovati. Ova funkcija zove se i funkcija srednjekvadratne greške (en. mean-squared error function).

$$\begin{aligned} J(\theta_0, \theta_1) &= \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 \\ & \underset{\theta_0, \theta_1}{\text{minimize}} J(\theta_0, \theta_1) \end{aligned}$$

ALGORITAM OPADAJUĆEG GRADIJENTA

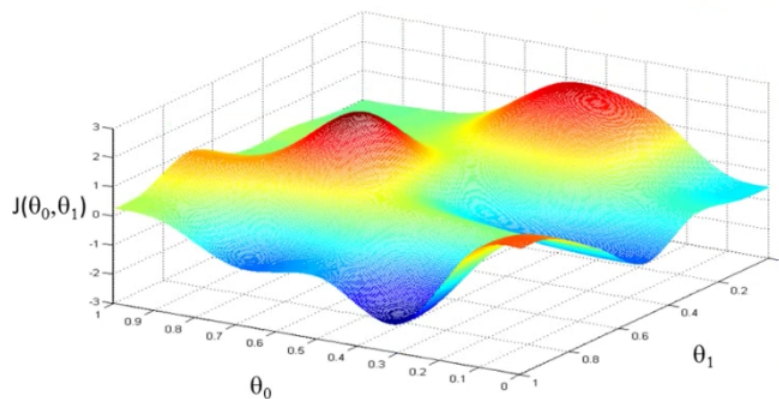
Najčešće korišćen algoritam za linearne modele mašinskog učenja jeste opadajući gradijent.

Jedan od najčešće korišćenih algoritama za minimiziranje funkcije greške jeste algoritam opadajućeg gradijenta (en. gradient descent algorithm).

Koraci algoritma jesu sledeći

- Poći sa nekim početnim vrednostima parametra θ_0 i θ_1 ,
- Menjati parametre da se smanji funkcija greške J dok se ne dođe do minimuma.

Na slici se mogu videti vrednosti funkcije J za različite vrednosti parametara. Algoritam iterativno radi dok ne dođe do minimuma funkcije.



Slika 3.3 Prikaz odnosa funkcije greške i parametara. [Izvor: coursera.org]

Matematički, algoritam opadajućeg gradijenta radi sledeće:

ponoviti do konvergencije :

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1), \quad \text{za } j = 0 \text{ i } j = 1$$

Parametar α predstavlja brzinu učenja, dok je drugi činilac izvod funkcije greške po θ_0 i θ_1 .

Potrebno je izračunati nove vrednosti za oba parametra, pa onda istovremeno izvršiti dodelu novih vrednosti.

$$temp0 := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$$

$$temp1 := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$$

$$\theta_0 := temp0$$

$$\theta_1 := temp1$$

PRIMENA ALGORITMA NA LINEARNU REGRESIJU

Računa se posebno izvod za prvi, a posebno za drugi parametar.

Algoritam opadajućeg gradijenta se primenjuje na linearnu regresiju sa jednom promenljivom:

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1), \quad \text{za } j = 0 \text{ i } j = 1$$

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$$

Računa se posebno izvod za prvi, a posebno za drugi parametar.

$$\begin{aligned} \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1) &= \frac{\partial}{\partial \theta_0} \times \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x) - y)^2 \\ &= \frac{\partial}{\partial \theta_0} \times \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x)^2 - 2h_{\theta}(x)y + y^2) \\ &= \frac{\partial}{\partial \theta_0} \times \frac{1}{2m} \sum_{i=1}^m ((\theta_0 + \theta_1 x)^2 - 2(\theta_0 + \theta_1 x)y + y^2) \\ &= \frac{\partial}{\partial \theta_0} \times \frac{1}{2m} \sum_{i=1}^m (\theta_0^2 + 2\theta_0\theta_1 x + \theta_1^2 x^2 - 2\theta_0 y - 2\theta_1 xy + y^2) \\ &= \frac{1}{2m} \sum_{i=1}^m (2\theta_0 + 2\theta_1 x - 2y) = \frac{1}{2m} \sum_{i=1}^m (2(\theta_0 + \theta_1 x - y)) \\ &= \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x) - y) \end{aligned}$$

$$\begin{aligned} \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1) &= \frac{\partial}{\partial \theta_1} \times \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x) - y)^2 \\ &= \frac{\partial}{\partial \theta_1} \times \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x)^2 - 2h_{\theta}(x)y + y^2) \\ &= \frac{\partial}{\partial \theta_1} \times \frac{1}{2m} \sum_{i=1}^m ((\theta_0 + \theta_1 x)^2 - 2(\theta_0 + \theta_1 x)y + y^2) \\ &= \frac{\partial}{\partial \theta_1} \times \frac{1}{2m} \sum_{i=1}^m (\theta_0^2 + 2\theta_0\theta_1 x + \theta_1^2 x^2 - 2\theta_0 y - 2\theta_1 xy + y^2) \\ &= \frac{1}{2m} \sum_{i=1}^m (2\theta_0 x + 2\theta_1 x^2 - 2xy) = \frac{1}{2m} \sum_{i=1}^m (2x(\theta_0 + \theta_1 x - y)) \\ &= \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x) - y)x \end{aligned}$$

IZVOĐENJE IZRAZA ZA LINEARNA REGRESIJA

U nastavku je video za izvođenje izraza za linearnu regresiju.

Ova lekcija sadrži video materijal. Ukoliko želite da pogledate ovaj video morate da otvorite LAMS lekciju.

▼ Poglavlje 4

Logistička regresija

POJAM LOGISTIČKE REGRESIJE

Logistička regresija predstavlja osnovni algoritam nadgledanog učenja za diskretne podatke.

Logistička regresija predstavlja osnovni algoritam nadgledanog učenja za diskretne podatke.

Kod logističke regresije, na osnovu postojećih ulaznih i izlaznih podataka, algoritam treba da predvidi izlaz za nove podatke koje su u opštem slučaju **diskretne vrednosti**, najčešće 0 i 1.

$$y \in \{0, 1\}$$

Kao kod svih algoritama nadgledanog učenja skup podataka se deli na trening i test skupove.

Za razliku od linearne regresije, gde estimirane vrednosti za y mogu bili realne vrednosti različite veličine, kod logističke regresije estimirane vrednosti treba da budu od 0 do 1

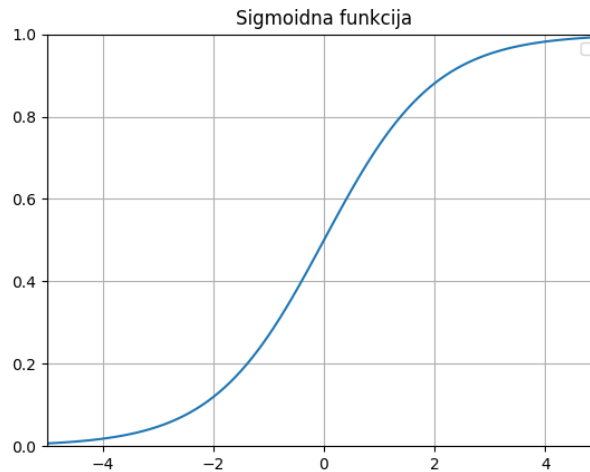
$$0 \leq h_{\theta}(x) \leq 1$$

Hipoteza kod logističke regresije je sledećeg oblika:

$$h_{\theta}(x) = g(\theta^T x)$$

U prethodnoj jednačini funkcija **$g(\cdot)$** je sigmoidna funkcija:

$$g(z) = \frac{1}{1 + e^{-z}}$$



Slika 4.1 Sigmoidna funkcija. [Izvor: Autor]

HIPOTEZA LOGISTIČKE REGRESIJE

Hipoteza koja je predstavljena sigmoidnom funkcijom na ovaj način predstavlja estimiranu verovatnoću da je $y=1$ za dati ulaz x .

Hipoteza koja je predstavljena sigmoidnom funkcijom na ovaj način predstavlja **estimiranu verovatnoću** da je **$y=1$** za dati ulaz **x** , koji je parametrizovan sa **θ** .

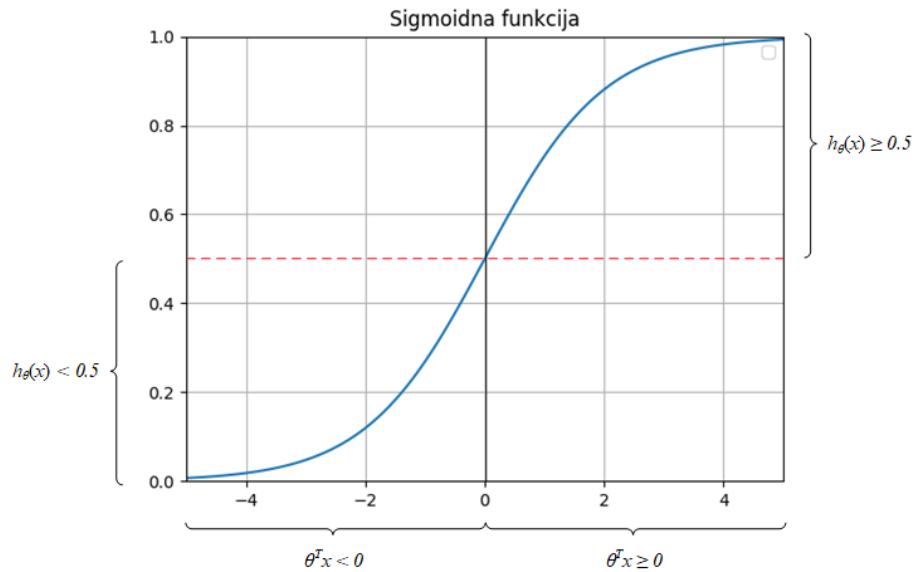
$$h_{\theta}(x) = P(y = 1|x; \theta)$$

Kako je zbir svih verovatnoća 1, onda se lako može izračunati verovatnoća da je **$y=0$** .

$$P(y = 0|x; \theta) + P(y = 1|x; \theta) = 1$$

$$P(y = 0|x; \theta) = 1 - P(y = 1|x; \theta)$$

Po izgledu sigmoidne funkcije može se zaključiti da kad je vrednost parametra (u ovom slučaju **$\theta^T x$**) veće od nule, onda je vrednost funkcije veća od 0.5.



Slika 4.2 Vrednosti granica za sigmoidnu funkciju. [Izvor: Autor]

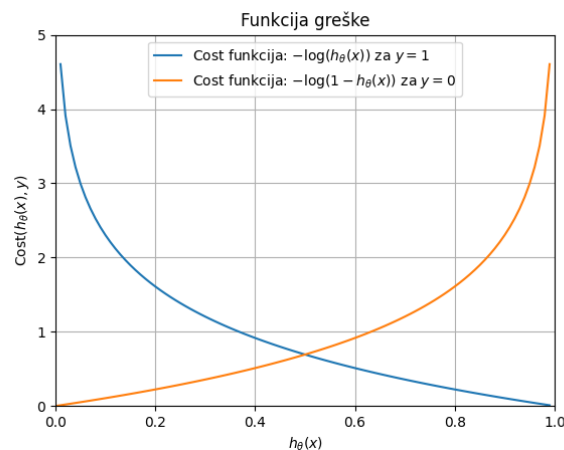
FUNKCIJA GREŠKE KOD LOGISTIČKE REGRESIJE

Kod logističke regresije, zbog nelinearne prirode sigmoidne funkcije, nije poželjno koristiti srednjekvadratnu grešku kao funkciju greške.

Kod logističke regresije, zbog nelinearne prirode sigmoidne funkcije, nije poželjno koristiti srednjekvadratnu grešku kao funkciju greške.

Umesto toga, funkcija greške koja se koristi u logističkoj regresiji jeste sledeća:

$$\text{Cost}(h_{\theta}(x), y) = \begin{cases} -\log(h_{\theta}(x)) & \text{za } y = 1 \\ -\log(1 - h_{\theta}(x)) & \text{za } y = 0 \end{cases}$$



Slika 4.3 Funkcija greške. [Izvor: Autor]

Potrebno je spojiti izraz za funkciju greške za $y=1$ i $y=0$ u jedinstven izraz.. Nakon toga, treba sabrati za sve vrednosti trening skupa, i podeliti sa brojem elemenata m da bi se izraz usrednjio.

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m y^{(i)} \log(h_{\theta}(x^{(i)})) + (1 - y^{(i)}) \log(1 - h_{\theta}(x^{(i)}))$$

ALGORITAM OPADAJUĆEG GRADIJENTA KOD LOGISTIČKE REGRESIJE

Kod logističke regresije koristi se sigmoidna funkcija.

Kao i kod linearne regresije, potrebno je minimizovati funkciju greške. Korišćenjem algoritma opadajućeg gradijenta dobija se sledeći izraz:

ponoviti do konvergencije :

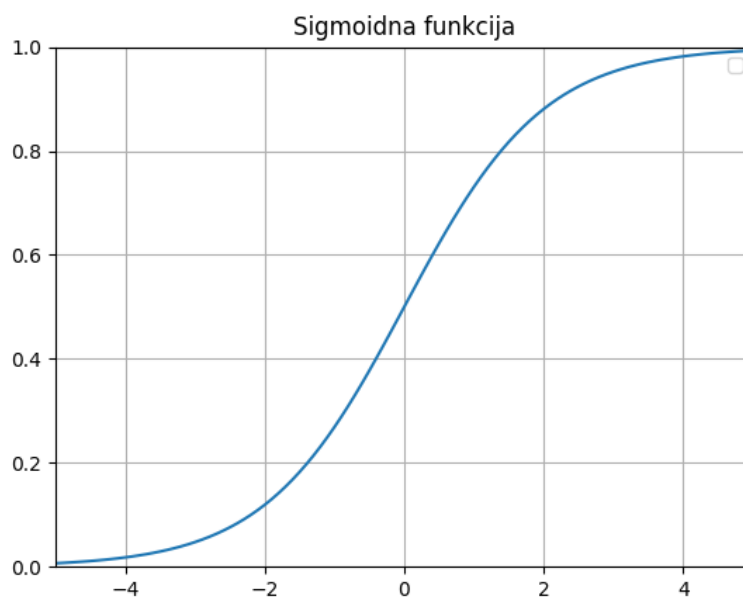
$$\theta_j := \theta_j - \alpha \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) x_j^{(i)}$$

Hipoteza kod logističke regresije je sledećeg oblika:

$$h_{\theta}(x) = g(\theta^T x)$$

U prethodnoj jednačini funkcija $g(\cdot)$ je sigmoidna funkcija:

$$g(z) = \frac{1}{1 + e^{-z}}$$



Slika 4.4 Sigmoidna funkcija. [Izvor: Autor]

▼ Poglavlje 5

Paket scikit-learn

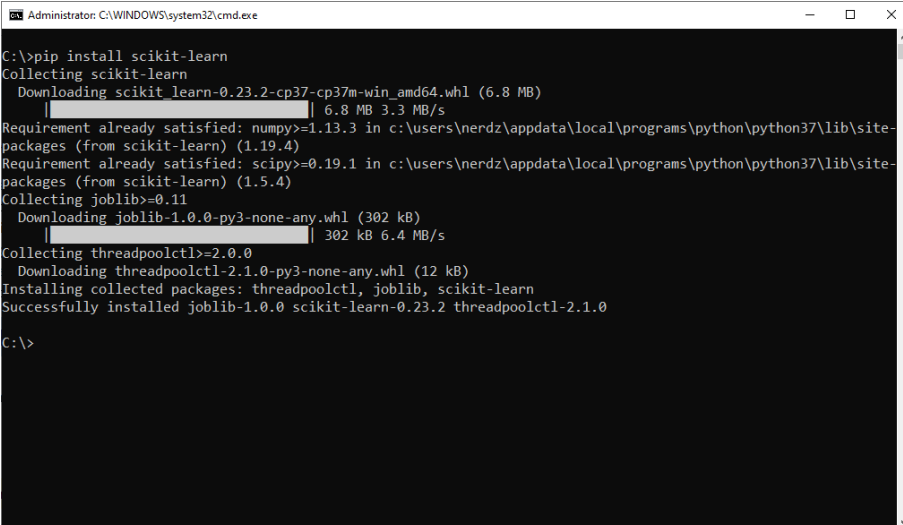
INSTALACIJA SCIKIT-LEARN PAKETA

Paket za Python jezik koji sadrži estimatore mašinskog učenja jeste scikit-learn.

