

Mašinsko učenje, Septembar1 rok, 8. septembar 2018.

Sa adrese

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

preuzeti arhivu **ML_septembar1_2018_materijali.zip** koja sadrži Jupyter sveske i materijale potrebne za rad. Na *Desktop*-u napraviti direktorijum sa imenom

ML_septembar1_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. **(6 poena)** U datotekama *positive_tweets.txt* i *negative_tweets.txt* nalaze se Tviter poruke koje su korisnici doživeli kao pozitivne tj. kao negativne.

- a) Pročitati podatke koji se nalaze u ovim datotekama i na osnovu njih formirati skupove za treniranje i testiranje. Podelu podataka izvršiti u razmeri 4:1 i pritom voditi računa o balansiranosti skupova. Za *random_state* parametar uzeti vrednost 7.
- b) Pripremiti tekstualne podatke za obradu koristeći *CountVectorizer* koji se oslanja na frekvencije pojavljivanja reči. Uzeti u obzir samo prvih 100 najfrekventnijih reči. Prilikom izdvajanja reči ne praviti razliku između reči koje su zapisane malim ili velikim slovima ili njihovim kombinacijama.
- c) Koristeći algoritam logističke regresije napraviti klasifikacioni model i dati ocenu tačnosti.
- d) Izdvojiti sve poruke skupa za testiranje koje pripadaju skupu lažno pozitivnih poruka.

2. **(8 poena)**

U datoteci *USD_INR.csv* se nalaze informacije o kursu indijskih rupija u periodu od 2. januara 1980. godine do 10. avgusta 2017. godine u odnosu na američki dolar. Potrebno je napraviti model koji će vršiti predikciju vrednosti kursa sa što većom tačnošću.

- a) Kreirati skupove za treniranje i testiranje koji će sadržati odgovarajuće vrednosti kurseva za ulazne i izlazne vrednosti. Vrednosti su sadržane u koloni *Price*. Za datum podele podataka uzeti 1. januar 2010. godine.
- b) Korišćenjem Keras biblioteke kreirati neuronsku mrežu sa jednim gustim slojem veličine 15 neurona i jednim gustim slojem sa jednim neuronom. Koristiti srednjekvadratnu grešku za funkciju greške i Adam optimizator. Takođe, prilikom treniranja koristiti i optimizaciju *ranim zaustavljanjem* sa *patience* parametrom postavljenim na 2. Za broj epoha uzeti 20.
- c) Korišćenjem Keras biblioteke kreirati neuronsku mrežu sa jednim rekurentnim slojem sa 7 LSTM neurona i jednim gustim slojem sa jednim neuronom. Koristiti srednjekvadratnu grešku za funkciju greške i Adam optimizator. Takođe, prilikom treniranja koristiti i optimizaciju *ranim zaustavljanjem* sa *patience* parametrom postavljenim na 5. Za broj epoha uzeti 20.
- d) Nacrtati grafik predikcija za podatke sadržane u test skupu za svaki od modela i dati ocenu greške.

3. **(6 poena)**

Okruženje *FrozenLake-v0* OpenIA Gym biblioteke omogućava kretanje agenta gore, dole, levo ili desno duž mreže dimenzija 4x4 u kojoj je početno polje bezbedno (S), neka polja su zamrznuta (F) i preko njih se može preći, neka predstavljaju rupe (H), a jedno od polja je cilj (G). Epizoda se završava kada agent upadne u rupu ili kada stigne do cilja. Ukoliko agent stigne do cilja, dobija nagradu čija je vrednost 1. Inače, ne dobija nagradu.

Korišćenjem *Q-learning* algoritma sa epsilon istraživanjem naučiti kretanje agenta u ovom okruženju. Za vrednost koraka učenja uzeti 0.8, za metaparametar umenjenja uzeti vrednost 0.9, a za početnu vrednost epsilon parametra 0.1. Broj epizoda za učenje ograničiti sa 2000.