

Završni zadatak, PR-R-LV1

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 0.0.10 Datoteka winequality-red.csv sadrži kemijska mjerena crnih vina i njihovu kvalitetu. Upoznajte se s datasetom. Više informacija nalazi se u datoteci winequality.names.

Dodajte programski kod u skriptu pomoću kojeg možete odgovoriti na sljedeća pitanja:

- a) Na koliko je crnih vina provedeno mjerjenje?
- b) Prikažite stupčastim dijagramima (na jednoj slici) ovisnost kvalitete vina o alkoholnoj jakosti. Dodajte naziv dijagrama i nazine osi. Interpretirajte rezultate prikazane dijagramom.
- c) Koliki broj uzoraka vina ima kvalitetu manju od 5, a koliki ima 5 i veću?
- d) Izračunajte i prikažite korelaciju svih veličina dostupnih u datasetu. Interpretirajte rezultate.

Zadatak 0.0.11 Datoteka winequality-red.csv sadrži kemijska mjerena crnih vina i njihovu kvalitetu. Upoznajte se s datasetom. Više informacija nalazi se u datoteci winequality.names. Učitajte dane podatke. Podijelite skup na ulazne podatke X i izlazne podatke y predstavljene kvalitetom vina. Zamijenite sve vrijednosti kvalitete vina manje od 5 s 0, a one koje imaju vrijednost 5 ili veću s 1 kako biste dobili dvije izlazne klase. Podijelite podatke na skup za učenje i skup za testiranje modela u omjeru 80:20. Standardizirajte podatke. Dodajte programski kod u skriptu pomoću kojeg možete odgovoriti na sljedeća pitanja:

- a) Izgradite linearni regresijski model. Ispišite parametre modela.
- b) Izvršite procjenu izlazne veličine na temelju ulaznih veličina skupa za testiranje. Prikažite pomoću dijagrama raspršenja odnos između stvarnih vrijednosti izlazne veličine i procjene dobivene modelom. Interpretirajte dobivene rezultate.

- c) Izvršite vrednovanje modela na način da izračunate vrijednosti regresijskih metrika (*RMSE*, *MAE*, *MAPE* i R^2) na skupu podataka za testiranje. Interpretirajte dobivene rezultate.

Zadatak 0.0.12 Datoteka `winequality-red.csv` sadrži kemijska mjerena crnih vina i njihovu kvalitetu. Upoznajte se s datasetom. Više informacija nalazi se u datoteci `winequality.names`. Učitajte dane podatke. Podijelite skup na ulazne podatke X i izlazne podatke y predstavljene kvalitetom vina. Zamijenite sve vrijednosti kvalitete vina manje od 5 s 0, a one koje imaju vrijednost 5 ili veću s 1 kako biste dobili dvije izlazne klase. Podijelite podatke na skup za učenje i skup za testiranje modela u omjeru 80:20. Pripredite podatke za učenje.

- a) Izgradite neuronsku mrežu sa sljedećim karakteristikama:
 - model očekuje ulazne podatke X
 - prvi skriveni sloj ima 20 neurona i koristi `relu` aktivacijsku funkciju
 - drugi skriveni sloj ima 12 neurona i koristi `relu` aktivacijsku funkciju
 - treći skriveni sloj ima 4 neurona i koristi `relu` aktivacijsku funkciju
 - izlazni sloj ima jedan neuron i koristi `sigmoid` aktivacijsku funkciju.Ispišite informacije o mreži u terminal.
- b) Podesite proces treniranja mreže sa sljedećim parametrima:
 - **loss** argument: `binary_crossentropy`
 - **optimizer**: `adam`
 - **metrika**: `accuracy`.
- c) Pokrenite učenje mreže sa proizvoljnim brojem epoha (pokušajte s 1000) i proizvoljnom veličinom `batch-a` (pokušajte s 50).
- 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. Interpretirajte dobivene rezultate.