

Završni zadatak, PR-P-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 0.0.1 Datoteka pima-indians-diabetes.csv sadrži mjerena provedena u svrhu otkrivanja dijabetesa, pri čemu se u devetom stupcu nalazi klasa 0 (nema dijabetes) ili klasa 1 (ima dijabetes). Učitajte dane podatke u obliku numpy polja data. Dodajte programski kod u skriptu pomoću kojeg možete odgovoriti na sljedeća pitanja:

- a) Na temelju veličine numpy polja data, na koliko osoba su izvršena mjerena?
- b) Postoje li izostale ili duplicitne vrijednosti u stupcima s mjerjenjima dobi i indeksa tjelesne mase (BMI)? Obrišite ih ako postoje. Koliko je sada uzorka mjerena preostalo?
- c) Prikažite odnos dobi i indeksa tjelesne mase (BMI) osobe pomoću scatter dijagrama. Dodajte naziv dijagrama i nazive osi s pripadajućim mjernim jedinicama. Komentirajte odnos dobi i BMI prikazan dijagramom.
- d) Izračunajte i ispišite u terminal minimalnu, maksimalnu i srednju vrijednost indeksa tjelesne mase (BMI) u ovom podatkovnom skupu.
- e) Ponovite zadatak pod d), ali posebno za osobe kojima je dijagnosticiran dijabetes i za one kojima nije. Kolikom je broju ljudi dijagnosticiran dijabetes? Komentirajte dobivene vrijednosti.

Zadatak 0.0.2 Datoteka pima-indians-diabetes.csv sadrži mjerena provedena u svrhu otkrivanja dijabetesa, pri čemu se u devetom stupcu nalazi izlazna veličina, predstavljena klasom 0 (nema dijabetes) ili klasom 1 (ima dijabetes).

Učitajte dane podatke u obliku numpy polja data. Podijelite ih na ulazne podatke X i izlazne

podatke y . Podijelite podatke na skup za učenje i skup za testiranje modela u omjeru 80:20. Dodajte programski kod u skriptu pomoću kojeg možete odgovoriti na sljedeća pitanja:

- a) Izgradite model logističke regresije pomoću scikit-learn biblioteke na temelju skupa podataka za učenje.
- b) Provedite klasifikaciju skupa podataka za testiranje pomoću izgrađenog modela logističke regresije.
- c) Izračunajte i prikažite matricu zabune na testnim podacima. Komentirajte dobivene rezultate.
- d) Izračunajte točnost, preciznost i odziv na skupu podataka za testiranje. Komentirajte dobivene rezultate.

Zadatak 0.0.3 Datoteka `pima-indians-diabetes.csv` sadrži mjerjenja provedena u svrhu otkrivanja dijabetesa, pri čemu je prvih 8 stupaca ulazna veličina, a u devetom stupcu se nalazi izlazna veličina: klasa 0 (nema dijabetes) ili klasa 1 (ima dijabetes).

Učitajte dane podatke. Podijelite ih na ulazne podatke X i izlazne podatke y . Podijelite podatke na skup za učenje i skup za testiranje modela u omjeru 80:20.

- a) Izgradite neuronsku mrežu sa sljedećim karakteristikama:
 - model očekuje ulazne podatke s 8 varijabli
 - prvi skriveni sloj ima 12 neurona i koristi `relu` aktivacijsku funkciju
 - drugi skriveni sloj ima 8 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: `cross_entropy`
 - **optimizer**: `adam`
 - **metrika**: `accuracy`.
- c) Pokrenite učenje mreže sa proizvoljnim brojem epoha (pokušajte sa 150) i veličinom batch-a 10.
- 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.