Scikit-learn paket jeste paket koji sadrži različite algoritme za mašinsku učenje unutar Python programskog jezika.

Instalira se preko pip instalera:

```
pip install scikit-learn
```



```
Administrator: C:\WINDOWS\system32\cmd.exe
C:\>pip install scikit-learn
Collecting scikit-learn
  Downloading scikit_learn-0.23.2-cp37m-win_amd64.whl (6.8 MB)
    |#####| 6.8 MB 3.3 MB/s
Requirement already satisfied: numpy>=1.13.3 in c:\users\nerdz\appdata\local\programs\python\python37\lib\site-packages (from scikit-learn) (1.19.4)
Requirement already satisfied: scipy>=0.19.1 in c:\users\nerdz\appdata\local\programs\python\python37\lib\site-packages (from scikit-learn) (1.5.4)
Collecting joblib>=0.11
  Downloading joblib-1.0.0-py3-none-any.whl (302 kB)
    |#####| 302 kB 6.4 MB/s
Collecting threadpoolctl>=2.0.0
  Downloading threadpoolctl-2.1.0-py3-none-any.whl (12 kB)
Installing collected packages: threadpoolctl, joblib, scikit-learn
Successfully installed joblib-1.0.0 scikit-learn-0.23.2 threadpoolctl-2.1.0
C:\>
```

Slika 5.1 Instalacija scikit-learn paketa. [Izvor: Autor]

Scikit-learn u sebi sadrži više modula, od kojih je svaki modul jedan algoritam mašinskog učenja, ili estimator.

Scikit-learn paket se uvozi u Python radno okruženje naredbom:

```
# uvoz scikit-learn paketa
import sklearn
```

Scikit-learn radi sa numpy nizovima, te je neophodno uvesti i numpy paket pri pravljenju programa za mašinsko učenje.

LINEARNA REGRESIJA U SCIKIT-LEARN PAKETU

Linerana regresija je unutar modula `linear_models`

Algoritmi linearne i logističke regresije spadaju u linearne modele, te je neophodno uvesti odgovarajuću klasu iz modula **`linear_model`**.

```
from sklearn import linear_model
```

Linearna regresija

Kada se učitava skup podataka (sa ulazima i izlazima), najbolje je da se deo skupa podeli na trening skup, a ostatak na test skup, koji će testirati model.

Po konvenciji, ovi skupovi se nazivaju **`X_train`**, **`X_test`**, **`y_train`** i **`y_test`**.

Nakon podele skupova, pravi se objekat klase **`LinearRegression()`**

```
regr = linear_model.LinearRegression()
```

Pozivom metode **`.fit()`** treba ubaciti test parametre (**`X_train`** i **`y_train`**)

```
regr.fit(X_train,y_train)
```

Sada je model istreniran, i njegova tačnost se može izračunati pozivom metode **`.score()`**, a kao ulazni i izlazni podaci mogu se staviti **`X_test`** i **`y_test`**.

```
accuracy = regr.score(X_test, y_test)
```

Estimirane vrednosti se mogu dobiti pozivom metode **`.predict()`** koje za ulazni parametar uzima **`X_test`** skup.

```
y_pred = regr.predict(X_test)
```

Model se može ponovo istrenirati, ponovnim pozivom metode **`.fit()`** ukoliko se smatra da je nedovoljna tačnost. Rad na istom skupu podataka obično će dati istu ili sličnu tačnost.

Parametri **θ_0** i **θ_1** se respektivno mogu dobiti pozivanjem atributa **`intercept_`** i **`coef_`**

```
# stampanje parametara  
print(regr.coef_)  
print(regr.intercept_)
```

Kod univarijatne linearne regresije, moguće je vizuelno i predstaviti odstupanje estimiranih (**`y_pred`**) vrednosti sa pravim (**`y_test`**) pozivom **`scatter`** i **`plot`** funkcija iz paketa **`matplotlib`**.

```
# graficki prikaz  
plt.scatter(X_test, y_test)  
plt.plot(X_test, y_pred, color='red', label='Linearna regresija')
```



```
plt.grid()  
plt.show()
```

LOGISTIČKA REGRESIJA U SCIKIT-LEARN PAKETU

Logistička regresija je unutar modula `linear_models`

Algoritmi linearne i logističke regresije spadaju u linearne modele, te je neophodno uvesti odgovarajuću klasu iz modula *linear_model*.

```
from sklearn import linear_model
```

Linearna regresija

Kada se učitava skup podataka (sa ulazima i izlazima), najbolje je da se deo skupa podeli na trening skup, a ostatak na test skup, koji će testirati model.

Po konvenciji, ovi skupovi se nazivaju *X_train*, *X_test*, *y_train* i *y_test*.

Nakon podele skupova, pravi se objekat klase *LogisticRegression()*

```
regr = linear_model.LogisticRegression()
```

Pozivom metode *.fit()* treba ubaciti test parametre (*X_train* i *y_train*)

```
regr.fit(X_train,y_train)
```

Sada je model istreniran, i njegova tačnost se može izračunati pozivom metode *.score()*, a kao ulazni i izlazni podaci mogu se staviti *X_test* i *y_test*.

```
accuracy = regr.score(X_test, y_test)
```

Estimirane vrednosti se mogu dobiti pozivom metode *.predict()* koje za ulazni parametar uzima *X_test* skup.

```
y_pred = regr.predict(X_test)
```

Model se može ponovo istrenirati, ponovnim pozivom metode *.fit()* ukoliko se smatra da je nedovoljna tačnost. Rad na istom skupu podataka obično će dati istu ili sličnu tačnost.

Parametri θ_0 i θ_1 se respektivno mogu dobiti pozivanjem atributa *intercept_* i *coef_*

```
# stampanje parametara  
print(regr.coef_)  
print(regr.intercept_)
```

Kod logističke regresije moguće je izbrojati greške preko *numpy* nizova.

```
# brojanje gresaka  
errors = np.sum(np.where(y_test != y_pred, 1, 0))
```

▼ Poglavlje 6

Pokazne vežbe #12

ZADATAK #1

Zadatak #1 se odnosi na linearnu regresiju i radi se okvirno 25 minuta

Zadatak #1 (25 minuta)

Iz datoteke **student-mat.csv** učitati kolone G1 i G3. Koristiti kolonu G1 kao skup ulaznih vrednosti, a kolonu G3 kao skup izlaznih vrednosti.

Istrenirati model linearne regresije koja uzima 90% skupa za trening skup, a ostalo za test skup.

Štampati koeficijente, tačnost, i vizuelno pokazati linearni model estimacije.

Pokušati sa više ulaznih parametara.

```
import numpy as np
import pandas as pd
from matplotlib import pyplot as plt
import sklearn
from sklearn import linear_model

df = pd.read_csv('student-mat.csv', sep=' ')

# stampanje samo imena kolona
# print(df.columns)

# stampanje imena kolona i prvih 5 redova
# print(df.head())

# biramo ono sto zelimo da pratimo
data_list = ['G1', 'G3']

data = df[data_list]

# izvlacimo promenljivu koju zelimo da predvidjamo
predict = 'G3'

# dftmp = df.drop([predict], 1)
# print(dftmp)

# novi niz bez onoga sto predvidjamo
x = np.array(data.drop([predict], 1))
```

```
# novi niz samo sa onim sto predvidjamo
y = np.array(data[predict])

# delimo skup u train i test skupove
x_train, x_test, y_train, y_test = sklearn.model_selection.train_test_split(x, y,
test_size=0.1)

print(len(x_train), len(x_test))

# pravimo objekat klase LinearRegression
linear = linear_model.LinearRegression()

# fit-ujemo prema trening podacima
linear.fit(x_train, y_train)

# predikcija na test podacima

predictions = linear.predict(x_test)

new_df = pd.DataFrame(x_test, columns=['G1'])
new_df['Y'] = y_test
new_df['Predikcija'] = predictions

print(new_df)

# racunamo tacnost, poredimo sa test podacima
accuracy = linear.score(x_test, y_test)
print(accuracy)

# stampanje koeficijenata
print(linear.coef_)
print(linear.intercept_)

# Plot outputs
plt.scatter(new_df['G1'], y_test, color='blue')
plt.plot(new_df['G1'], predictions, color='red')
plt.show()
```

```
school sex age address famsize Pstatus Medu Fedu Mjob Fjob reason guardian
traveltime studytime failures schoolsup famsup paid activities nursery higher
internet romantic famrel freetime goout Dalc Walc health absences G1 G2 G3
"GP" "F" 18 "U" "GT3" "A" 4 4 "at_home" "teacher" "course" "mother" 2 2 0 "yes"
"no" "no" "no" "yes" "yes" "no" "no" 4 3 4 1 1 3 6 "5" "6" 6
"GP" "F" 17 "U" "GT3" "T" 1 1 "at_home" "other" "course" "father" 1 2 0 "no" "yes"
"no" "no" "no" "yes" "yes" "no" 5 3 3 1 1 3 4 "5" "5" 6
"GP" "F" 15 "U" "LE3" "T" 1 1 "at_home" "other" "other" "mother" 1 2 3 "yes" "no"
"yes" "no" "yes" "yes" "yes" "no" 4 3 2 2 3 3 10 "7" "8" 10
"GP" "F" 15 "U" "GT3" "T" 4 2 "health" "services" "home" "mother" 1 3 0 "no" "yes"
"yes" "yes" "yes" "yes" "yes" "yes" 3 2 2 1 1 5 2 "15" "14" 15
"GP" "F" 16 "U" "GT3" "T" 3 3 "other" "other" "home" "father" 1 2 0 "no" "yes"
"yes" "no" "yes" "yes" "no" "no" 4 3 2 1 2 5 4 "6" "10" 10
"GP" "M" 16 "U" "LE3" "T" 4 3 "services" "other" "reputation" "mother" 1 2 0 "no"
```

```

"yes" "yes" "yes" "yes" "yes" "yes" "no" 5 4 2 1 2 5 10 "15" "15" 15
"GP" "M" 16 "U" "LE3" "T" 2 2 "other" "other" "home" "mother" 1 2 0 "no" "no" "no"
"no" "yes" "yes" "yes" "no" 4 4 4 1 1 3 0 "12" "12" 11
"GP" "F" 17 "U" "GT3" "A" 4 4 "other" "teacher" "home" "mother" 2 2 0 "yes" "yes"
"no" "no" "yes" "yes" "no" "no" 4 1 4 1 1 1 6 "6" "5" 6
"GP" "M" 15 "U" "LE3" "A" 3 2 "services" "other" "home" "mother" 1 2 0 "no" "yes"
"yes" "no" "yes" "yes" "yes" "no" 4 2 2 1 1 1 0 "16" "18" 19
"GP" "M" 15 "U" "GT3" "T" 3 4 "other" "other" "home" "mother" 1 2 0 "no" "yes"
"yes" "yes" "yes" "yes" "yes" "no" 5 5 1 1 1 5 0 "14" "15" 15
"GP" "F" 15 "U" "GT3" "T" 4 4 "teacher" "health" "reputation" "mother" 1 2 0 "no"
"yes" "yes" "no" "yes" "yes" "yes" "no" 3 3 3 1 2 2 0 "10" "8" 9
"GP" "F" 15 "U" "GT3" "T" 2 1 "services" "other" "reputation" "father" 3 3 0 "no"
"yes" "no" "yes" "yes" "yes" "yes" "no" 5 2 2 1 1 4 4 "10" "12" 12
"GP" "M" 15 "U" "LE3" "T" 4 4 "health" "services" "course" "father" 1 1 0 "no"
"yes" "yes" "yes" "yes" "yes" "yes" "no" 4 3 3 1 3 5 2 "14" "14" 14
"GP" "M" 15 "U" "GT3" "T" 4 3 "teacher" "other" "course" "mother" 2 2 0 "no" "yes"
"yes" "no" "yes" "yes" "yes" "no" 5 4 3 1 2 3 2 "10" "10" 11
"GP" "M" 15 "U" "GT3" "A" 2 2 "other" "other" "home" "other" 1 3 0 "no" "yes" "no"
"no" "yes" "yes" "yes" "yes" 4 5 2 1 1 3 0 "14" "16" 16
"GP" "F" 16 "U" "GT3" "T" 4 4 "health" "other" "home" "mother" 1 1 0 "no" "yes"
"no" "no" "yes" "yes" "yes" "no" 4 4 4 1 2 2 4 "14" "14" 14
"GP" "F" 16 "U" "GT3" "T" 4 4 "services" "services" "reputation" "mother" 1 3 0
"no" "yes" "yes" "yes" "yes" "yes" "yes" "no" 3 2 3 1 2 2 6 "13" "14" 14
"GP" "F" 16 "U" "GT3" "T" 3 3 "other" "other" "reputation" "mother" 3 2 0 "yes"
"yes" "no" "yes" "yes" "yes" "no" "no" 5 3 2 1 1 4 4 "8" "10" 10
"GP" "M" 17 "U" "GT3" "T" 3 2 "services" "services" "course" "mother" 1 1 3 "no"
"yes" "no" "yes" "yes" "yes" "yes" "no" 5 5 5 2 4 5 16 "6" "5" 5
"GP" "M" 16 "U" "LE3" "T" 4 3 "health" "other" "home" "father" 1 1 0 "no" "no"
"yes" "yes" "yes" "yes" "yes" "no" 3 1 3 1 3 5 4 "8" "10" 10
"GP" "M" 15 "U" "GT3" "T" 4 3 "teacher" "other" "reputation" "mother" 1 2 0 "no"
"no" "no" "no" "yes" "yes" "yes" "no" 4 4 1 1 1 1 0 "13" "14" 15
"GP" "M" 15 "U" "GT3" "T" 4 4 "health" "health" "other" "father" 1 1 0 "no" "yes"
"yes" "no" "yes" "yes" "yes" "no" 5 4 2 1 1 5 0 "12" "15" 15
"GP" "M" 16 "U" "LE3" "T" 4 2 "teacher" "other" "course" "mother" 1 2 0 "no" "no"
"no" "yes" "yes" "yes" "yes" "no" 4 5 1 1 3 5 2 "15" "15" 16
"GP" "M" 16 "U" "LE3" "T" 2 2 "other" "other" "reputation" "mother" 2 2 0 "no"
"yes" "no" "yes" "yes" "yes" "yes" "no" 5 4 4 2 4 5 0 "13" "13" 12
"GP" "F" 15 "R" "GT3" "T" 2 4 "services" "health" "course" "mother" 1 3 0 "yes"
"yes" "yes" "yes" "yes" "yes" "yes" "no" 4 3 2 1 1 5 2 "10" "9" 8
"GP" "F" 16 "U" "GT3" "T" 2 2 "services" "services" "home" "mother" 1 1 2 "no"
"yes" "yes" "no" "no" "yes" "yes" "no" 1 2 2 1 3 5 14 "6" "9" 8
"GP" "M" 15 "U" "GT3" "T" 2 2 "other" "other" "home" "mother" 1 1 0 "no" "yes"
"yes" "no" "yes" "yes" "yes" "no" 4 2 2 1 2 5 2 "12" "12" 11
"GP" "M" 15 "U" "GT3" "T" 4 2 "health" "services" "other" "mother" 1 1 0 "no" "no"
"yes" "no" "yes" "yes" "yes" "no" 2 2 4 2 4 1 4 "15" "16" 15
"GP" "M" 16 "U" "LE3" "A" 3 4 "services" "other" "home" "mother" 1 2 0 "yes" "yes"
"no" "yes" "yes" "yes" "yes" "no" 5 3 3 1 1 5 4 "11" "11" 11
"GP" "M" 16 "U" "GT3" "T" 4 4 "teacher" "teacher" "home" "mother" 1 2 0 "no" "yes"
"yes" "yes" "yes" "yes" "yes" "yes" 4 4 5 5 5 5 16 "10" "12" 11
"GP" "M" 15 "U" "GT3" "T" 4 4 "health" "services" "home" "mother" 1 2 0 "no" "yes"
"yes" "no" "no" "yes" "yes" "no" 5 4 2 3 4 5 0 "9" "11" 12
"GP" "M" 15 "U" "GT3" "T" 4 4 "services" "services" "reputation" "mother" 2 2 0
"no" "yes" "no" "yes" "yes" "yes" "yes" "no" 4 3 1 1 1 5 0 "17" "16" 17

