

Domača naloga - 1.del

Časovne vrste

Vito Rozman

21. junij 2023

1 Predstavitev uporabljenih metod

Pri iskanju njaboljšega modela napovedovanje časovnih vrst sem testiral različne vrste rekurenčnih nevronske mreže (RNN). Najprej sem začel z navadnim RNN, ki smo ga zgradili na vajah. Model imam eno skrito stanje, ki služi spominu preteklih primerov. Nato sem testiral model kjer je uporabljena GRU celica, ta nam omogoča da se model dinamično odloča kaj naj si zapomne in kaj naj pozabi. Na zadnje še nevronske mreže z imenom Long short-term memory (LSTM), ki omogoča uporabo kratkoročnega spomina in dolgoročnega spomina za grajenje prihodnjih napovedi. LSTM model se je izkazal za najboljšega.

Pri samih podatkih sem poizkusil več možnih pretvorb za boljše učenje.

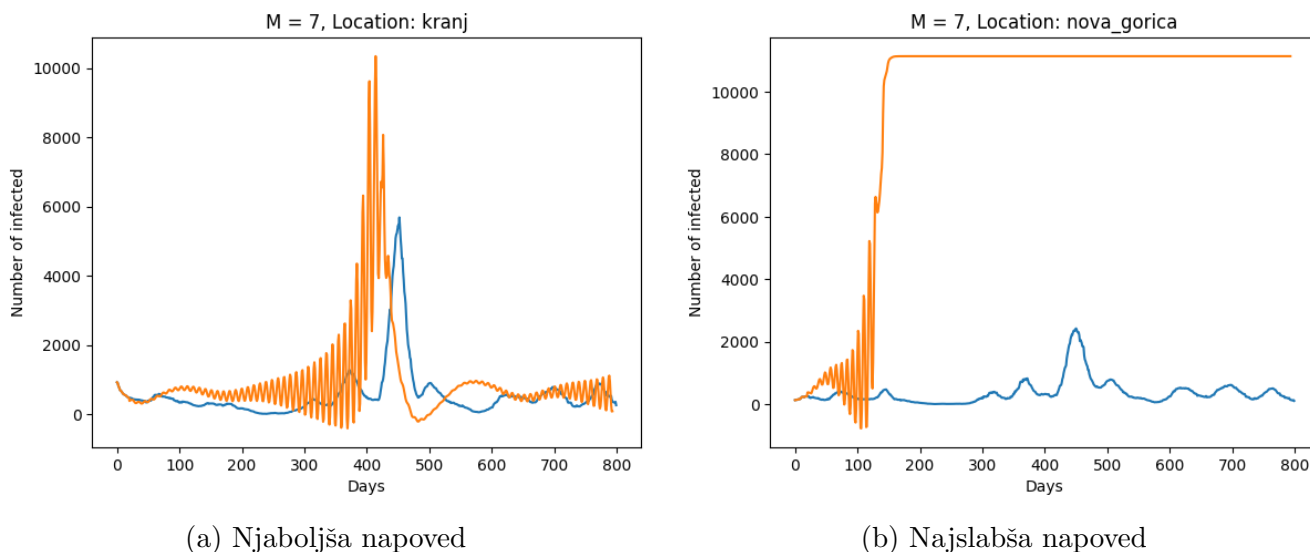
- Navadni podatki
- Diference med dnevi (število novo ukuženih)
- Logaritmična skala števila ukuženih
- Skalirane podatke z MinMaxScaler

Najbolji pristop je bil skaliranje podatkov z MinMaxScaler, kar sem tudi ugotovil pri iskanju modelov, ki že obstajajo za napovedovanje okuženih. Za LSTM je priporočljivo da so podatki med -1 in 1 zaradi vrat ki nastopajo v arhitekturi za prepust novih podatkov in pozabo prejšnjih.

