13M051MU, 4. domaći zadatak 2023/24
Izbor odlika. Stabla. Ansambli.

Podaci sa kojima radite su u datotekama `data_1.csv` i `data_2.csv`. Poslednja kolona označava klasu, a ostale kolone sadrže vrednosti prediktora.

1 Izbor prediktora

Za podatke iz `data_1.csv`, treba da sortirate prediktore od najboljeg do najgoreg, na 2 različita načina:
1) na osnovu koeficijenata korelacije sa ciljnom promenljivom,
2) pomoću "omotač-algoritma", sa klasifikatorom po želji.

Osmislite i nacrtajte grafik koji ilustruje kvalitete prediktora na po jednoj slici, za svaku od navedenih metoda.

2 Obučavanje stabla

Za jedan par prediktora iz `data_1.csv` obučite klasifikaciono stablo. Dozvoljeno je korišćenje ugrađenih f-ja i klasa: `sklearn.tree.DecisionTreeClassifier` u Pythonu, odnosno `ClassificationTree` u Matlabu.

Osmislite i nacrtajte grafike koji demonstriraju uticaj maksimalne dubine (ili maksimalnog broja čvorova, ako biblioteka koju koristite ne omogućava zadavanje maksimalne dubine) na kvalitet dobijenog modela. Treba da ilustrujete 3 različita slučaja: kada je model podobučen, kada je preobučen, i kada nije ni podobučen ni preobučen. Prikažite i granice odlučivanja u ravni prediktora za svaki od ova 3 slučaja.

3 Ansambli

Na podacima iz `data_2.csv` ispitajte kako hiper-parametri utiču na performanse sledeća dva algoritama: 1) Random Forest (RF), 2) Gradient Boosting (GB).

U oba algoritma koristite stabla odlučivanja kao članove ansambla. Dozvoljeno je korišćenje ugrađenih f-ja za obučavanje klasifikatora: `fitcensemble` (sa `Method = 'LogitBoost'`) u Matlabu, odnosno `RandomForestClassifier` i `GradientBoostingClassifier` u Pyhon modulu `sklearn.ensemble`.

Osmislite i nacrtajte grafike koji ilustruju uticaj hiper-parametara na kvalitet obučenog modela. Minimalan skup hiper-parametara koje treba da analizirate je
- veličina ansambla,
- veličina stabala,
- maksimalan broj odlika koji se razmatraju pri dodavanju čvora stabla (za RF),
- stopa učenja (za GB).

Po završenom treningu, grafički prikažite estimirane značajnosti prediktora za bar jedan od dva analizirana algoritma.

Kôd u Pythonu ili Matlab/Octave, i izveštaj sa traženim graficima u pdf formatu, predaje se putem MS Teamsa. Ako predajete `ipynb`, uz njega dostavite i `html`. *Ne zaboravite da kliknete na Turn In!*