```

```
"GP" "M" 15 "R" "GT3" "T" 4 3 "teacher" "at_home" "course" "mother" 1 2 0 "no"
"yes" "no" "yes" "yes" "yes" "yes" "yes" 4 5 2 1 1 5 0 "17" "16" 16
"GP" "M" 15 "U" "LE3" "T" 3 3 "other" "other" "course" "mother" 1 2 0 "no" "no"
"no" "yes" "no" "yes" "yes" "no" 5 3 2 1 1 2 0 "8" "10" 12
"GP" "M" 16 "U" "GT3" "T" 3 2 "other" "other" "home" "mother" 1 1 0 "no" "yes"
"yes" "no" "no" "yes" "yes" "no" 5 4 3 1 1 5 0 "12" "14" 15
"GP" "F" 15 "U" "GT3" "T" 2 3 "other" "other" "other" "father" 2 1 0 "no" "yes"
"no" "yes" "yes" "yes" "no" "no" 3 5 1 1 1 5 0 "8" "7" 6
"GP" "M" 15 "U" "LE3" "T" 4 3 "teacher" "services" "home" "mother" 1 3 0 "no" "yes"
"no" "yes" "yes" "yes" "yes" "no" 5 4 3 1 1 4 2 "15" "16" 18
"GP" "M" 16 "R" "GT3" "A" 4 4 "other" "teacher" "reputation" "mother" 2 3 0 "no"
"yes" "no" "yes" "yes" "yes" "yes" "yes" 2 4 3 1 1 5 7 "15" "16" 15
"GP" "F" 15 "R" "GT3" "T" 3 4 "services" "health" "course" "mother" 1 3 0 "yes"
"yes" "yes" "yes" "yes" "yes" "yes" "no" 4 3 2 1 1 5 2 "12" "12" 11
"GP" "F" 15 "R" "GT3" "T" 2 2 "at_home" "other" "reputation" "mother" 1 1 0 "yes"
"yes" "yes" "yes" "yes" "yes" "no" "no" 4 3 1 1 1 2 8 "14" "13" 13
"GP" "F" 16 "U" "LE3" "T" 2 2 "other" "other" "home" "mother" 2 2 1 "no" "yes" "no"
"yes" "no" "yes" "yes" "yes" 3 3 3 1 2 3 25 "7" "10" 11
"GP" "M" 15 "U" "LE3" "T" 4 4 "teacher" "other" "home" "other" 1 1 0 "no" "yes"
"no" "no" "no" "yes" "yes" "yes" 5 4 3 2 4 5 8 "12" "12" 12
"GP" "M" 15 "U" "GT3" "T" 4 4 "services" "teacher" "course" "father" 1 2 0 "no"
"yes" "no" "yes" "yes" "yes" "yes" "no" 4 3 3 1 1 5 2 "19" "18" 18
"GP" "M" 15 "U" "GT3" "T" 2 2 "services" "services" "course" "father" 1 1 0 "yes"
"yes" "no" "no" "yes" "yes" "yes" "no" 5 4 1 1 1 1 0 "8" "8" 11
"GP" "F" 16 "U" "LE3" "T" 2 2 "other" "at_home" "course" "father" 2 2 1 "yes" "no"
"no" "yes" "yes" "yes" "yes" "no" 4 3 3 2 2 5 14 "10" "10" 9
"GP" "F" 15 "U" "LE3" "A" 4 3 "other" "other" "course" "mother" 1 2 0 "yes" "yes"
"yes" "yes" "yes" "yes" "yes" "yes" 5 2 2 1 1 5 8 "8" "8" 6
"GP" "F" 16 "U" "LE3" "A" 3 3 "other" "services" "home" "mother" 1 2 0 "no" "yes"
"no" "no" "yes" "yes" "yes" "no" 2 3 5 1 4 3 12 "11" "12" 11
"GP" "M" 16 "U" "GT3" "T" 4 3 "health" "services" "reputation" "mother" 1 4 0 "no"
"no" "no" "yes" "yes" "yes" "yes" "no" 4 2 2 1 1 2 4 "19" "19" 20
"GP" "M" 15 "U" "GT3" "T" 4 2 "teacher" "other" "home" "mother" 1 2 0 "no" "yes"
"yes" "no" "yes" "yes" "no" "no" 4 3 3 2 2 5 2 "15" "15" 14
"GP" "F" 15 "U" "GT3" "T" 4 4 "services" "teacher" "other" "father" 1 2 1 "yes"
"yes" "no" "yes" "no" "yes" "yes" "no" 4 4 4 1 1 3 2 "7" "7" 7
"GP" "F" 16 "U" "LE3" "T" 2 2 "services" "services" "course" "mother" 3 2 0 "no"
"yes" "yes" "no" "yes" "yes" "yes" "no" 4 3 3 2 3 4 2 "12" "13" 13
"GP" "F" 15 "U" "LE3" "T" 4 2 "health" "other" "other" "mother" 1 2 0 "no" "yes"
"yes" "no" "yes" "yes" "yes" "no" 4 3 3 1 1 5 2 "11" "13" 13
"GP" "M" 15 "U" "LE3" "A" 4 2 "health" "health" "other" "father" 2 1 1 "no" "no"
"no" "no" "yes" "yes" "no" "no" 5 5 5 3 4 5 6 "11" "11" 10
"GP" "F" 15 "U" "GT3" "T" 4 4 "services" "services" "course" "mother" 1 1 0 "yes"
"yes" "yes" "no" "yes" "yes" "yes" "no" 3 3 4 2 3 5 0 "8" "10" 11
"GP" "F" 15 "U" "LE3" "A" 3 3 "other" "other" "other" "mother" 1 1 0 "no" "no"
"yes" "no" "yes" "yes" "yes" "no" 5 3 4 4 4 1 6 "10" "13" 13
"GP" "F" 16 "U" "GT3" "A" 2 1 "other" "other" "other" "mother" 1 2 0 "no" "no"
"yes" "yes" "yes" "yes" "yes" "yes" 5 3 4 1 1 2 8 "8" "9" 10
"GP" "F" 15 "U" "GT3" "A" 4 3 "services" "services" "reputation" "mother" 1 2 0
"no" "yes" "yes" "yes" "yes" "yes" "yes" "no" 4 3 2 1 1 1 0 "14" "15" 15
"GP" "M" 15 "U" "GT3" "T" 4 4 "teacher" "health" "reputation" "mother" 1 2 0 "no"
"yes" "no" "yes" "yes" "yes" "no" "no" 3 2 2 1 1 5 4 "14" "15" 15
"GP" "M" 15 "U" "LE3" "T" 1 2 "other" "at_home" "home" "father" 1 2 0 "yes" "yes"
```

```

"no" "yes" "yes" "yes" "yes" "no" 4 3 2 1 1 5 2 "9" "10" 9
"GP" "F" 16 "U" "GT3" "T" 4 2 "services" "other" "course" "mother" 1 2 0 "no" "yes"
"no" "no" "yes" "yes" "yes" "no" 4 2 3 1 1 5 2 "15" "16" 16
"GP" "F" 16 "R" "GT3" "T" 4 4 "health" "teacher" "other" "mother" 1 2 0 "no" "yes"
"no" "yes" "yes" "yes" "no" "no" 2 4 4 2 3 4 6 "10" "11" 11
"GP" "F" 16 "U" "GT3" "T" 1 1 "services" "services" "course" "father" 4 1 0 "yes"
"yes" "no" "yes" "no" "yes" "yes" "yes" 5 5 5 5 5 6 "10" "8" 11
"GP" "F" 16 "U" "LE3" "T" 1 2 "other" "services" "reputation" "father" 1 2 0 "yes"
"no" "no" "yes" "yes" "yes" "yes" "no" 4 4 3 1 1 1 4 "8" "10" 9
"GP" "F" 16 "U" "GT3" "T" 4 3 "teacher" "health" "home" "mother" 1 3 0 "yes" "yes"
"yes" "yes" "yes" "yes" "yes" "no" 3 4 4 2 4 4 2 "10" "9" 9
"GP" "F" 15 "U" "LE3" "T" 4 3 "services" "services" "reputation" "father" 1 2 0
"yes" "no" "no" "yes" "yes" "yes" "yes" "yes" 4 4 4 2 4 2 0 "10" "10" 10
"GP" "F" 16 "U" "LE3" "T" 4 3 "teacher" "services" "course" "mother" 3 2 0 "no"
"yes" "no" "yes" "yes" "yes" "yes" "no" 5 4 3 1 2 1 2 "16" "15" 15
"GP" "M" 15 "U" "GT3" "A" 4 4 "other" "services" "reputation" "mother" 1 4 0 "no"
"yes" "no" "yes" "no" "yes" "yes" "yes" 1 3 3 5 5 3 4 "13" "13" 12
"GP" "F" 16 "U" "GT3" "T" 3 1 "services" "other" "course" "mother" 1 4 0 "yes"
"yes" "yes" "no" "yes" "yes" "yes" "no" 4 3 3 1 2 5 4 "7" "7" 6
"GP" "F" 15 "R" "LE3" "T" 2 2 "health" "services" "reputation" "mother" 2 2 0 "yes"
"yes" "yes" "no" "yes" "yes" "yes" "no" 4 1 3 1 3 4 2 "8" "9" 8
"GP" "F" 15 "R" "LE3" "T" 3 1 "other" "other" "reputation" "father" 2 4 0 "no"
"yes" "no" "no" "no" "yes" "yes" "no" 4 4 2 2 3 3 12 "16" "16" 16
"GP" "M" 16 "U" "GT3" "T" 3 1 "other" "other" "reputation" "father" 2 4 0 "no"
"yes" "yes" "no" "yes" "yes" "yes" "no" 4 3 2 1 1 5 0 "13" "15" 15
"GP" "M" 15 "U" "GT3" "T" 4 2 "other" "other" "course" "mother" 1 4 0 "no" "no"
"no" "no" "yes" "yes" "yes" "no" 3 3 3 1 1 3 0 "10" "10" 10
"GP" "F" 15 "R" "GT3" "T" 1 1 "other" "other" "reputation" "mother" 1 2 2 "yes"
"yes" "no" "no" "no" "yes" "yes" "yes" 3 3 4 2 4 5 2 "8" "6" 5
"GP" "M" 16 "U" "GT3" "T" 3 1 "other" "other" "reputation" "mother" 1 1 0 "no" "no"
"no" "yes" "yes" "yes" "no" "no" 5 3 2 2 2 5 2 "12" "12" 14
"GP" "F" 16 "U" "GT3" "T" 3 3 "other" "services" "home" "mother" 1 2 0 "yes" "yes"
"yes" "yes" "yes" "yes" "yes" "no" 4 3 3 2 4 5 54 "11" "12" 11
"GP" "M" 15 "U" "GT3" "T" 4 3 "teacher" "other" "home" "mother" 1 2 0 "no" "yes"
"yes" "yes" "yes" "yes" "yes" "no" 4 3 3 2 3 5 6 "9" "9" 10
"GP" "M" 15 "U" "GT3" "T" 4 0 "teacher" "other" "course" "mother" 2 4 0 "no" "no"
"no" "yes" "yes" "yes" "yes" "no" 3 4 3 1 1 1 8 "11" "11" 10
"GP" "F" 16 "U" "GT3" "T" 2 2 "other" "other" "reputation" "mother" 1 4 0 "no" "no"
"yes" "no" "yes" "yes" "yes" "yes" 5 2 3 1 3 3 0 "11" "11" 11
"GP" "M" 17 "U" "GT3" "T" 2 1 "other" "other" "home" "mother" 2 1 3 "yes" "yes"
"no" "yes" "yes" "no" "yes" "no" 4 5 1 1 1 3 2 "8" "8" 10
"GP" "F" 16 "U" "GT3" "T" 3 4 "at_home" "other" "course" "mother" 1 2 0 "no" "yes"
"no" "no" "yes" "yes" "yes" "no" 2 4 3 1 2 3 12 "5" "5" 5
"GP" "M" 15 "U" "GT3" "T" 2 3 "other" "services" "course" "father" 1 1 0 "yes"
"yes" "yes" "yes" "no" "yes" "yes" "yes" 3 2 2 1 3 3 2 "10" "12" 12
"GP" "M" 15 "U" "GT3" "T" 2 3 "other" "other" "home" "mother" 1 3 0 "yes" "no"
"yes" "no" "no" "yes" "yes" "no" 5 3 2 1 2 5 4 "11" "10" 11
"GP" "F" 15 "U" "LE3" "T" 3 2 "services" "other" "reputation" "mother" 1 2 0 "no"
"yes" "yes" "no" "yes" "yes" "yes" "no" 4 4 4 1 1 5 10 "7" "6" 6
"GP" "M" 15 "U" "LE3" "T" 2 2 "services" "services" "home" "mother" 2 2 0 "no" "no"
"yes" "yes" "yes" "yes" "yes" "no" 5 3 3 1 3 4 4 "15" "15" 15
"GP" "F" 15 "U" "GT3" "T" 1 1 "other" "other" "home" "father" 1 2 0 "no" "yes" "no"
"yes" "no" "yes" "yes" "no" 4 3 2 2 3 4 2 "9" "10" 10

