

# Naloga 5 - Napovedovanje prihodov avtobusov

## LPP

Melanija Kraljevska

10 December 2019

### 1 Predtekmovanje

**1.1 predtekmovanje1.py - 149.53986**

**1.2 predtekmovanje2.py - 140.79592**

**1.3 predtekmovanje3\*.py - 138.46979**

Vsak primer iz učnih podatkih je predstavljen z vsako vstico iz matrike X. Pri izdelavi matrike X je uporabljen čas odhoda avtobusa, in dodatne informacije (prazniki, vremenska prognoza). Ena vrstica vsebuje 489 atributov: prvih 480 atributov so ure razdeljene na intervalih po 3 minute (00:03, 00:6, 00:09,...), naslednji 7 atributov vsebujejo kateri dan v tednu je, naslednji pa vsebuje informacijo ali je tisti dan praznik. Vse do sedaj omejene attribute so binarne, torej vsebujejo 0 ali 1. Matrika X v tem primeru je nadopolnjena s tem da v prvih 288 stolpcih je čas razdeljen na intervale od 5 minut, potem vsebuje podatke kateri dan v tednu je, pa ali je praznik. Naslednji stolpec vsebuje 0 ali 1, ki predstavljajo dez/sneg v tistem dnevu - v primeru ko je dez nad 5mm ali sneg nad 5cm je 1, sicer 0 (vir: [meteo.arso.gov.si](http://meteo.arso.gov.si)). Ciljna spremenljivka y predstavlja trajanje vožnje od začetne do končne postaje.

### 2 Tekmovanje

Pri učnih podatkih, kjer imamo podatke o različnih linij, za vsako linijo je izgrajen model - linearna (polinomska) regresija. Pri branju podatkov je na začetku narejen slovar ki na podlagi imena rute avtobusa (stolpec Route), vrne čas odhoda in prihoda za določeno linijo. Z izbiranjem Route upoštevamo smer pa tudi ločimo med več linij z isto številko (primer: 6 in 6B).

**2.1 tekMOVanje1.py - 184.11504**

Vsak primer iz učnih podatkih je predstavljen z vsako vstico iz matrike X. Pri izdelavi matrike X je uporabljen čas odhoda avtobusa, in dodatne informacije

(prazniki, vremenska prognoza). Ena vrstica vsebuje 58 atributov: prvih 48 atributov so ure razdeljene na pol ure (00:00, 00:30, 01:00,...), naslednji 7 atributov vsebujejo kateri dan v tednu je, naslednji pa vsebuje informacijo ali je tisti dan praznik. Vse do sedaj omejene attribute so binarne, torej vsebujejo 0 ali 1. Ciljna spremenljivka  $y$  predstavlja trajanje vožnje od začetne do končne postaje.

## **2.2 tekmovanje2.py - 174.15551**

Matrika  $X$  v tem primeru je nadopolnjena s tem da v prvih 288 stolpcev je čas razdeljen na intervale od 5 minut, potem vsebuje podatke kateri dan v tednu je, pa ali je praznik. Naslednji stolpec vsebuje 0 ali 1, ki predstavljajo dez/sneg v tistem dnevu - v primeru ko je dez nad 5mm ali sneg nad 5cm je 1, sicer 0 (vir: [meteo.arso.gov.si](http://meteo.arso.gov.si)).

## **2.3 tekmovanje3\*.py - 171.97208**

Matrika  $X$  je ista kot pri tekmovanje2.py, samo je čas razdeljen na interval od 3 minute.