

# Mašinsko učenje, junski rok, 15. jun 2018.

Sa adrese

<http://poincare.matf.bg.ac.rs/~andjelkaz/pexw>

preuzeti arhivu **ML\_jun1\_2018\_materijali.zip** koja sadrži Jupyter sveske i materijale potrebne za rad. Na *Desktop*-u napraviti direktorijum sa imenom

ML\_jun1\_2018\_ImePrezime\_BrojIndeksa

i njemu čuvati rešenja zadataka. Sveska sa imenom **literatura.ipynb** sadrži linkove do zvaničnih dokumentacija koje je dozvoljeno koristiti u toku rada.

1. **(8 poena)** U datoteci *consumer\_complaints.csv* se nalaze primedbe korisnika na račun usluga jedne banke. Cilj je napraviti klasifikator koji će na osnovu teksta primedbe moći da odredi na koju kategoriju usluga se primedba odnosi.

- a) Učitati podatke koji se nalaze u zadatoj datoteci, a potom izdvojiti podatke koji se nalaze u kolonama "Consumer complaint narrative" i "Product" i eliminisati nedostajuće vrste. Podaci sadržani u koloni "Consumer complaint narrative" će se koristiti kao ulaz klasifikatora, a kategorije sadržane u koloni "Product" kao izlaz klasifikatora. Koliko ima različitih kategorija usluga? Prikazati histogram broja primedbi po kategorijama.
- b) Podeliti skup podataka na trening i test skup u razmeri 2:1. Za *random\_state* parametar uzeti vrednost 7.
- c) Koristeći TF-IDF vektorizaciju pripremiti tekstualne podatke za obradu. Ignorirati sve reči koje se pojavljuju manje od 5 puta i sve *stop* reči engleskog jezika.
- d) Izvršiti mapiranje izlaza klasifikatora u odgovarajuće numeričke kategorije.
- e) Koristeći linearni SVM klasifikator sa politikom "jedan protiv svih" naučiti višeklasni klasifikator.
- e) Dati ocenu tačnosti klasifikatora na test skupu.
- e) Kako se menja tačnost linearnog SVM klasifikatora ako se dodaju podešavanja koja su prilagođena nebalansiranom skupu podataka?

2. **(5 poena)**

Korišćenjem Keras biblioteke napraviti neuronsku mrežu koja će se koristiti u klasifikaciji iris cvetova.

- a) Pročitati podatke koji se nalaze u iris skupu *sklearn* paketa, a zatim preduzeti sve neophodne pripremne korake i podeliti skup podataka na skup za treniranje i skup za testiranje u razmeri 2:1. Za *random\_state* parametar uzeti vrednost 7.
- b) Napraviti neuronsku mrežu koja se sastoji od:
  - ulaznog sloja sa brojem neurona koji odgovara broju atributa ulaznog skupa
  - gustog sloja koji sadži 16 neurona sa sigmoidnom aktivacijom
  - izlaznog sloja sa brojem neurona koji odgovara broju različitih kategorija cvetova; aktivacija ovog sloja je meki maksimum

Mreži pridružiti Adam optimizator, kategoričku unakrsnu entropiju kao funkciju gubitka i tačnost kao metriku.

- c) Nacrtati grafik zavisnosti tačnosti klasifikatora u odnosu na broj epoha u toku treniranja mreže korišćenjem paketa veličine 8. Za broj epoha uzeti 40.
- d) Dati ocenu klasifikatora na skupu za testiranje.
- e) Kakva je ocena klasifikatora na skupu za testiranje ukoliko se prilikom učenja koristi regularizaciona tehnika ranog zaustavljanja sa *patience* parametrom sa vrednošću 4?

3. **(7 poena)**

Jedno od okruženja koje postoji u OpenIA Gym biblioteci je *Taxi-v2* okruženje u kojem je cilj maksimizovati nagradu koja se dobija preuzimanjem putnika i njihovim prevoženjem na 4 različite lokacije. Za svakog prevezenog putnika se dobija nagrada od 20 poena, a za svaki korak transfera se gubi 1 poen. Takođe, postoji penal od 10 poena za svako nepropisno preuzimanje ili ostavljanje putnika. Korišćenjem *Q-learning* algoritma u toku 100 epizoda naučiti optimalno kretanje taksija. Za vrednost koraka učenja uzeti 0.8, a za metaparametar umenjenja uzeti vrednost 0.95.