```

```

"GP" "F" 15 "U" "GT3" "T" 4 4 "services" "services" "reputation" "father" 2 2 2
"no" "no" "yes" "no" "yes" "yes" "yes" "yes" 4 4 4 2 3 5 6 "7" "9" 8
"GP" "F" 16 "U" "LE3" "T" 2 2 "at_home" "other" "course" "mother" 1 2 0 "no" "yes"
"no" "no" "yes" "yes" "no" "no" 4 3 4 1 2 2 4 "8" "7" 6
"GP" "F" 15 "U" "GT3" "T" 4 2 "other" "other" "reputation" "mother" 1 3 0 "no"
"yes" "no" "yes" "yes" "yes" "yes" "no" 5 3 3 1 3 1 4 "13" "14" 14
"GP" "M" 16 "U" "GT3" "T" 2 2 "services" "other" "reputation" "father" 2 2 1 "no"
"no" "yes" "yes" "no" "yes" "yes" "no" 4 4 2 1 1 3 12 "11" "10" 10
"GP" "M" 16 "U" "LE3" "A" 4 4 "teacher" "health" "reputation" "mother" 1 2 0 "no"
"yes" "no" "no" "yes" "yes" "no" "no" 4 1 3 3 5 5 18 "8" "6" 7
"GP" "F" 16 "U" "GT3" "T" 3 3 "other" "other" "home" "mother" 1 3 0 "no" "yes"
"yes" "no" "yes" "yes" "yes" "yes" 4 3 3 1 3 4 0 "7" "7" 8
"GP" "F" 15 "U" "GT3" "T" 4 3 "services" "other" "reputation" "mother" 1 1 0 "no"
"no" "yes" "yes" "yes" "yes" "yes" "no" 4 5 5 1 3 1 4 "16" "17" 18
"GP" "F" 16 "U" "LE3" "T" 3 1 "other" "other" "home" "father" 1 2 0 "yes" "yes"
"no" "no" "yes" "yes" "no" "no" 3 3 3 2 3 2 4 "7" "6" 6
"GP" "F" 16 "U" "GT3" "T" 4 2 "teacher" "services" "home" "mother" 2 2 0 "no" "yes"
"yes" "yes" "yes" "yes" "yes" "no" 5 3 3 1 1 1 0 "11" "10" 10
"GP" "M" 15 "U" "LE3" "T" 2 2 "services" "health" "reputation" "mother" 1 4 0 "no"
"yes" "no" "yes" "yes" "yes" "yes" "no" 4 3 4 1 1 4 6 "11" "13" 14
"GP" "F" 15 "R" "GT3" "T" 1 1 "at_home" "other" "home" "mother" 2 4 1 "yes" "yes"
"yes" "yes" "yes" "yes" "yes" "no" 3 1 2 1 1 1 2 "7" "10" 10
"GP" "M" 16 "R" "GT3" "T" 4 3 "services" "other" "reputation" "mother" 2 1 0 "yes"
"yes" "no" "yes" "no" "yes" "yes" "no" 3 3 3 1 1 4 2 "11" "15" 15
"GP" "F" 16 "U" "GT3" "T" 2 1 "other" "other" "course" "mother" 1 2 0 "no" "yes"
"yes" "no" "yes" "yes" "no" "yes" 4 3 5 1 1 5 2 "8" "9" 10
"GP" "F" 16 "U" "GT3" "T" 4 4 "other" "other" "reputation" "mother" 1 1 0 "no" "no"
"no" "yes" "no" "yes" "yes" "no" 5 3 4 1 2 1 6 "11" "14" 14
"GP" "F" 16 "U" "GT3" "T" 4 3 "other" "at_home" "course" "mother" 1 3 0 "yes" "yes"
"yes" "no" "yes" "yes" "yes" "no" 5 3 5 1 1 3 0 "7" "9" 8
"GP" "M" 16 "U" "GT3" "T" 4 4 "services" "services" "other" "mother" 1 1 0 "yes"
"yes" "yes" "yes" "yes" "yes" "yes" "no" 4 5 5 5 5 4 14 "7" "7" 5
"GP" "M" 16 "U" "GT3" "T" 4 4 "services" "teacher" "other" "father" 1 3 0 "no"
"yes" "no" "yes" "yes" "yes" "yes" "yes" 4 4 3 1 1 4 0 "16" "17" 17
"GP" "M" 15 "U" "GT3" "T" 4 4 "services" "other" "course" "mother" 1 1 0 "no" "yes"
"no" "yes" "no" "yes" "yes" "no" 5 3 3 1 1 5 4 "10" "13" 14
"GP" "F" 15 "U" "GT3" "T" 3 2 "services" "other" "home" "mother" 2 2 0 "yes" "yes"
"yes" "no" "yes" "yes" "yes" "no" 4 3 5 1 1 2 26 "7" "6" 6
"GP" "M" 15 "U" "GT3" "A" 3 4 "services" "other" "course" "mother" 1 2 0 "no" "yes"
"yes" "yes" "yes" "yes" "yes" "no" 5 4 4 1 1 1 0 "16" "18" 18
"GP" "F" 15 "U" "GT3" "A" 3 3 "other" "health" "reputation" "father" 1 4 0 "yes"
"no" "no" "no" "yes" "yes" "no" "no" 4 3 3 1 1 4 10 "10" "11" 11
"GP" "F" 15 "U" "GT3" "T" 2 2 "other" "other" "course" "mother" 1 4 0 "yes" "yes"
"yes" "no" "yes" "yes" "yes" "no" 5 1 2 1 1 3 8 "7" "8" 8
"GP" "M" 16 "U" "GT3" "T" 3 3 "services" "other" "home" "father" 1 3 0 "no" "yes"
"no" "yes" "yes" "yes" "yes" "no" 5 3 3 1 1 5 2 "16" "18" 18
"GP" "M" 15 "R" "GT3" "T" 4 4 "other" "other" "home" "father" 4 4 0 "no" "yes"
"yes" "yes" "yes" "yes" "yes" "yes" 1 3 5 3 5 1 6 "10" "13" 13
"GP" "F" 16 "U" "LE3" "T" 4 4 "health" "health" "other" "mother" 1 3 0 "no" "yes"
"yes" "yes" "yes" "yes" "yes" "yes" 5 4 5 1 1 4 4 "14" "15" 16
"GP" "M" 15 "U" "LE3" "A" 4 4 "teacher" "teacher" "course" "mother" 1 1 0 "no" "no"
"no" "yes" "yes" "yes" "yes" "no" 5 5 3 1 1 4 6 "18" "19" 19
"GP" "F" 16 "R" "GT3" "T" 3 3 "services" "other" "reputation" "father" 1 3 1 "yes"

