

# DOMAĆI 2

## 1. ZADATAK 1: PCA analiza

Prvi deo zadatka odnosi se na PCA analizu. Prikupljeni su podaci za 50 država u Americi o broju hapšenja na 100 000 stanovnika za ubistva, napade i silovanja. Dat je takođe i podatak o procentu populacije u urbanim zonama.

- Primeniti PCA na dati skup podataka.
- Vizuelizovati podatke pomoću biplot-a (pogledati funkciju u MATLAB-u).
- U izveštaju analizirati dobijeni biplot, kao i zavisnost dela objašnjene varijanse od broja PCA komponenti.

## 2. ZADATAK 2: kNN klasifikacija (praznični zadatak)

Napraviti klasifikator koji će na osnovu prisustva/odsustva određenih sastojaka odrediti da li je dati recept za pravljenje kolačića (*Cookies*), peciva (*Pastries*) ili pice (*Pizzas*). Prikupljeno je 1738 recepata i za svaki je naznačeno prisustvo (1) ili odsustvo (0) svakog od 133 sastojka.

Na odbrani će klasifikator biti testiran na određenom broju recepata, a student čiji klasifikator postigne najbolji rezultat, dobiće 3 bonus boda. Ukoliko bude postojalo više istih rešenja, bonus bodovi neće biti dodeljeni.

- Napraviti kNN klasifikator koristeći sva obeležja na sledeći način:
  - Odvojiti deo podataka (15%) za testiranje, pazeći pritom da od svake klase ima dovoljno uzoraka u test skupu.
  - Parametar  $k$  (broj najbližih suseda) varirati od 1 do 10. Da li možete još nešto da varirate?
  - Izbor modela: krosvalidacijom odabrati najbolji kNN klasifikator (u izveštaju bi trebalo da se navedu vrednosti za tačnost za svako korišćeno  $k$  ili svaku korišćenu kombinaciju parametara).
  - Obučiti model nad svim trening podacima, potom nad test podacima izračunati tačnost, preciznost, odziv i specifičnost za svaku od klasa, kao i prosečne mere za celokupan klasifikator.
- Nad čitavim setom podataka za svaku od klasa napraviti prosečan obrazac pojavljivanja sastojaka (za svaki sastojak naći koliko se često pojavljuje u receptima date klase) i prikazati na 3 odvojena histograma.
- Rezultate diskutovati u izveštaju.

\*korisne MATLAB funkcije: *fitcknn*, *crossval*, *cvpartition*, kao i primeri u *help-u*.