2 Učenje, testiranje in izbira modela

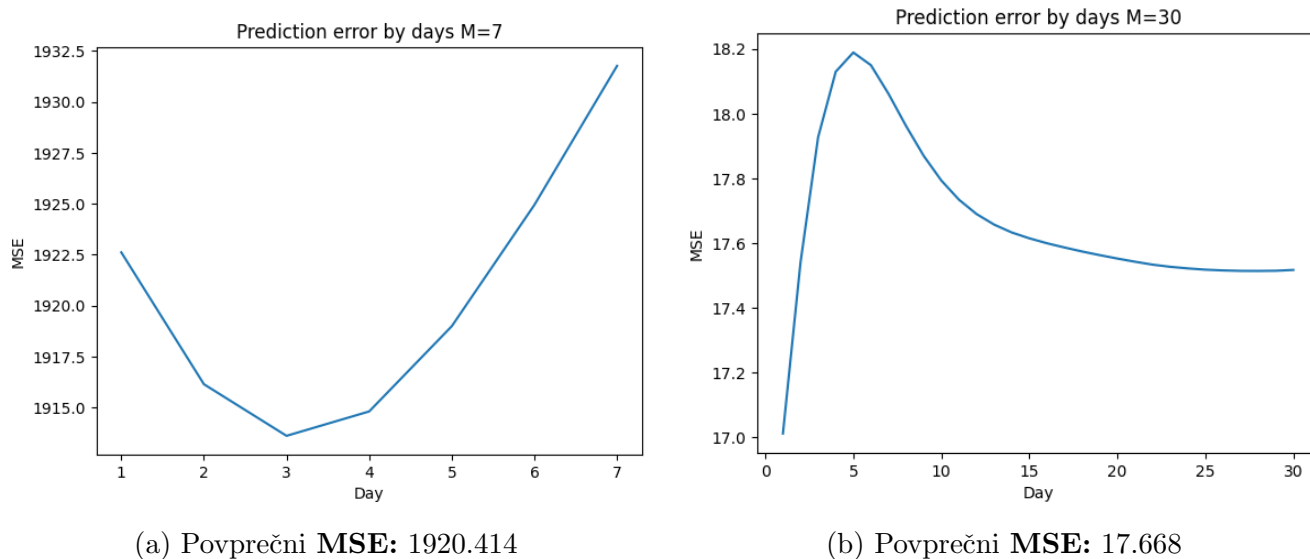
Model sem učil na skaliranih podatkih za vsako mesto posebaj. Najprej sem naredil učno zanko čez vseh 800 dni potem pa še testno čez vse dni. Uporabil se pristip s šaržami (ang. batch), kar pomeni da sem ob vzratnem posodabljanju uteži najprej izračunal napako za primere v šarži in šele na to posodobil uteži. Dan pristop se je izkazal za veliko boljše in hitrejše kot brez šarž. V spodnji tabeli so navedeni hiperparametri ki sem jih preizkusil za izbran model in izbiro parametra za vsakega od modelov napovedovanja (7 dni oz. 30 dni).

Hiperparameter	Prostor hiperparametrov	M=7	M=30
N	{7, 10, 14, 20, 30, 40}	20	30
epoch	{10, 100}	100	100
batch	{8, 16, 32, 64, 128}	32	32
hidden size	{20, 32, 64}	32	64
learning rate	{0.1, 0.01, 0.001, 0.0001}	0.001	0.001



Slika 1: Napovedi modela

Ko sem računal napako modela na testni in na učni množici sem uporabil MSE (Mean squared error). Čudno se mi je zdelo, da je napak za napoved 7 dni tako velika v primerjavi z napako za 30 dni, kljub temu, da izgledajo grafi napovedi za 7 dni veliko boljše. Na sliki 1 je izris napovedi modela (z oranžno) in prava vrednost okuženih (modra). Kot njaboljša napoved se je izkazala pri občini Kranj, najslabš pa občina Nova gorica. Na sliki 2 je izris povprečne napake odvisnosti od oddaljenosti napovedi.



Slika 2: Napaka modele glede na oddaljenost napovedi

Možne izboljšave

Za grajenje boljšega modela bi lahko sestavil model ki bi napovedoval vse občine hkrati, ali pa vsako zaporedje vhodnih primereov pridobil iz drugega kraja. Torej šarže bi generirala iz vsakega kraja takao da bi vsaka šarža vsebovala 12 vhodni zaporedij za učenje. Tako bi model ob posodobitvi videl več primernih primerov.