```

```
"yes" "no" "yes" "yes" "yes" "yes" "no" 4 1 2 1 1 2 0 "7" "10" 10
"GP" "F" 16 "U" "GT3" "T" 2 2 "at_home" "other" "home" "mother" 1 2 1 "yes" "no"
"no" "yes" "yes" "yes" "yes" "no" 3 1 2 1 1 5 6 "10" "13" 13
"GP" "M" 15 "U" "LE3" "T" 4 2 "teacher" "other" "course" "mother" 1 1 0 "no" "no"
"no" "no" "yes" "yes" "yes" "no" 3 5 2 1 1 3 10 "18" "19" 19
"GP" "M" 15 "R" "GT3" "T" 2 1 "health" "services" "reputation" "mother" 1 2 0 "no"
"no" "no" "yes" "yes" "yes" "yes" "yes" 5 4 2 1 1 5 8 "9" "9" 9
"GP" "M" 16 "U" "GT3" "T" 4 4 "teacher" "teacher" "course" "father" 1 2 0 "no"
"yes" "no" "yes" "yes" "yes" "yes" "no" 5 4 4 1 2 5 2 "15" "15" 16
"GP" "M" 15 "U" "GT3" "T" 4 4 "other" "teacher" "reputation" "father" 2 2 0 "no"
"yes" "no" "yes" "yes" "yes" "no" "no" 4 4 3 1 1 2 2 "11" "13" 14
"GP" "M" 16 "U" "GT3" "T" 3 3 "other" "services" "home" "father" 2 1 0 "no" "no"
"no" "yes" "yes" "yes" "yes" "no" 5 4 2 1 1 5 0 "13" "14" 13
"GP" "M" 17 "R" "GT3" "T" 1 3 "other" "other" "course" "father" 3 2 1 "no" "yes"
"no" "yes" "yes" "yes" "yes" "no" 5 2 4 1 4 5 20 "9" "7" 8
"GP" "M" 15 "U" "GT3" "T" 3 4 "other" "other" "reputation" "father" 1 1 0 "no" "no"
"no" "no" "yes" "yes" "yes" "no" 3 4 3 1 2 4 6 "14" "13" 13
"GP" "F" 15 "U" "GT3" "T" 1 2 "at_home" "services" "course" "mother" 1 2 0 "no"
"no" "no" "no" "no" "yes" "yes" "no" 3 2 3 1 2 1 2 "16" "15" 15
"GP" "M" 15 "U" "GT3" "T" 2 2 "services" "services" "home" "father" 1 4 0 "no"
"yes" "yes" "yes" "yes" "yes" "yes" "no" 5 5 4 1 2 5 6 "16" "14" 15
"GP" "F" 16 "U" "LE3" "T" 2 4 "other" "health" "course" "father" 2 2 0 "no" "yes"
"yes" "yes" "yes" "yes" "yes" "yes" 4 2 2 1 2 5 2 "13" "13" 13
"GP" "M" 16 "U" "GT3" "T" 4 4 "health" "other" "course" "mother" 1 1 0 "no" "yes"
"no" "yes" "yes" "yes" "yes" "no" 3 4 4 1 4 5 18 "14" "11" 13
"GP" "F" 16 "U" "GT3" "T" 2 2 "other" "other" "home" "mother" 1 2 0 "no" "no" "yes"
"no" "yes" "yes" "yes" "yes" 5 4 4 1 1 5 0 "8" "7" 8
"GP" "M" 15 "U" "GT3" "T" 3 4 "services" "services" "home" "father" 1 1 0 "yes"
"no" "no" "no" "yes" "yes" "yes" "no" 5 5 5 3 2 5 0 "13" "13" 12
"GP" "F" 15 "U" "LE3" "A" 3 4 "other" "other" "home" "mother" 1 2 0 "yes" "no" "no"
"yes" "yes" "yes" "yes" "yes" 5 3 2 1 1 1 0 "7" "10" 11
"GP" "F" 19 "U" "GT3" "T" 0 1 "at_home" "other" "course" "other" 1 2 3 "no" "yes"
"no" "no" "no" "no" "no" "no" 3 4 2 1 1 5 2 "7" "8" 9
"GP" "M" 18 "R" "GT3" "T" 2 2 "services" "other" "reputation" "mother" 1 1 2 "no"
"yes" "no" "yes" "yes" "yes" "yes" "no" 3 3 3 1 2 4 0 "7" "4" 0
"GP" "M" 16 "R" "GT3" "T" 4 4 "teacher" "teacher" "course" "mother" 1 1 0 "no" "no"
"yes" "yes" "yes" "yes" "yes" "no" 3 5 5 2 5 4 8 "18" "18" 18
"GP" "F" 15 "R" "GT3" "T" 3 4 "services" "teacher" "course" "father" 2 3 2 "no"
"yes" "no" "no" "yes" "yes" "yes" "yes" 4 2 2 2 2 5 0 "12" "0" 0
"GP" "F" 15 "U" "GT3" "T" 1 1 "at_home" "other" "course" "mother" 3 1 0 "no" "yes"
"no" "yes" "no" "yes" "yes" "yes" 4 3 3 1 2 4 0 "8" "0" 0
"GP" "F" 17 "U" "LE3" "T" 2 2 "other" "other" "course" "father" 1 1 0 "no" "yes"
"no" "no" "yes" "yes" "yes" "yes" 3 4 4 1 3 5 12 "10" "13" 12
"GP" "F" 16 "U" "GT3" "A" 3 4 "services" "other" "course" "father" 1 1 0 "no" "no"
"no" "no" "yes" "yes" "yes" "no" 3 2 1 1 4 5 16 "12" "11" 11
"GP" "M" 15 "R" "GT3" "T" 3 4 "at_home" "teacher" "course" "mother" 4 2 0 "no"
"yes" "no" "no" "yes" "yes" "no" "yes" 5 3 3 1 1 5 0 "9" "0" 0
"GP" "F" 15 "U" "GT3" "T" 4 4 "services" "at_home" "course" "mother" 1 3 0 "no"
"yes" "no" "yes" "yes" "yes" "yes" "yes" 4 3 3 1 1 5 0 "11" "0" 0
"GP" "M" 17 "R" "GT3" "T" 3 4 "at_home" "other" "course" "mother" 3 2 0 "no" "no"
"no" "no" "yes" "yes" "no" "no" 5 4 5 2 4 5 0 "10" "0" 0
"GP" "F" 16 "U" "GT3" "A" 3 3 "other" "other" "course" "other" 2 1 2 "no" "yes"
"no" "yes" "no" "yes" "yes" "yes" 4 3 2 1 1 5 0 "4" "0" 0
```



```

"GP" "M" 16 "U" "LE3" "T" 1 1 "services" "other" "course" "mother" 1 2 1 "no" "no"
"no" "no" "yes" "yes" "no" "yes" 4 4 4 1 3 5 0 "14" "12" 12
"GP" "F" 15 "U" "GT3" "T" 4 4 "teacher" "teacher" "course" "mother" 2 1 0 "no" "no"
"no" "yes" "yes" "yes" "yes" "no" 4 3 2 1 1 5 0 "16" "16" 15
"GP" "M" 15 "U" "GT3" "T" 4 3 "teacher" "services" "course" "father" 2 4 0 "yes"
"yes" "no" "no" "yes" "yes" "yes" "no" 2 2 2 1 1 3 0 "7" "9" 0
"GP" "M" 16 "U" "LE3" "T" 2 2 "services" "services" "reputation" "father" 2 1 2
"no" "yes" "no" "yes" "yes" "yes" "yes" "no" 2 3 3 2 2 2 8 "9" "9" 9
"GP" "F" 15 "U" "GT3" "T" 4 4 "teacher" "services" "course" "mother" 1 3 0 "no"
"yes" "yes" "yes" "yes" "yes" "yes" "no" 4 2 2 1 1 5 2 "9" "11" 11
"GP" "F" 16 "U" "LE3" "T" 1 1 "at_home" "at_home" "course" "mother" 1 1 0 "no" "no"
"no" "no" "yes" "yes" "yes" "no" 3 4 4 3 3 1 2 "14" "14" 13
"GP" "M" 17 "U" "GT3" "T" 2 1 "other" "other" "home" "mother" 1 1 3 "no" "yes" "no"
"no" "yes" "yes" "yes" "no" 5 4 5 1 2 5 0 "5" "0" 0
"GP" "F" 15 "U" "GT3" "T" 1 1 "other" "services" "course" "father" 1 2 0 "no" "yes"
"yes" "no" "yes" "yes" "yes" "no" 4 4 2 1 2 5 0 "8" "11" 11
"GP" "F" 15 "U" "GT3" "T" 3 2 "health" "services" "home" "father" 1 2 3 "no" "yes"
"no" "no" "yes" "yes" "yes" "no" 3 3 2 1 1 3 0 "6" "7" 0
"GP" "F" 15 "U" "GT3" "T" 1 2 "at_home" "other" "course" "mother" 1 2 0 "no" "yes"
"yes" "no" "no" "yes" "yes" "no" 4 3 2 1 1 5 2 "10" "11" 11
"GP" "M" 16 "U" "GT3" "T" 4 4 "teacher" "teacher" "course" "mother" 1 1 0 "no"
"yes" "no" "no" "yes" "no" "yes" "yes" 3 3 2 2 1 5 0 "7" "6" 0
"GP" "M" 15 "U" "LE3" "A" 2 1 "services" "other" "course" "mother" 4 1 3 "no" "no"
"no" "no" "yes" "yes" "yes" "no" 4 5 5 2 5 5 0 "8" "9" 10
"GP" "M" 18 "U" "LE3" "T" 1 1 "other" "other" "course" "mother" 1 1 3 "no" "no"
"no" "no" "yes" "no" "yes" "yes" 2 3 5 2 5 4 0 "6" "5" 0
"GP" "M" 16 "U" "LE3" "T" 2 1 "at_home" "other" "course" "mother" 1 1 1 "no" "no"
"no" "yes" "yes" "yes" "no" "yes" 4 4 4 3 5 5 6 "12" "13" 14
"GP" "F" 15 "R" "GT3" "T" 3 3 "services" "services" "reputation" "other" 2 3 2 "no"
"yes" "yes" "yes" "yes" "yes" "yes" "yes" 4 2 1 2 3 3 8 "10" "10" 10
"GP" "M" 19 "U" "GT3" "T" 3 2 "services" "at_home" "home" "mother" 1 1 3 "no" "yes"
"no" "no" "yes" "no" "yes" "yes" 4 5 4 1 1 4 0 "5" "0" 0
"GP" "F" 17 "U" "GT3" "T" 4 4 "other" "teacher" "course" "mother" 1 1 0 "yes" "yes"
"no" "no" "yes" "yes" "no" "yes" 4 2 1 1 1 4 0 "11" "11" 12
"GP" "M" 15 "R" "GT3" "T" 2 3 "at_home" "services" "course" "mother" 1 2 0 "yes"
"no" "yes" "yes" "yes" "yes" "no" "no" 4 4 4 1 1 1 2 "11" "8" 8
"GP" "M" 17 "R" "LE3" "T" 1 2 "other" "other" "reputation" "mother" 1 1 0 "no" "no"
"no" "no" "yes" "yes" "no" "no" 2 2 2 3 3 5 8 "16" "12" 13
"GP" "F" 18 "R" "GT3" "T" 1 1 "at_home" "other" "course" "mother" 3 1 3 "no" "yes"
"no" "yes" "no" "yes" "no" "no" 5 2 5 1 5 4 6 "9" "8" 10
"GP" "M" 16 "R" "GT3" "T" 2 2 "at_home" "other" "course" "mother" 3 1 0 "no" "no"
"no" "no" "no" "yes" "no" "no" 4 2 2 1 2 3 2 "17" "15" 15
"GP" "M" 16 "U" "GT3" "T" 3 3 "other" "services" "course" "father" 1 2 1 "no" "yes"
"yes" "no" "yes" "yes" "yes" "yes" 4 5 5 4 4 5 4 "10" "12" 12
"GP" "M" 17 "R" "LE3" "T" 2 1 "at_home" "other" "course" "mother" 2 1 2 "no" "no"
"no" "yes" "yes" "no" "yes" "yes" 3 3 2 2 2 5 0 "7" "6" 0
"GP" "M" 15 "R" "GT3" "T" 3 2 "other" "other" "course" "mother" 2 2 2 "yes" "yes"
"no" "no" "yes" "yes" "yes" "yes" 4 4 4 1 4 3 6 "5" "9" 7
"GP" "M" 16 "U" "LE3" "T" 1 2 "other" "other" "course" "mother" 2 1 1 "no" "no"
"no" "yes" "yes" "yes" "no" "no" 4 4 4 2 4 5 0 "7" "0" 0
"GP" "M" 17 "U" "GT3" "T" 1 3 "at_home" "services" "course" "father" 1 1 0 "no"
"no" "no" "no" "yes" "no" "yes" "no" 5 3 3 1 4 2 2 "10" "10" 10
"GP" "M" 17 "R" "LE3" "T" 1 1 "other" "services" "course" "mother" 4 2 3 "no" "no"

```

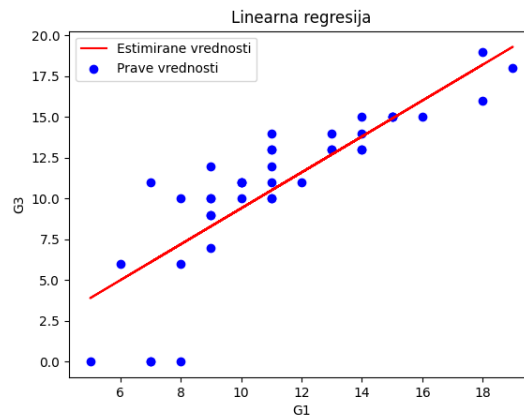
```

"no" "yes" "yes" "no" "no" "yes" 5 3 5 1 5 5 0 "5" "8" 7
"GP" "M" 16 "U" "GT3" "T" 3 2 "services" "services" "course" "mother" 2 1 1 "no"
"yes" "no" "yes" "no" "no" "no" "no" 4 5 2 1 1 2 16 "12" "11" 12
"GP" "M" 16 "U" "GT3" "T" 2 2 "other" "other" "course" "father" 1 2 0 "no" "no"
"no" "no" "yes" "no" "yes" "no" 4 3 5 2 4 4 4 "10" "10" 10
"GP" "F" 16 "U" "GT3" "T" 4 2 "health" "services" "home" "father" 1 2 0 "no" "no"
"yes" "no" "yes" "yes" "yes" "yes" 4 2 3 1 1 3 0 "14" "15" 16
"GP" "F" 16 "U" "GT3" "T" 2 2 "other" "other" "home" "mother" 1 2 0 "no" "yes"
"yes" "no" "no" "yes" "yes" "no" 5 1 5 1 1 4 0 "6" "7" 0
"GP" "F" 16 "U" "GT3" "T" 4 4 "health" "health" "reputation" "mother" 1 2 0 "no"
"yes" "yes" "no" "yes" "yes" "yes" "yes" 4 4 2 1 1 3 0 "14" "14" 14
"GP" "M" 16 "U" "GT3" "T" 3 4 "other" "other" "course" "father" 3 1 2 "no" "yes"
"no" "yes" "no" "yes" "yes" "no" 3 4 5 2 4 2 0 "6" "5" 0
"GP" "M" 16 "U" "GT3" "T" 1 0 "other" "other" "reputation" "mother" 2 2 0 "no"
"yes" "yes" "yes" "yes" "yes" "yes" "yes" 4 3 2 1 1 3 2 "13" "15" 16
"GP" "M" 17 "U" "LE3" "T" 4 4 "teacher" "other" "reputation" "mother" 1 2 0 "no"
"yes" "yes" "yes" "yes" "yes" "yes" "no" 4 4 4 1 3 5 0 "13" "11" 10
"GP" "F" 16 "U" "GT3" "T" 1 3 "at_home" "services" "home" "mother" 1 2 3 "no" "no"
"no" "yes" "no" "yes" "yes" "yes" 4 3 5 1 1 3 0 "8" "7" 0
"GP" "F" 16 "U" "LE3" "T" 3 3 "other" "other" "reputation" "mother" 2 2 0 "no"
"yes" "yes" "yes" "yes" "yes" "yes" "no" 4 4 5 1 1 4 4 "10" "11" 9
"GP" "M" 17 "U" "LE3" "T" 4 3 "teacher" "other" "course" "mother" 2 2 0 "no" "no"
"yes" "yes" "yes" "yes" "yes" "no" 4 4 4 4 4 4 4 "10" "9" 9
"GP" "F" 16 "U" "GT3" "T" 2 2 "services" "other" "reputation" "mother" 2 2 0 "no"
"no" "yes" "yes" "no" "yes" "yes" "no" 3 4 4 1 4 5 2 "13" "13" 11
"GP" "M" 17 "U" "GT3" "T" 3 3 "other" "other" "reputation" "father" 1 2 0 "no" "no"
"no" "yes" "no" "yes" "yes" "no" 4 3 4 1 4 4 4 "6" "5" 6
"GP" "M" 16 "R" "GT3" "T" 4 2 "teacher" "services" "other" "mother" 1 1 0 "no"
"yes" "no" "yes" "yes" "yes" "yes" "yes" 4 3 3 3 4 3 10 "10" "8" 9
"GP" "M" 17 "U" "GT3" "T" 4 3 "other" "other" "course" "mother" 1 2 0 "no" "yes"
"no" "yes" "yes" "yes" "yes" "yes" 5 2 3 1 1 2 4 "10" "10" 11
"GP" "M" 16 "U" "GT3" "T" 4 3 "teacher" "other" "home" "mother" 1 2 0 "no" "yes"
"yes" "yes" "yes" "yes" "yes" "no" 3 4 3 2 3 3 10 "9" "8" 8
"GP" "M" 16 "U" "GT3" "T" 3 3 "services" "other" "home" "mother" 1 2 0 "no" "no"
"yes" "yes" "yes" "yes" "yes" "yes" 4 2 3 1 2 3 2 "12" "13" 12
"GP" "F" 17 "U" "GT3" "T" 2 4 "services" "services" "reputation" "father" 1 2 0
"no" "yes" "no" "yes" "yes" "yes" "no" "no" 5 4 2 2 3 5 0 "16" "17" 17
"GP" "F" 17 "U" "LE3" "T" 3 3 "other" "other" "reputation" "mother" 1 2 0 "no"
"yes" "no" "yes" "yes" "yes" "yes" "yes" 5 3 3 2 3 1 56 "9" "9" 8
"GP" "F" 16 "U" "GT3" "T" 3 2 "other" "other" "reputation" "mother" 1 2 0 "no"
"yes" "yes" "no" "yes" "yes" "yes" "no" 1 2 2 1 2 1 14 "12" "13" 12
"GP" "M" 17 "U" "GT3" "T" 3 3 "services" "services" "other" "mother" 1 2 0 "no"
"yes" "no" "yes" "yes" "yes" "yes" "yes" 4 3 4 2 3 4 12 "12" "12" 11
"GP" "M" 16 "U" "GT3" "T" 1 2 "services" "services" "other" "mother" 1 1 0 "no"
"yes" "yes" "yes" "yes" "yes" "yes" "yes" 3 3 3 1 2 3 2 "11" "12" 11
"GP" "M" 16 "U" "LE3" "T" 2 1 "other" "other" "course" "mother" 1 2 0 "no" "no"
"yes" "yes" "yes" "yes" "yes" "yes" 4 2 3 1 2 5 0 "15" "15" 15
"GP" "F" 17 "U" "GT3" "A" 3 3 "health" "other" "reputation" "mother" 1 2 0 "no"
"yes" "no" "no" "no" "yes" "yes" "yes" 3 3 3 1 3 3 6 "8" "7" 9
"GP" "M" 17 "R" "GT3" "T" 1 2 "at_home" "other" "home" "mother" 1 2 0 "no" "no"
"no" "no" "yes" "yes" "no" "no" 3 1 3 1 5 3 4 "8" "9" 10
"GP" "F" 16 "U" "GT3" "T" 2 3 "services" "services" "course" "mother" 1 2 0 "no"
"no" "no" "no" "yes" "yes" "yes" "no" 4 3 3 1 1 2 10 "11" "12" 13

