

Tema A3: Aproksimacija cena kuća u Bostonu

Na formiranje cene kuće utiče veliki broj faktora: blizina autoputa, udaljenost od grada, stepen zagađenosti, stopa kriminala i dr. Jedno takvo istraživanje, koje prati uticaj 13 numeričkih faktora na formiranje cene stambenih objekata, sprovedeno je u Bostonu. Vrednosti faktora uzimane su na nivou različitih delova grada, tj. predstavljaju prosečne vrednosti faktora za sve kuće u tom delu grada. U nastavku će biti korišćen mali podskup skupa podataka koji je dostupan na <https://scikit-learn.org/stable/datasets/index.html#boston-house-prices-dataset>.

U tabeli 1 nalaze se prosečne vrednosti sledećih faktora:

- stepen zagađenosti, tj. koncentracija azot-monoksida (NOX),
- težinska udaljenost od glavnih poslovnih centara (DIS),
- procenat siromašog dela populacije (LSTAT).

Poslednja kolona MEDV (Median Value) je kolona prosečnih vrednosti cena kuća iz istog dela grada u \$1000 .

| Index | NOX | DIS | LSTAT | MEDV |
|-------|--------|--------|-------|------|
| 0 | 0.5240 | 6.2267 | 13.27 | 18.9 |
| 1 | 0.7400 | 2.0026 | 18.05 | 9.6 |
| 2 | 0.6310 | 1.3325 | 3.26 | 50.0 |
| 3 | 0.6090 | 1.7554 | 23.97 | 7.0 |
| 4 | 0.5100 | 2.9153 | 6.92 | 29.9 |
| 5 | 0.8710 | 1.7494 | 15.12 | 17.0 |
| 6 | 0.6050 | 1.9709 | 1.73 | 50.0 |
| 7 | 0.4580 | 6.0622 | 5.33 | 36.2 |
| 8 | 0.4880 | 2.5979 | 9.45 | 36.2 |
| 9 | 0.7400 | 1.8206 | 34.02 | 8.4 |
| 10 | 0.6140 | 1.9512 | 13.11 | 21.4 |
| 11 | 0.7400 | 1.8956 | 23.79 | 10.8 |
| 12 | 0.7400 | 2.1980 | 16.44 | 12.6 |
| 13 | 0.4280 | 5.4011 | 4.32 | 30.8 |
| 14 | 0.6710 | 1.3163 | 13.44 | 15.0 |
| 15 | 0.5380 | 4.1750 | 13.04 | 14.5 |
| 16 | 0.4429 | 5.2447 | 4.59 | 35.4 |
| 17 | 0.5850 | 2.3999 | 15.10 | 17.5 |
| 18 | 0.6050 | 2.1000 | 11.10 | 23.8 |
| 19 | 0.6240 | 2.1107 | 16.96 | 18.1 |
| 20 | 0.4530 | 7.2255 | 6.73 | 22.2 |
| 21 | 0.4530 | 7.9809 | 9.50 | 25.0 |
| 22 | 0.5240 | 5.5605 | 12.43 | 22.9 |
| 23 | 0.4490 | 4.4272 | 12.86 | 22.5 |
| 24 | 0.4030 | 7.6534 | 4.45 | 32.9 |

Tabela 1: Rezultati 25 naseljenih delova grada

Zadatak:

1. Naći polinom prvog stepena koji u srednjekvadratnom smislu najbolje aproksimira zavisnost cene kuće od faktora DIS.
2. Naći polinom prvog stepena koji u srednjekvadratnom smislu najbolje aproksimira zavisnost cene kuće od faktora DIS i NOX. Koja je razlika u odnosu na aproksimaciju dobijenu pod 1?
3. Naći polinom drugog stepena koji u srednjekvadratnom smislu najbolje aproksimira zavisnost cene kuće od sva 3 faktora.
4. Uporediti dobijene aproksimacije.

Podaci iz tabele 1 dostupni su na <http://www.matf.bg.ac.rs/p/files/82-boston.csv>.