

# Završni zadatak, PR-P-LV3, PR-R-LV2

## Upute za rješavanje

1. Riješite dane zadatke koristeći programski jezik Python te pripadajuće biblioteke i metode strojnog učenja koje ste učili u okviru laboratorijskih vježbi. Zadatke rješavajte u skripti prezime\_ispit.py.
2. Rješenja zadataka, u obliku skripte prezime\_ispit.py, kao i popunjeni list za odgovore u obliku prezime\_ispit.pdf, uploadajte na Merlin u jednom .zip folderu naziva prezime\_ispit.

**Zadatak 0.0.7** Iris Dataset sastoji se od informacija o laticama i čašicama tri različita cvijeta irisa (Setosa, Versicolour i Virginica). Dostupan je u sklopu biblioteke scikitlearn:

```
from sklearn import datasets  
iris = datasets.load_iris()
```

Upoznajte se s datasetom i dodajte programski kod u skriptu pomoću kojeg možete odgovoriti na sljedeća pitanja:

- a) Prikažite odnos duljine latice i čašice svih pripadnika klase Versicolour pomoću scatter dijagrama. Dodajte na isti dijagram odnos duljine latice i čašice svih pripadnika klase Virginica, drugom bojom. Dodajte naziv dijagrama i nazine osi te legendu. Komentirajte prikazani dijagram.
- b) Pomoću stupčastog dijagrama prikažite prosječnu vrijednost širine čašice za sve tri klase cvijeta. Dodajte naziv dijagrama i nazine osi. Komentirajte prikazani dijagram.
- c) Koliko jedinki pripadnika klase Virginica ima veću širinu čašice od prosječne širine čašice te klase?

**Zadatak 0.0.8** Iris Dataset sastoji se od informacija o laticama i čašicama tri različita cvijeta irisa (Setosa, Versicolour i Virginica). Dostupan je u sklopu biblioteke scikitlearn:

```
from sklearn import datasets
```

```
iris = datasets.load_iris()
```

Upoznajte se s datasetom. Pripremite podatke za učenje. Dodajte programski kod u skriptu pomoću kojeg možete odgovoriti na sljedeća pitanja:

- a) Pronadite optimalni broj klastera  $K$  za klasifikaciju cvijeta irisa algoritmom K srednjih vrijednosti.
- b) Grafički prikažite lakat metodu.
- c) Primijenite algoritam K srednjih vrijednosti koji će pronaći grupe u podatcima. Koristite vrijednost  $K$  dobivenu u prethodnom zadatku.
- d) Dijagramom raspršenja prikažite dobivene klastere. Obojite ih različitim bojama. Centroide obojite crvenom bojom. Dodajte nazine osi, naziv dijagrama i legendu. Komentirajte prikazani dijagram.
- e) Usporedite dobivene klase sa njihovim stvarnim vrijednostima. Izračunajte točnost klasifikacije.

**Zadatak 0.0.9** Iris Dataset sastoji se od informacija o laticama i čašicama tri različita cvijeta irisa (Setosa, Versicolour i Virginica). Dostupan je u sklopu biblioteke `scikitlearn`:

```
from sklearn import datasets
iris = datasets.load_iris()
```

Upoznajte se s datasetom. Podijelite ga na ulazne podatke  $X$  i izlazne podatke  $y$  predstavljene klasom cvijeta. Pripremite podatke za učenje neuronske mreže (kategoričke veličine, skaliranje...). Podijelite podatke na skup za učenje i skup za testiranje modela u omjeru 75:25. Dodajte programski kod u skriptu pomoću kojeg možete odgovoriti na sljedeća pitanja:

- a) Izgradite neuronsku mrežu sa sljedećim karakteristikama:
  - model očekuje ulazne podatke  $X$
  - prvi skriveni sloj ima 10 neurona i koristi `relu` aktivacijsku funkciju
  - drugi skriveni sloj ima 7 neurona i koristi `relu` aktivacijsku funkciju
  - treći skriveni sloj ima 5 neurona i koristi `relu` aktivacijsku funkciju
  - izlazni sloj ima 3 neurona i koristi `softmax` aktivacijsku funkciju.između prvog i drugog te drugog i trećeg sloja dodajte Dropout sloj s 30% izbačenih neurona  
Ispišite informacije o mreži u terminal.
- b) Podesite proces treniranja mreže sa sljedećim parametrima:
  - **loss** argument: `categorical_crossentropy`
  - **optimizer**: `adam`
  - **metrika**: `accuracy`.
- c) Pokrenite učenje mreže sa proizvoljnim brojem epoha (pokušajte sa 500) i veličinom `batch-a` 7.
- d) Pohranite model na tvrdi disk te preostale zadatke izvršite na temelju učitanog modela.
- e) Izvršite evaluaciju mreže na testnom skupu podataka.
- f) Izvršite predikciju mreže na skupu podataka za testiranje. Prikažite matricu zabune za skup podataka za testiranje. Komentirajte dobivene rezultate i predložite kako biste ih poboljšali, ako je potrebno.