```

```
"GP" "F" 17 "U" "GT3" "T" 1 1 "at_home" "services" "course" "mother" 1 2 0 "no"
"no" "no" "yes" "yes" "yes" "yes" "no" 5 3 3 1 1 3 0 "8" "8" 9
"GP" "M" 17 "U" "GT3" "T" 1 2 "at_home" "services" "other" "other" 2 2 0 "no" "no"
"yes" "yes" "no" "yes" "yes" "no" 4 4 4 4 5 5 12 "7" "8" 8
"GP" "M" 16 "R" "GT3" "T" 3 3 "services" "services" "reputation" "mother" 1 1 0
"no" "yes" "no" "yes" "yes" "yes" "yes" "no" 4 3 2 3 4 5 8 "8" "9" 10
"GP" "M" 16 "U" "GT3" "T" 2 3 "other" "other" "home" "father" 2 1 0 "no" "no" "no"
"no" "yes" "yes" "yes" "no" 5 3 3 1 1 3 0 "13" "14" 14
"GP" "F" 17 "U" "LE3" "T" 2 4 "services" "services" "course" "father" 1 2 0 "no"
"no" "no" "yes" "yes" "yes" "yes" "yes" 4 3 2 1 1 5 0 "14" "15" 15
"GP" "M" 17 "U" "GT3" "T" 4 4 "services" "teacher" "home" "mother" 1 1 0 "no" "no"
"no" "no" "yes" "yes" "yes" "no" 5 2 3 1 2 5 4 "17" "15" 16
"GP" "M" 16 "R" "LE3" "T" 3 3 "teacher" "other" "home" "father" 3 1 0 "no" "yes"
"yes" "yes" "yes" "yes" "yes" "no" 3 3 4 3 5 3 8 "9" "9" 10
"GP" "F" 17 "U" "GT3" "T" 4 4 "services" "teacher" "home" "mother" 2 1 1 "no" "yes"
"no" "no" "yes" "yes" "yes" "no" 4 2 4 2 3 2 24 "18" "18" 18
"GP" "F" 16 "U" "LE3" "T" 4 4 "teacher" "teacher" "reputation" "mother" 1 2 0 "no"
"yes" "yes" "no" "yes" "yes" "yes" "no" 4 5 2 1 2 3 0 "9" "9" 10
"GP" "F" 16 "U" "GT3" "T" 4 3 "health" "other" "home" "mother" 1 2 0 "no" "yes"
"no" "yes" "yes" "yes" "yes" "no" 4 3 5 1 5 2 2 "16" "16" 16
"GP" "F" 16 "U" "GT3" "T" 2 3 "other" "other" "reputation" "mother" 1 2 0 "yes"
"yes" "yes" "yes" "yes" "yes" "no" "no" 4 4 3 1 3 4 6 "8" "10" 10
"GP" "F" 17 "U" "GT3" "T" 1 1 "other" "other" "course" "mother" 1 2 0 "no" "yes"
"yes" "no" "no" "yes" "no" "no" 4 4 4 1 3 1 4 "9" "9" 10
"GP" "F" 17 "R" "GT3" "T" 2 2 "other" "other" "reputation" "mother" 1 1 0 "no"
"yes" "no" "no" "yes" "yes" "yes" "no" 5 3 2 1 2 3 18 "7" "6" 6
"GP" "F" 16 "R" "GT3" "T" 2 2 "services" "services" "reputation" "mother" 2 4 0
"no" "yes" "yes" "yes" "no" "yes" "yes" "no" 5 3 5 1 1 5 6 "10" "10" 11
"GP" "F" 17 "U" "GT3" "T" 3 4 "at_home" "services" "home" "mother" 1 3 1 "no" "yes"
"yes" "no" "yes" "yes" "yes" "yes" 4 4 3 3 4 5 28 "10" "9" 9
"GP" "F" 16 "U" "GT3" "A" 3 1 "services" "other" "course" "mother" 1 2 3 "no" "yes"
"yes" "no" "yes" "yes" "yes" "no" 2 3 3 2 2 4 5 "7" "7" 7
"GP" "F" 16 "U" "GT3" "T" 4 3 "teacher" "other" "other" "mother" 1 2 0 "no" "no"
"yes" "yes" "yes" "yes" "yes" "yes" 1 3 2 1 1 1 10 "11" "12" 13
"GP" "F" 16 "U" "GT3" "T" 1 1 "at_home" "other" "home" "mother" 2 1 0 "no" "yes"
"yes" "no" "yes" "yes" "no" "no" 4 3 2 1 4 5 6 "9" "9" 10
"GP" "F" 17 "R" "GT3" "T" 4 3 "teacher" "other" "reputation" "mother" 2 3 0 "no"
"yes" "yes" "yes" "yes" "yes" "yes" "yes" 4 4 2 1 1 4 6 "7" "7" 7
"GP" "F" 19 "U" "GT3" "T" 3 3 "other" "other" "reputation" "other" 1 4 0 "no" "yes"
"yes" "yes" "yes" "yes" "yes" "no" 4 3 3 1 2 3 10 "8" "8" 8
"GP" "M" 17 "U" "LE3" "T" 4 4 "services" "other" "home" "mother" 1 2 0 "no" "yes"
"yes" "no" "yes" "yes" "yes" "yes" 5 3 5 4 5 3 13 "12" "12" 13
"GP" "F" 16 "U" "GT3" "A" 2 2 "other" "other" "reputation" "mother" 1 2 0 "yes"
"yes" "yes" "no" "yes" "yes" "yes" "no" 3 3 4 1 1 4 0 "12" "13" 14
"GP" "M" 18 "U" "GT3" "T" 2 2 "services" "other" "home" "mother" 1 2 1 "no" "yes"
"yes" "yes" "yes" "yes" "yes" "no" 4 4 4 2 4 5 15 "6" "7" 8
"GP" "F" 17 "R" "LE3" "T" 4 4 "services" "other" "other" "mother" 1 1 0 "no" "yes"
"yes" "no" "yes" "yes" "no" "no" 5 2 1 1 2 3 12 "8" "10" 10
"GP" "F" 17 "U" "LE3" "T" 3 2 "other" "other" "reputation" "mother" 2 2 0 "no" "no"
"yes" "no" "yes" "yes" "yes" "no" 4 4 4 1 3 1 2 "14" "15" 15
"GP" "F" 17 "U" "GT3" "T" 4 3 "other" "other" "reputation" "mother" 1 2 2 "no" "no"
"yes" "no" "yes" "yes" "yes" "yes" 3 4 5 2 4 1 22 "6" "6" 4
"GP" "M" 18 "U" "LE3" "T" 3 3 "services" "health" "home" "father" 1 2 1 "no" "yes"
```

```
"yes" "no" "yes" "yes" "yes" "no" 3 2 4 2 4 4 13 "6" "6" 8
"GP" "F" 17 "U" "GT3" "T" 2 3 "at_home" "other" "home" "father" 2 1 0 "no" "yes"
"yes" "no" "yes" "yes" "no" "no" 3 3 3 1 4 3 3 "7" "7" 8
"GP" "F" 17 "U" "GT3" "T" 2 2 "at_home" "at_home" "course" "mother" 1 3 0 "no"
"yes" "yes" "yes" "yes" "yes" "yes" "no" 4 3 3 1 1 4 4 "9" "10" 10
"GP" "F" 17 "R" "GT3" "T" 2 1 "at_home" "services" "reputation" "mother" 2 2 0 "no"
"yes" "no" "yes" "yes" "yes" "yes" "no" 4 2 5 1 2 5 2 "6" "6" 6
"GP" "F" 17 "U" "GT3" "T" 1 1 "at_home" "other" "reputation" "mother" 1 3 1 "no"
"yes" "no" "yes" "yes" "yes" "no" "yes" 4 3 4 1 1 5 0 "6" "5" 0
"GP" "F" 16 "U" "GT3" "T" 2 3 "services" "teacher" "other" "mother" 1 2 0 "yes"
"no" "no" "no" "yes" "yes" "yes" "no" 2 3 1 1 1 3 2 "16" "16" 17
"GP" "M" 18 "U" "GT3" "T" 2 2 "other" "other" "home" "mother" 2 2 0 "no" "yes"
"yes" "no" "yes" "yes" "yes" "no" 3 3 3 5 5 4 0 "12" "13" 13
"GP" "F" 16 "U" "GT3" "T" 4 4 "teacher" "services" "home" "mother" 1 3 0 "no" "yes"
"no" "yes"
```



Slika 6.1 Grafik linearne regresije. [Izvor: Autor]

ZADATAK #2

Zadatak #2 se odnosi na logističku regresiju i radi se okvirno 20 minuta

Zadatak #2 (20 minuta)

Učitati datoteku job.csv. Koristiti kolone GMAT, GPA, Experience kao skup ulaznih vrednosti, a kolonu Admitted kao skup izlaznih vrednosti.

Istrenirati model logističke regresije koja uzima 75% skupa za trening skup, a ostalo za test skup.

Štampati u koloni prave vrednosti, estimirane vrednosti. Izračunati i štampati broj grešaka.

Zatim, učitati datoteku new_candidates.csv i na osnovu ulaznih parametara i istreniranog modela, napraviti estimaciju za **y** i štampati kao tabelu.

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression

df = pd.read_csv('jobs.csv', sep='\t')

X = df[['GMAT', 'GPA', 'Experience']]
y = df['Admitted']

#train is based on 75% of the dataset, test is based on 25% of dataset
X_train,X_test,y_train,y_test =
train_test_split(X,y,test_size=0.25,random_state=0)

logistic_regression= LogisticRegression()

logistic_regression.fit(X_train,y_train)

y_pred=logistic_regression.predict(X_test)

# print
df2 = pd.DataFrame(X_test)
df2['Actual'] = y_test
df2['Predicted'] = y_pred
print(df2)
errors = np.sum(np.where(y_test != y_pred, 1, 0))
print("Broj gresaka: ", errors)

df2 = pd.read_csv('new_candidates.csv', sep='\t')

y_pred=logistic_regression.predict(df2)

df2['Predictions'] = y_pred
print(df2)
```

GMAT	GPA	Experience	Admitted
780	4.0	3	1
750	3.9	4	1
690	3.3	3	0
710	3.7	5	1
680	3.9	4	0
730	3.7	6	1
690	2.3	1	0
720	3.3	4	1
740	3.3	5	1
690	1.7	1	0
610	2.7	3	0
690	3.7	5	1
710	3.7	6	1
680	3.3	4	0
770	3.3	3	1

```
610 3.0 1 0
580 2.7 4 0
650 3.7 6 1
540 2.7 2 0
590 2.3 3 0
620 3.3 2 1
600 2.0 1 0
550 2.3 4 0
550 2.7 1 0
570 3.0 2 0
670 3.3 6 1
660 3.7 4 1
580 2.3 2 0
650 3.7 6 1
660 3.3 5 1
640 3.0 1 0
620 2.7 2 0
660 4.0 4 1
660 3.3 6 1
680 3.3 5 1
650 2.3 1 0
670 2.7 2 0
580 3.3 1 0
590 1.7 4 0
690 3.7 5 1
```

```
# new values
new_candidates = {'GMAT': [590,740,680,610,710],
                  'GPA': [2,3.7,3.3,2.3,3],
                  'Experience': [3,4,6,1,5]
                  }
```

	GMAT	GPA	Experience	Actual	Predicted
22	550	2.3	4	0	0
20	620	3.3	2	1	0
25	670	3.3	6	1	1
4	680	3.9	4	0	1
10	610	2.7	3	0	0
15	610	3.0	1	0	0
28	650	3.7	6	1	1
11	690	3.7	5	1	1
18	540	2.7	2	0	0
29	660	3.3	5	1	1

Broj gresaka: 2

	GMAT	GPA	Experience	Predictions
0	590	2.0	3	0
1	740	3.7	4	1
2	680	3.3	6	1
3	610	2.3	1	0
4	710	3.0	5	1

PS D:\CS324\L12>

Slika 6.2 Izlaz logističke regresije. [Izvor: Autor]

VIDEO OBJAŠNJENJA ZA POKAZNE VEŽBE

Slede video objašnjenja za zadatke pokaznih vežbi

Ova lekcija sadrži video materijal. Ukoliko želite da pogledate ovaj video morate da otvorite LAMS lekciju.

Ova lekcija sadrži video materijal. Ukoliko želite da pogledate ovaj video morate da otvorite LAMS lekciju.

▼ Poglavlje 7

Individualne vežbe #12

ZADACI INDIVIDUALNIH VEŽBI

Zadaci individualnih vežbi odnosi se na linearnu i logističku regresiju i rade se okvirno po 45 min

Zadatak #1 (45 minuta)

Učitati datoteku **LinReg_ocene.csv** i istrenirati model linearne regresije tako da ulazni podaci budu kolone:

- samo kolona '**cs101_ocena**'
- samo kolona '**cs102_izostanci**'
- kolone '**cs101_ocena**' i '**cs102_izostanci**'

Izlazni podaci u svakom od slučaja jeste kolona '**cs102_ocena**' I njene vrednosti treba da se estimiraju.

Istrenirati modele linearne regresije (koristiti 75%, a zatim 90% skupa za trening skup) za svaku od traženih kolona, naći broj grešaka i tačnost za svaki, i zaključiti koji je model najbolje koristiti.

Kada se koristi samo jedna kolona, nacrtati i grafike estimiranih vrednosti naspram pravih vrednosti.

Zadatak #2 (45 minuta)

Učitati datoteku **LogReg_ocene.csv** i istrenirati model logističke regresije tako da ulazni podaci budu kolone:

- samo kolona '**broj_polozenih_ispita**',
- samo kolona '**cs101_izostanci**',
- samo kolona '**cs101_polozen**',
- kolone '**broj_polozenih_ispita**' i '**cs101_izostanci**'
- kolone '**broj_polozenih_ispita**', '**cs101_izostanci**', '**cs101_polozen**', '**cs102_izostanci**'

Izlazni podaci u svakom od slučaja jeste kolona '**cs102_polozen**' I njene vrednosti treba da se estimiraju.

Istrenirati modele logističke regresije (koristiti 75%, a zatim 90% skupa za trening skup) za svaku od traženih kolona, naći broj grešaka i tačnost za svaki, i zaključiti koji je model najbolje koristiti.

VREDNOSTI CSV DATOTEKA ZA ZADATKE INDIVIDUALNIH VEŽBI.

*U nastavku su polovične vrednosti CSV datoteka (cele datoteke su u
Dodatnim aktivnostima)*

Vrednosti za Zadatak #1

