

Mašinsko učenje, Jun2 rok, 3. jul 2018.

Sa adrese

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

preuzeti arhivu **ML_jun2_2018_materijali.zip** koja sadrži Jupyter sveske i materijale potrebne za rad. Na *Desktop*-u napraviti direktorijum sa imenom

ML_jun2_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) Na tržištu se cena dijamanta ocenjuje na osnovu takozvanih **4C** karakteristika: kvaliteta isecanja (eng. cut), čistoće (eng. clarity), boje (eng. color) i finoće koja se izražava u karatima (eng. carat). Potrebno je odrediti regresioni model kojim se na osnovu podataka sadržanih u datoteci *diamonds.csv* može predvideti cena dijamanta.

- a) Učitati podatke koji se nalaze u zadatoj datoteci, a potom izdvojiti podatke koji se nalaze u kolonama *carat*, *cut*, *color*, *clarity* i *price*.
- b) Nacrtati grafike zavisnosti cene dijamanta od svakog od navedenih atributa.
- c) Izvršiti podesne transformacije kategoričkih atributa.
- d) U razmeri 2:1 podeliti podatke na skup za treniranje i skup za testiranje. Za *random_state* parametar uzeti vrednost 7.
- e) Odrediti grebeni regresioni model koji najbolje modeluje zadate podatke. Vrednost *alfa* metaparametra birati iz skupa [0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 1000]. Koji model je najbolji?
- f) Dati ocenu greške najboljeg regresionog modela.

2. (8 poena)

Skup *fashion_mnist* sadrži sličice odevnih predmeta. Potrebno je napraviti klasifikator koji za zadatu sliku odevnog predmeta određuje kojoj kategoriji pripada.

- a) Korišćenjem Keras biblioteke napraviti konvolutivnu neuronsku mrežu čija arhitektura odgovara slici priloženoj uz Jupyter sveske.
- b) Pročitati podatke koji se nalaze u skupu *fashion_mnist* Keras kolekcije, a zatim preduzeti sve neophodne pripremne korake.
- c) Nacrtati grafik zavisnosti tačnosti klasifikatora na skupovima za treniranje i validaciju u odnosu na broj epoha. U toku treniranja mreže koristiti 20 procenata polaznog skupa za validaciju, pakete veličine 64, a za broj epoha uzeti 10.
- d) Kako se menja tačnost mreže kada se nakon svakog agregacionog sloja doda *dropout* sloj sa verovatnoćom isključivanja neurona 0.25? Nacrtati grafik zavisnosti tačnosti klasifikatora na skupovima za treniranje i validaciju u odnosu na broj epoha.
- e) Za koju klasu odevnih predmeta se mreža najgore ponaša?
- f) Sačuvati konfiguraciju mreže u datoteci sa imenom *fashion_mnist*.

3. (6 poena)

Okruženje *MountainCar-v0* OpenIA Gym biblioteke simulira kretanje automobila koji se nalazi u dolini, a koji treba ispeti na desni vrh uvala. Automobil u svakom trenutku karakterišu pozicija i brzina, a radnje koje se mogu preuzeti su pomeranje u levo, pomeranje u desno ili zadržavanje pozicije. Epizoda se završava ukoliko se automobil popne na vrh ili, alternativno, ukoliko se dođe do 200 iteracija. Korišćenjem *Q-learning* algoritma sa epsilon istraživanjem naučiti kretanje automobila. Za vrednost koraka učenja uzeti 0.5, za metaparametar umenjenja uzeti vrednost 0.9, a za početnu vrednost epsilon parametra 0.1.