

Mašinsko učenje - Praktičan deo - Domaći zadaci

1. Organizacija

Praktičan deo kursa je organizovan sa ciljem da se kroz timski rad i rešavanje realnih problema savladaju koncepti, elementi i tehnike mašinskog učenja. Domaći zadaci pokrivaju zasebne (pod)oblasti mašinskog učenja obrađene na predavanjima.

Svaki od 5 zadataka dolazi sa svojim skupom podataka i kriterijumom koji je potrebno ispuniti da bi se zadatak smatrao uspešno urađenim. Gradivo obrađeno na predavanjima je dovoljno kako bi se zadatak uspešno uradio. Dodatnim istraživanjem (pod)oblasti i problema moguće je ostvariti bolje rezultate. Rešenja (modeli) će se evaluirati na skrivenom testnom skupu podataka na online platformi. Opis same platforme dat je u sekciji 2.

Određeni zadaci će zahtevati samostalno implementiranje algoritama mašinskog učenja. Pomenuti algoritmi će biti detaljno obrađeni na predavanjima i materijali sa predavanja će sadržati njihove pseudokodove. Studentima se savetuje da redovno prate predavanja, pošto će se na vežbama podrazumevati da se gradivo sa predavanja zna. **Za izradu preostalih zadataka studentima će na raspolaganju biti Python biblioteke za mašinsko učenje.**

Uz svaki (urađeni) zadatak, timovi će predavati i propratni izveštaj na srpskom ili engleskom jeziku. Opširnije o samom izveštaju u sekciji 3.

Najuspešniji timovi će biti oslobođeni izrade projekta. Kriterijumi na osnovu kojih će se određivati najuspešniji timovi su:

- **Ostvareni rezultati i kako se do njih došlo:**
 - Pristup problemima
 - Korišćeni algoritmi
 - Određivanje (hiper)parametara algoritama
 - Rad sa trening skupom podataka
- **Propratni izveštaji:**
 - Sadržaj propratnih izveštaja
 - Usklađenost izveštaja i izvornih kodova rešenja
- **Diskusija:**
 - Presentovanje rešenja i odgovori na pitanja prilikom prezentovanja.

Takođe, **postoji mogućnost oslobađanja polaganja teorijskog dela ispita za najuspešnije timove koji pokažu izuzetno znanje i razumevanje prilikom prezentovanja i odbrane svojih rešenja.**

2. Platforma

Za izradu zadataka koristiti **Python 3.10.x**. Sama platforma “trči” na **Python 3.10.13**.

Timovima je dostupan **trening skup podataka**. Rešenja se evaluiraju na **testnom skupu podataka**. Testni skup ima isti format i zaglavlje kao i trening skup podataka.

Prilikom *submit*-a Python skripte sa rešenjem, platforma izvršava skriptu pomoću sledeće komande:

```
python <ime_skripte>.py <train_set_path> <test_set_path>
```

Lokacijama skupova podataka se može pristupiti pomoću argumenata komandne linije (**sys.argv**).

Neophodno je da vaša skripta kao rezultat ispiše samo broj koji predstavlja rezultat, bez ikakvog pratećeg teksta:

- **Nerpavilno: RMSE: 10.23**
- **Nepravilno: Dobijena vrednost je 10.23 (RMSE)**
- **Pravilno: 10.23**

Takođe, potrebno je zakomentarisati sva iscrtavanja (*plot*-ove) i one ispise (*print*-ove) koji ne predstavljaju ispis konačnog rešenja (rezultata).

Svaki tim može najviše tri puta submit-ovati svoje rešenje za svaki zadatak. Rešenje mora uključivati i trening i evaluaciju modela.

Skripte se na platformi izvršavaju **asinhrono**, i izvršavanje je ograničeno na **240 sekundi**. Ako se taj period prekorači, **pokušaj će se računati kao neuspešan**. Platforma “trči” na serveru koji ima *Intel Xeon* procesor (4 x 2.4 GHz) i 16 GB RAM memorije.

Nakon što se skripta izvrši, ukoliko je sve prošlo bez problema, u desnom delu prozora (*My Submissions*) će biti prikazan ostvareni rezultat sa zelenom bojom kao pozadinom. Ukoliko nema rezultata, već samo crvena boja kao pozadina - pogledajte sekciju **2.2**.