```
cs101_ocena,cs102_izostanci,cs102_ocena
9,0,9
10,2,10
10,0,10
10,0,10
10,0,10
10,0,10
9,1,9
10,0,10
10,0,10
10,0,10
9,1,9
10,0,10
9,0,9
10,0,10
9,0,9
10,0,10
10,1,10
9,0,9
10,0,10
9,0,9
10,2,9
10,0,10
9,2,9
10,0,10
10,0,10
10,2,9
9,0,9
10,0,10
9,0,9
10,0,10
9,0,9
10,2,9
9,0,9
10,0,10
9,0,9
10,0,10
10,0,10
10,0,10
9,1,9
10,0,10
10,0,10
10,0,10
9,0,9
```

10,0,10
9,0,9
10,0,10
9,0,9
10,0,10
10,0,10
10,0,10
9,0,9
10,0,10
10,0,10
9,0,9
9,0,9
10,0,10
9,0,9
10,0,10
10,0,10
10,0,9
10,0,10
9,0,9
10,0,10
10,0,10
9,0,9
10,0,10
10,0,10
9,0,9
9,0,9
10,1,10
10,0,10
10,0,10
10,0,10
10,0,10
10,0,10
10,0,10
9,0,9
10,1,10
10,0,10
10,1,9
10,0,10
10,0,10
10,0,9
10,0,10
10,0,10
10,0,9
10,0,10
10,1,10
10,0,10
10,0,9
10,0,10
10,0,10
10,0,10
10,0,10
10,0,10
10,0,10
10,0,10
10,0,10

10,0,10
10,0,10
10,1,10
10,1,10
8,2,8
8,2,8
8,2,8
8,1,8
8,2,8
7,1,7
9,1,9
8,1,8
8,1,8
7,2,7
8,2,8
7,1,7
8,1,8
8,2,8
8,2,8
6,1,6
8,2,8
6,1,6
7,1,7
9,1,9
8,2,8
8,2,8
8,2,7
8,2,8
8,2,7
7,3,7
8,1,8
8,1,7
8,1,8
8,1,8
8,3,8
6,2,7
7,2,7
7,1,7
8,2,8
8,3,8
8,2,8
9,2,9
8,3,8
8,1,8
8,1,8
7,2,7
7,2,7
9,2,9
8,2,8
8,2,8
7,3,7
7,2,7
8,1,7

8,2,8
7,3,7
8,2,8
8,2,8
8,2,8
8,2,8
8,2,8
7,2,7
8,2,8
7,2,7
8,2,8
8,2,8
8,2,8
7,2,7
7,2,7
8,2,8
8,2,8
7,2,7
8,3,8
8,2,8
8,2,8
7,2,7
7,2,7
8,3,7
9,2,9
8,2,8
7,2,7
8,2,8
7,2,7
8,2,8
8,2,8
8,2,8
9,3,9
8,3,8
7,2,7
8,2,8
8,2,8
8,2,8
8,2,8
7,2,7
8,3,8
7,2,7
7,3,7
8,2,8
8,2,8
7,2,8
8,2,8
8,2,8
8,2,8
7,2,7
7,4,7
6,3,6

5,4,5
7,4,7
7,3,7
7,3,7
7,3,7
6,2,6
7,4,7
7,0,7
7,2,7
7,4,7
7,2,7
7,4,7
7,2,7
7,1,7
7,2,7
7,2,7
6,4,6
7,4,6
7,4,7
8,2,8
6,1,6
7,2,7
7,1,7
7,3,7
7,4,7
5,4,6
5,4,5
7,2,6
6,1,6
7,2,7
5,2,6
7,4,6
6,2,6
6,3,6
7,1,7
5,4,5
5,4,6
5,3,5
5,4,5
5,4,5
6,3,6
7,4,7
7,2,7
5,4,5
5,4,5
5,3,6
7,3,7
7,3,7
5,1,5
6,4,6
7,3,7
6,4,6
6,1,6

5,1,6
6,2,6
5,4,5
5,1,5
6,2,6
5,4,5
7,2,7
7,4,7
5,4,5
7,2,7
7,4,7
7,2,7
7,4,7
7,4,7
7,4,6
7,2,7
7,4,6
7,3,7
7,4,6
7,2,7
7,4,7
7,3,7
6,2,6
7,4,7
7,4,7
7,3,6
7,3,7
7,4,7
7,2,7
6,4,6
7,3,7
7,4,7
6,2,6
7,4,7
7,4,7
7,4,7
7,4,7
6,2,6
7,4,7
7,4,7
7,1,7
6,4,6
7,2,7
7,4,7
6,4,6
6,3,6
7,4,7
7,4,7
7,4,7
7,4,7
7,3,7
7,4,7
7,4,7

7,4,7
7,4,7
7,4,7
7,4,6
7,3,6
7,4,7
7,3,6
7,4,7
7,2,7
7,4,6
7,3,6
7,4,7
7,4,7
7,2,7
7,3,7
7,3,7
7,4,7
6,4,6
7,4,7
7,4,7
7,4,7
6,3,6
7,1,7
7,4,7
7,4,7
7,1,7
6,4,6
6,4,6
7,4,7
7,4,7
7,1,6
7,0,7
7,1,7
6,3,6
5,4,5
7,4,7
7,1,7
7,1,7
7,3,7
7,4,7
6,4,6
7,4,7
7,3,7
6,2,6
7,4,7
6,4,6
7,4,7
7,1,7
7,4,6
7,3,7
7,4,7
6,4,6
6,4,6

7,4,7
7,4,7
7,3,7
7,4,7
7,4,7
6,1,6
6,4,6
7,4,7
6,4,6
6,4,6
7,4,7
5,2,5
6,4,6
7,2,7
6,2,6
6,4,6
7,4,7
5,4,5
7,1,7
6,4,6
6,1,6
6,7,6
7,7,7
6,4,6
6,3,6
7,7,7
7,7,6
7,2,7
7,4,7
6,4,6
7,7,7
7,2,7
7,4,7
7,1,7
7,7,7
7,7,7
6,4,6
6,4,6
7,7,7
5,5,5
5,2,6
5,5,5
5,5,5
5,2,5
5,7,5
5,3,6
5,5,5
5,5,5
5,3,5
5,2,5
5,2,5
5,2,5
5,5,5

5,5,5
5,5,5
5,3,5
5,2,5
5,2,6
5,2,5
5,5,5
5,2,5
5,3,6
5,3,6
5,5,5
5,2,5
5,5,5
5,2,6
5,2,5
5,2,5
5,3,5
5,5,5
5,5,5
5,2,5
5,5,5
5,3,5
5,2,5
5,5,5
5,5,5
5,5,5
5,5,5
5,5,5
5,5,5
5,2,5
5,5,5
5,2,5
5,4,5
5,5,5
5,5,5
5,5,5
5,2,5
5,5,5
5,4,5
5,5,6
5,5,5
5,2,6
5,2,5
5,5,5
5,1,5
5,5,6
5,5,6
5,1,6
5,7,5
5,7,6
5,2,5
5,5,5
5,4,6

6,5,6
6,2,6
5,4,6
6,4,6
6,5,6
5,7,5
5,5,5
6,2,6
6,5,6
6,7,6
5,4,5
6,7,6
6,7,6
5,5,5
5,5,5
6,5,6
5,7,5
6,2,6
5,3,5
6,5,6
5,5,6
5,5,5
5,3,5
6,5,6
6,7,6
5,3,5
5,7,5
6,5,6
5,5,5
5,3,5
6,3,6
5,7,5
5,5,5

Vrednosti za Zadatak #2

