

3. domača naloga

Tlen Venko (63140280)

30. november 2016

1 Uvod

Cilj tretje naloge je bil, da na podlagi podatkov o vožnjah mestnih avtobusov iz obdobja januar - november 2012, napovemo trajanje vožnje avtobusov za mesec december 2012.

2 Podatki

Podatki na katerih se je program učil se nahajajo v datoteki train.csv.gz. Podatki obsegajo podatke vožnje o 28 mestnih avtobusnih linijah v obdobju od 1.1.2012 do 30.11.2012. Datoteka zavzema podatke o registrski tablici avtobusa, IDju voznika, Stevilki proge, začetna in končna postaja, ter začetni in končni čas vožnje. Za nalogo so bile pomembni podatki številka proge, začetna in končna postaja in pa čas začetka in konca vožnje.

3 Metode

3.1 Ocenjevanje točnosti

Ocenjevanje pravilnosti moje napovedi sem realiziral z MAE in RMSE v razredih MAE in RMSE. Razreda sprejmeta vrednosti, ki sem jih z linearno regresijo pridelal jaz in pa realne podatke enega meseca iz učne množice (učim se na enem mesecu manj in na njem izvajam teste). ZA MAE seštejem absolutne vrednosti razlik napovedi in jih delim s številom napovedi. Pri RMSE pa seštejem kvadrate razlik napovedi, jih delim s številom napovedi in vse skupaj korenim.

3.2 Napovedni model

model Funkcija, ki predela podatke tako, da jih lahko sprejme linearna regresija. Učimo se samo na podatkih iz novembra in ločimo med sabo uro dneva in pa dan v tednu.

LinearLearner Funkcijo LinearLearner sem predelal tako, da sprejema vse podatke in se nato v njej gradi model X, tako da kliče funkcijo model() in pa vektor y.

loci_po_mesecu Funkcija, ki loci podatke na katerih se učimo tako, da izloči mesec november in se uči samo na podatkih od januarja do oktobra, november pa nam služi za interno testiranje.

read_file Funkcija za branje vhodnih podatkov.

main Najprej preberemo testne podatke, nato pa zgradimo model, tako da ločimo linije med sabo z razredom `SeparateBySetLearner`, linearno regresijo pa delamo s `LinearLearner`. Rezultate nato zapišemo v datoteko. Sledi še interno preverjanje, kjer se učimo na podatkih od januarja do oktobra in jih preverjamo na novembru z metodo RMSE in MAE.

4 Rezultati

Tabela 1: Rezultati			
ime metode	oddaja	ocena s preverjanjem na učnih podatkih	ocena na tekmovanju
celo_letno	2016-11-29 23:47:53	170.8	210.77369
*pretekli_mesec	2016-11-30 01:04:09	184.26	191.43976
polinom8	2016-11-30 10:42:27	175.7	207.80280
polinom5	/	176.84	/

Rezultati za pretekli mesec so na internem testiranju izdelane na oktobru in ne na novembru, kot pri pravem testiranju.

5 Izjava o izdelavi domače naloge

Domačo nalogo in pripadajoče programe sem izdelal sam.