

Metode i primjene vještačke inteligencije

Lab 2 Izvještaj

Haris Dajic
Broj indeksa: 1543/17198

November 2019

1 Implementacija KNN algoritma

1. Parametri koji se šalju su sljedeći:
 - data - Matrica sa dvije kolone, u našem slučaju integer i float tipovi
 - query - tačka za koju ispituje
 - k - broj koliko tačaka uzimamo u obzir
 - funkcija
 - funkcija
2. Funkcija na izlazu vraća dva elementa, prvi je vektor parova, udaljenost između tačaka i indeks (pozicija) tačke u ulaznom vektoru, a druga vrijednost je vrijednost labele tačaka kojih ima najviše u prethodno spomenutom vektoru.
3. Izračunava se kao udaljenost između dvije tačke.
4. Koristi se lista parova
5. Udaljenost između ulazne tačke i naših trening podataka i indeks
6. Onoliko koliko imamo u našem dataset-u podataka
7. Algoritam za sotritanje koji koristi funkcija sorted je Timsort
8. Prvih k elemenata iz varijable neighbor_distances_and_indices, tačnije k najbližih tačaka
9. Urađeno u fajlu data.py gdje pozivamo funkcije koje čitaju iz csv datoteke.
10. Urađeno u funkciji classification u fajlu knn.py
11. Prvo za svaku tačku iz naše tabele (csv datoteke) računamo distancu u odnosu na ulaznu tačku (33) prema sljedećoj formuli $\sqrt{(x - y)^2}$. Npr.
$$\sqrt{(34 - 33)^2} = 1.$$
Nakon što smo izračunali udaljenost za sve tačke, tada imamo listu parova čiji elementi predstavljaju udaljenost i indeks elementa iz tabele. Tu listu ćemo sortirati od manjeg ka većem poredeći distance. Nakon toga uzimamo prvih $k = 3$ elemenata. Labelu koju ćemo dodijeliti našoj tački jeste ona koja se najviše puta ponavlja u prethodno odabranih k tačaka.
12. Već odgovoreno u 9-tom pitanju.
13. Urađeno u funkciji regresion u fajlu knn.py
14. Kao i kod pitanja pod rednim brojem 11., na isti način računamo tačke, jedina razlika je kod dodjeljivanja labela. Kod regresije dodjelu labela radimo na način da k najbližih tačaka koje smo odredili saberemo i podijelimo sa njihovom dužinom tj, sa k . Vrijednost koju smo dobili jeste naša tražena labela. K najbližih tačaka koje dobijemo za tačku 150.4 i $k = 3$ jesu [51.58, 45.99, 64.32]. Nakon što saberemo ove tri tačke i podijelimo ih sa njihovim brojem, tj sa 3 dobijemo labelu koja odgovara traženoj tački, u ovom slučaju 53.96333333333333.