```
broj_polozenih_ispita,cs101_izostanci,cs101_polozen,cs101_ocena,cs102_izostanci,cs102_polozen,cs102_ocena
11,7,0,5,4,1,10
11,11,0,5,5,1,8
3,12,1,6,10,0,5
8,11,1,9,12,1,7
7,10,0,5,9,0,5
7,2,1,9,0,1,6
4,8,1,9,12,1,10
1,13,1,8,3,1,7
11,13,1,6,12,1,7
3,9,0,5,6,0,5
13,8,0,5,15,1,9
1,15,1,9,11,1,6
9,4,1,8,9,1,8
8,13,0,5,5,1,9
7,6,0,5,14,1,6
```

2,8,0,5,5,1,6
 6,10,1,7,3,1,8
 10,0,1,10,4,1,9
 9,8,1,8,2,1,10
 8,4,1,7,2,1,10
 0,2,1,7,11,1,7
 10,5,0,5,6,0,5
 6,7,0,5,13,0,5
 8,12,0,5,12,0,5
 2,1,0,5,7,1,9
 3,1,1,7,8,1,8
 9,9,1,7,0,1,6
 2,4,1,7,14,0,5
 11,8,1,7,14,1,8
 1,3,1,10,0,1,8
 9,15,1,7,12,1,7
 9,0,1,7,0,1,10
 9,2,1,10,6,1,10
 11,13,0,5,9,1,9
 0,9,1,9,10,1,7
 0,10,0,5,4,1,7
 0,14,1,9,9,0,5
 8,6,1,9,9,1,8
 1,3,0,5,4,1,8
 6,3,1,9,11,1,6
 10,12,0,5,1,1,6
 3,14,0,5,15,1,6
 13,11,1,9,8,1,9
 9,15,0,5,10,0,5
 12,7,1,9,3,1,10
 10,6,0,5,3,1,9
 2,8,0,5,0,1,7
 9,4,1,6,15,1,8
 0,0,1,9,5,1,7
 2,8,1,6,15,1,8
 6,8,1,7,0,0,5
 9,15,1,8,11,0,5
 11,4,1,7,11,1,9
 12,13,1,9,2,1,10
 2,11,1,7,5,1,6
 10,11,1,9,0,1,6
 3,5,1,9,1,1,9
 6,2,1,8,3,1,9
 12,2,1,9,3,1,6
 12,1,1,6,8,1,7
 7,7,0,5,3,1,9
 11,1,1,6,11,1,8
 8,0,1,6,6,1,8
 6,5,1,9,13,1,8
 7,15,1,7,14,1,9
 0,1,1,8,11,1,8
 4,4,1,6,9,1,9
 8,9,1,6,13,1,7

7,4,1,6,10,0,5
 3,0,1,8,11,1,7
 5,2,1,10,11,1,7
 13,4,1,10,15,1,9
 9,14,1,8,8,0,5
 10,10,1,6,3,1,8
 5,7,1,9,7,1,7
 11,14,1,7,8,0,5
 5,10,1,8,9,0,5
 2,0,1,10,14,1,6
 1,15,1,8,0,1,9
 3,0,0,5,3,1,10
 12,2,1,10,6,1,7
 8,5,0,5,3,1,7
 7,7,1,6,4,0,5
 4,5,1,9,13,1,10
 1,5,0,5,3,0,5
 5,5,1,7,9,1,8
 13,1,1,7,5,1,9
 7,5,1,7,10,1,6
 0,15,1,9,7,1,10
 2,8,0,5,6,1,8
 7,3,1,9,0,1,8
 5,12,0,5,7,0,5
 10,15,0,5,1,1,8
 7,7,1,6,7,1,7
 10,4,1,10,6,1,7
 5,6,1,9,11,1,10
 10,11,0,5,4,1,9
 13,10,1,7,13,0,5
 9,13,0,5,8,1,6
 3,15,0,5,10,1,6
 7,10,1,9,1,1,9
 11,8,1,7,0,0,5
 9,10,0,5,10,0,5
 4,3,1,8,4,1,10
 0,12,1,6,11,0,5
 4,4,1,8,11,1,9
 4,1,1,7,8,0,5
 13,10,1,8,14,1,10
 13,14,1,7,4,1,6
 9,2,1,10,6,1,10
 13,6,1,7,10,0,5
 4,12,1,7,9,1,8
 6,8,0,5,3,1,8
 13,2,0,5,7,1,9
 5,6,1,9,0,1,10
 7,10,1,9,3,1,9
 10,9,0,5,5,0,5
 9,13,0,5,9,1,6
 5,4,1,6,15,1,7
 8,8,1,6,7,1,8
 10,14,0,5,1,1,9

6,9,1,6,10,0,5
6,4,1,8,2,1,6
5,0,1,10,14,1,10
1,10,1,8,10,1,6
0,0,1,8,8,1,8
13,2,1,6,0,1,9
13,6,1,9,9,1,7
11,14,0,5,11,1,9
0,10,0,5,7,0,5
0,5,0,5,8,0,5
12,5,1,6,7,0,5
12,11,0,5,8,1,7
3,5,0,5,5,0,5
1,13,1,8,4,1,9
13,9,1,6,5,1,7
2,4,1,10,13,1,10
0,15,0,5,5,1,8
5,8,1,8,2,1,10
8,2,0,5,7,1,8
0,11,1,8,14,1,10
9,2,1,8,5,1,9
6,12,1,7,4,1,6
3,9,0,5,0,1,10
4,4,1,8,9,1,7
9,8,1,7,4,1,8
11,8,1,6,9,1,6
2,4,1,8,8,1,10
5,10,0,5,12,0,5
12,7,0,5,8,0,5
9,0,1,7,11,1,8
11,10,1,6,3,1,7
9,13,1,7,1,1,9
0,8,1,7,12,1,8
9,8,1,9,13,1,10
8,5,1,6,15,1,8
10,3,1,6,8,1,6
8,4,1,10,13,1,10
4,2,1,8,4,1,8
10,9,1,7,0,1,6
2,14,1,6,4,0,5
12,10,1,7,3,0,5
0,6,1,8,6,1,7
3,10,1,7,15,1,8
9,6,1,6,1,1,8
3,5,1,8,0,1,10
11,3,1,9,11,1,6
5,8,1,6,15,0,5
2,14,0,5,11,1,9
4,13,1,7,10,1,7
1,12,1,8,2,1,10
3,15,1,8,3,1,10
1,7,1,7,9,0,5
13,9,1,9,7,1,6

12,3,1,8,10,1,9
2,2,1,7,14,1,8
1,1,1,6,10,1,7
7,4,1,10,13,1,9
11,12,1,7,7,1,9
6,5,1,9,4,1,7
4,7,1,6,9,0,5
2,6,0,5,12,0,5
9,13,1,6,13,1,9
9,0,1,8,5,1,10
0,10,0,5,10,1,6
0,3,1,9,3,1,7
6,14,1,9,10,1,9
4,8,0,5,6,1,9
9,5,0,5,11,1,6
8,3,1,8,1,1,10
2,7,0,5,11,1,7
0,0,1,9,13,1,6
10,11,0,5,13,0,5
5,7,1,8,7,1,7
5,2,1,10,9,1,6
12,9,0,5,13,1,6
0,10,1,7,8,1,8
9,8,1,6,9,1,8
5,6,1,6,7,0,5
10,0,0,5,13,1,6
8,7,1,6,4,1,6
1,1,1,7,13,1,6
9,13,0,5,5,1,7
10,1,1,6,15,1,9
12,12,1,8,5,0,5
3,9,0,5,3,1,8
9,14,0,5,1,1,7
4,2,1,6,5,1,8
6,6,0,5,3,1,7
6,6,1,8,2,1,7
4,15,1,9,4,1,8
7,14,1,7,7,1,8
5,9,0,5,0,0,5
2,9,0,5,5,0,5
4,1,1,7,15,1,6
12,3,1,9,0,1,7
10,3,0,5,7,1,8
12,3,1,6,15,1,6
2,7,0,5,13,1,9
9,0,1,10,4,1,10
12,13,1,8,6,1,9
4,4,1,10,14,1,10
3,0,1,8,0,1,6
6,5,1,9,0,1,9
8,13,0,5,11,1,6
5,0,1,6,12,1,9
6,2,1,6,6,1,6

7,13,1,7,10,0,5
 13,2,1,6,8,0,5
 2,14,0,5,2,1,6
 13,11,1,9,5,1,8
 1,0,1,6,1,1,9
 3,10,1,8,14,0,5
 8,14,0,5,14,1,9
 5,6,0,5,9,1,9
 12,9,1,9,1,1,6
 3,3,0,5,10,0,5
 1,15,1,7,7,1,7
 13,7,1,8,14,1,8
 1,4,1,6,5,0,5
 1,8,1,9,8,1,7
 0,3,1,7,3,1,7
 11,13,1,9,6,1,9
 5,10,1,7,4,1,6
 6,2,0,5,13,1,6
 1,0,1,6,1,0,5
 12,2,1,7,10,1,7
 13,15,1,7,4,1,10
 8,5,1,6,1,1,10
 4,9,0,5,15,1,6
 9,9,1,6,11,0,5
 12,3,1,7,10,1,9
 12,12,0,5,0,0,5
 5,4,0,5,13,0,5
 9,14,1,6,7,1,9
 13,2,0,5,14,1,6
 13,9,1,8,11,1,7
 10,15,1,6,15,1,8
 9,12,0,5,1,1,6
 0,11,1,8,9,1,6
 2,1,1,7,1,1,8
 13,0,1,10,8,1,6
 10,3,1,8,14,1,8
 5,5,0,5,10,1,7
 6,5,0,5,11,0,5
 10,3,1,7,0,1,8
 12,9,1,8,14,1,8
 10,0,1,6,0,1,8
 1,6,1,8,13,1,8
 13,2,1,6,3,0,5
 6,8,0,5,10,1,6
 9,7,1,7,4,1,8
 10,2,0,5,0,1,10
 12,4,1,6,4,1,7
 8,15,1,7,14,0,5
 5,8,1,8,3,1,6
 10,6,1,8,7,1,8
 6,6,1,8,14,1,8
 1,11,1,8,1,1,10
 10,10,1,6,14,0,5

9,7,0,5,6,1,7
 13,12,1,8,15,1,9
 0,4,1,8,6,1,10
 7,5,1,7,8,1,8
 6,4,0,5,10,1,8
 11,10,1,8,15,1,9
 2,7,1,8,14,1,6
 10,8,1,9,14,1,9
 10,0,1,8,8,1,6
 1,4,1,9,11,1,7
 1,11,1,6,10,0,5
 12,9,0,5,12,0,5
 7,10,0,5,6,1,6
 8,10,1,9,6,1,9
 11,5,1,8,0,1,10
 3,7,1,8,0,1,7
 10,8,1,7,2,1,8
 6,14,1,6,10,1,7
 10,15,0,5,4,0,5
 4,10,0,5,11,0,5
 8,4,1,8,7,1,6
 0,11,1,7,6,1,8
 4,7,1,9,5,1,7
 3,13,0,5,2,1,8
 3,12,1,9,6,1,8
 3,1,1,9,14,1,10
 2,10,0,5,12,0,5
 7,15,1,8,4,1,10
 13,9,0,5,0,0,5
 12,9,1,9,14,1,8
 7,3,1,6,13,0,5
 6,1,1,10,11,1,10
 2,9,1,8,15,1,9
 5,8,1,6,11,1,6
 10,8,0,5,7,1,7
 13,1,1,6,0,1,9
 8,6,1,8,12,1,8
 2,6,1,6,4,1,8
 7,5,0,5,15,1,7
 2,0,1,7,10,0,5
 6,6,1,6,5,1,9
 7,12,1,8,5,1,10
 9,13,1,6,0,1,9
 12,13,0,5,5,0,5
 0,5,1,9,5,1,9
 11,12,0,5,4,1,8
 8,12,1,9,15,1,10
 3,1,1,10,3,1,10
 8,3,1,8,10,1,6
 4,11,0,5,0,1,10
 0,11,0,5,2,1,7
 0,13,1,8,11,0,5
 3,12,1,8,5,1,7

11,5,1,8,1,1,9
 4,8,1,7,4,0,5
 10,11,1,9,5,1,7
 8,9,1,9,7,1,9
 9,3,1,10,3,1,9
 7,1,1,7,6,0,5
 10,13,1,8,14,1,9
 5,11,1,7,2,1,10
 7,14,1,9,8,1,6
 12,0,1,9,7,1,9
 11,0,1,10,9,1,9
 8,11,0,5,5,1,7
 6,2,1,6,9,1,6
 5,10,1,8,5,1,10
 12,15,1,7,6,0,5
 2,1,0,5,8,1,9
 4,3,1,6,10,1,9
 1,12,1,8,0,1,7
 8,3,1,6,7,1,6
 0,5,0,5,9,0,5
 5,15,1,8,11,1,6
 7,0,1,10,7,1,7
 6,15,0,5,15,1,7
 3,4,1,9,6,1,10
 1,12,0,5,14,0,5
 3,13,1,7,5,0,5
 1,15,1,6,3,0,5
 2,1,1,9,2,1,9
 11,4,1,8,1,1,10
 3,7,0,5,8,0,5
 1,9,0,5,1,0,5
 9,12,1,8,7,1,10
 8,7,1,8,10,0,5
 10,8,0,5,7,1,7
 12,10,0,5,10,1,6
 11,7,0,5,14,0,5
 13,11,0,5,3,0,5
 7,4,1,10,2,1,10
 4,1,1,7,9,0,5
 2,0,1,7,12,1,6
 0,11,1,6,3,1,8
 1,3,1,8,7,1,10
 4,11,0,5,5,0,5
 7,13,0,5,1,1,6
 7,5,0,5,3,1,9
 8,9,1,7,3,1,7
 9,10,0,5,2,0,5
 13,5,1,9,11,1,7
 9,1,1,10,12,1,9
 8,2,1,9,5,1,9
 4,4,0,5,15,1,7
 11,14,0,5,11,1,9
 6,14,1,8,3,1,9

2,3,1,7,1,1,8
11,14,1,6,8,0,5
6,2,0,5,2,1,6
2,12,1,9,9,1,6
7,11,1,6,2,1,8
6,7,0,5,15,0,5
6,10,1,8,3,1,9
4,6,1,8,10,1,6
10,7,1,7,10,1,7
8,5,1,6,9,1,9
4,2,1,9,10,1,9
10,13,1,8,6,0,5
8,8,1,7,4,1,6
13,4,1,9,13,1,9
9,12,1,8,8,1,7
4,1,1,6,15,1,7
10,13,0,5,5,1,8
6,9,1,6,12,1,7
8,3,1,7,1,1,9
13,5,1,9,11,1,6
12,13,1,9,13,1,7
11,8,1,7,9,1,9
3,7,0,5,0,1,8
11,5,1,7,1,0,5
7,0,1,6,14,1,9
3,3,1,10,9,1,7
4,6,1,8,0,1,10
4,0,0,5,4,0,5
3,11,1,9,14,1,9
13,14,1,7,8,0,5
2,8,1,6,7,1,6
2,6,0,5,5,0,5
13,9,1,6,3,1,7
2,0,1,8,5,0,5
4,2,1,6,11,0,5
6,6,0,5,3,0,5
6,12,1,6,8,1,6
0,3,1,8,11,1,10
0,9,1,8,14,1,7
6,15,1,9,9,1,9
0,3,1,10,11,1,6
0,5,1,8,8,1,8
9,0,1,9,4,1,9
10,1,1,6,4,1,6
8,6,0,5,0,1,6
10,5,1,9,15,1,10
11,0,1,7,2,1,10
12,6,1,6,3,1,8
5,5,0,5,15,0,5
10,5,1,7,2,1,8
13,10,1,9,14,1,8
2,8,1,6,7,0,5
4,15,1,6,8,1,6

12,4,1,6,10,1,8
 0,5,1,6,7,1,8
 13,15,0,5,9,0,5
 6,6,0,5,3,0,5
 7,9,0,5,0,1,8
 1,6,1,6,5,1,7
 13,3,1,10,5,1,9
 0,15,1,9,7,1,6
 5,10,1,9,1,1,9
 3,4,1,7,0,1,8
 3,9,0,5,4,1,9
 8,10,1,9,7,0,5
 6,1,1,6,10,1,9
 2,14,1,9,9,1,6
 12,15,1,8,3,1,9
 8,12,1,7,5,0,5
 2,14,1,9,3,1,10
 10,11,1,6,9,0,5
 2,0,1,7,15,1,8
 5,0,1,7,6,1,8
 1,3,1,6,0,1,7
 0,7,0,5,0,0,5
 4,1,1,8,5,1,9
 0,1,1,10,2,1,10
 8,2,0,5,0,0,5
 9,12,1,9,5,1,9
 0,10,1,7,8,1,9
 9,5,1,6,4,1,6
 6,7,1,7,9,1,7
 1,12,1,9,4,1,7
 12,5,1,6,9,1,7
 5,3,1,8,11,0,5
 10,5,0,5,5,1,6
 12,11,1,6,14,1,6
 4,4,1,7,9,0,5
 4,12,0,5,12,1,7
 8,5,1,7,12,1,6
 5,8,0,5,7,1,9
 7,15,1,9,2,1,8
 9,4,1,8,15,1,8
 10,5,1,8,2,1,10
 7,2,1,7,8,1,9
 3,10,1,8,6,1,7
 3,10,0,5,11,0,5
 7,8,0,5,13,1,8
 7,4,1,7,2,1,6
 9,11,1,9,6,1,6
 2,11,1,6,11,1,9
 5,13,0,5,13,1,7
 9,8,0,5,9,1,9
 10,11,1,9,8,1,10
 5,9,1,9,13,1,7
 11,12,1,6,0,1,9

2,6,0,5,1,1,9
3,4,1,6,3,1,9
5,0,1,10,10,1,9
6,10,0,5,4,1,10
0,4,0,5,0,0,5
7,11,1,9,2,1,6
12,9,1,6,9,1,6
7,13,1,6,1,1,7

▼ Poglavlje 8

Domaći zadatak

TEKST DOMAĆEG ZADATKA #12

Domaći zadatak #12 se radi okvirno 4.5 h.

Zadatak #1

Napisati program koji će generisati skup podataka koji će se koristiti za trening i test skupove. Podaci jesu sledeći:

- Kolona #1: Ime: **cs_101_ocena** (vrednost od 5 do 10)
- Kolona #2: Ime: **it_101_ocena** (vrednost od 5 do 10)
- Kolona #3: Ime: **ma_101_ocena** (vrednost od 5 do 10)
- Kolona #4: Ime: **cs_115_izostanci** (vrednosti od 0 do 15)
- Kolona #5: Ime **cs_115_položen** (vrednosti 0 ili 1)
- Kolona #6: Ime: **cs_115_ocena** (vrednosti od 5 do 10)

Kolone popuniti nasumično, i onda promeniti sledeće (ne ručno, već kroz kod):

Ukoliko je vrednost kolone #1 između 8 i 10, povećati vrednost kolone #6 za jedan. Ukoliko je vrednost kolone #4 između 5 i 12, smanjiti vrednost kolone #6 za jedan, a između 13 i 15, smanjiti vrednost kolone #6 za 2. Ukoliko je vrednost kolone #3 između 9 i 10, povećati vrednost kolone #6 za 1. Svuda gde je vrednost kolone #5 nula, vrednost kolone #6 je 5. Obratiti pažnju da pri promeni vrednosti kolona da minimalne i maksimalne vrednosti ostanu nepromenjene.

Svaka kolona ima 500 redova. Sačuvati tabelu kao **dataset.csv**.

Zadatak #2

Učitati skup **dataset.csv** zadatka #1.

Istrenirati model linearna regresije da estimira vrednosti kolone **cs_115_ocena**, ako su ulazni podaci:

- Samo kolona **cs_101 ocena**,
- Samo kolona **cs_115_izostanci**,
- Sve kolone (osim **cs_115_ocena** i **cs_115_polozen**, naravno)

Istrenirati modele **linearne regresije** (koristiti 75%, a zatim 90% skupa za trening skup) za svaku od traženih kolona, naći broj grešaka i tačnost za svaki, i zaključiti koji je model najbolje koristiti. Kada se koristi samo jedna kolona, nacrtati i grafike estimiranih vrednosti naspram pravih vrednosti.

Zadatak #3

Uraditi isto kao i kod **zadatka #2**, samo je model **logistička regresija**, a estimiraju se vrednosti kolone **cs_115_polozen**.

PREDAJA DOMAĆEG ZADATKA

Domaći zadaci se predaju 7 dana nakon predavanja za 100% poena (kod tradicionalnih studenata).

Tradicionalni studenti:

Domaći zadatak treba dostaviti najkasnije 7 dana nakon predavanja, za 100% poena. Nakon toga poeni se umanjuju za 50%.

Internet studenti:

Domaći zadatak treba dostaviti najkasnije 10 dana pred polaganje ispita. Domaći zadaci se brane!

Domaći zadatak poslati dr Nemanji Zdravkoviću: nemanja.zdravkovic@metropolitan.ac.rs

Obavezno koristiti uputstvo za izradu domaćeg zadatka.

Uz .doc dokument (koji treba sadržati i screenshot svakog urađenog zadatka kao i komentare za zadatak), poslati i izvorne i dodatne datoteke.

▼ Poglavlje 9

Zaključak

ZAKLJUČAK

Zaključak lekcije #12

Zaključak

U ovoj lekciji bilo je reči o osnovama veštačke inteligencije i mašinskog učenja. U uvodnom delu definisan je pojam veštačke inteligencije, kao i primene iste. Zatim je data definicija mašinskog učenja kao jednog od polja veštačke inteligencije.

Opisana je razlika između nadgledanog, nenadgledanog i pojačanog učenja kao osnovne podele mašinskog učenja.

Obrađeni su osnovni algoritmi nadgledanog učenja, i to linearna regresija i logistička regresija. Opisan je algoritam opadajućeg gradijenta kao algoritma koji umanjuje grešku pri linearnoj i logističkoj regresiji.

Literatura

- Zsolt Nagy, veštačke inteligencije i mašinskog učenja, Kompjuter biblioteka, 2019. (prevod iste knjige Packt Publishing-a)
- David Beazley, Brian Jones, Python Cookbook: Recipes for Mastering Python 3, 3rd edition, O'Reilly Press, 2013.
- Mark Lutz, Learning Python, 5th Edition, O'Reilly Press, 2013.
- Andrew Bird, Lau Cher Han, et. al, The Python Workshop, Packt Publishing, 2019.
- Al Sweigart, Automate the boring stuff with Python, 2nd Edition, No Starch Press, 2020.