Na stranici sa zadatkom, ispod dela za *submit*, nalazi se rang lista (*Leaderboard*). Moguće je videti rešenja drugih timova nakon isteka roka za izradu zadatka. Platforma na rang listi uvek prikazuje najbolji rezultat tima.

Ako nakon isteka roka za izradu zadatka na rang listi nema rezultata vašeg tima, to je indikator da je rešenje nakon provere od strane asistenta odbijeno. U tom slučaju na stranici *Submissions* možete videti da vaši *submit*-i za zadatak imaju crvenu boju pozadine, a na stranici *submit*-a sa najboljim rezultatom možete videti komentar asistenta zašto je rešenje odbijeno.

2.1. Biblioteke

Za prvi zadatak, na platformi će biti instalirane sledeće biblioteke (u zagradi su date tačne verzije biblioteka):

- **numpy (1.24.3)**
- **pandas (2.1.0).**

Za drugi zadatak, na platformi će biti instalirane sledeće biblioteke:

- **numpy (1.24.3)**
- **pandas (2.1.0)**
- **scipy (1.11.2).**

Počevši od trećeg zadatka, na platformi će biti instalirane sledeće biblioteke:

- **numpy (1.24.3)**
- **pandas (2.1.0)**
- **scipy (1.11.2)**
- **scikit-learn (1.3.0).**

Prilikom *submit*-a rešenja na platformu, studenti mogu da koriste sve *core* pakete Python-a 3.10.x i gorenavedene biblioteke. Prilikom analize problema, analize skupa podataka, vizualizacije podataka i izrade propratnog izveštaja studenti mogu da koriste biblioteke po izboru, ali je neophodno da njihovi *import*-i i upotreba u *submit* skripti budu zakomentarisani, kako bi se ona bez problema izvršila na platformi.

2.2. Najčešće greške

Ukoliko nakon *submit*-a skripte u desnom delu prozora ne dobijete rešenje, već samo crvenu boju pozadine, na stranici *submit*-a možete videti poruku greške (*Error message*). Značenja nekih poruka greške:

- ***Invalid output format*** - poslednji red ispisa nije broj koji predstavlja rezultat
- ***Execution time exceeded the limit*** - skripta se izvršavala duže od 240 sekundi
- ***No output*** - skripta se nije uspešno izvršila. Uzrok može biti sintaksna greška, korišćenje biblioteke koja nije instalirana na platformi...

Pored gorenavedenih, moguća je i poruka koja predstavlja *exception* koji je skripta "bacila".

Da bi se sprečile greške, a samim tim izgubili i dragoceni pokušaji za zadatak, preporučuje se da pre *submit*-a proverite sledeće stvari:

- Da li koristite odgovarajuću verziju Python-a;
- Da li se skripta može izvršiti putem komande:

python <ime_skripte>.py <train_set_path> <test_set_path>

npr.: **python skripta.py podaci/treniranje.csv podaci/testiranje.csv**

- Da li se kao rezultat izvršavanja vraća samo broj koji predstavlja rešenje, bez pratećeg teksta i suvišnih iscrtavanja i ispisa;
- Da li su zakomentarisani *import*-i i upotreba biblioteka koje nisu instalirane na platformi za taj zadatak;
- Da li izvršavanje skripte traje kraće od 240 sekundi.

Ukoliko i pored svega navedenog ne uspete da utvrdite problem, pošaljite asistentu e-mail.

3. Izveštaj

Uz svaki zadatak, a pre isteka roka za izradu istog, neophodno je poslati izveštaj u kome ćete opisati:

- Vaš pristup problemu
- Isprobane algoritme
- Ostvarene rezultate
- Konačno (odabrano) rešenje.

Grafici i opisi preprocesiranja podataka su dobrodošli.

Izveštaj slati u **PDF** formatu na **ftn.machine.learning@gmail.com** na sledeći način:

- *Email Subject:* **ML24-Z<redni_broj_zadatka>-<username_sa_platforme>**
- *Email Body:* prazan ili sa porukom po izboru
- *Attachment:* Izveštaj u PDF formatu
- PDF nazvati po šablonu:

ML24-Z<redni_broj_zadatka>-<username_sa_platforme>.pdf

Izveštaji poslati nakon isteka roka za izradu zadatka će biti odbačeni.

Rešenja bez izveštaja će biti odbačena.