Predicting hospital beds for COVID-19 patients

Team: Lara Du, Julian Rosen, Kai Wei

Building a model to predict the future number of hospital beds needed for COVID-19 patients based on current usage, vaccination rates and the number of new cases

https://github.com/czkaiweb/ErdosProject-COVID/

The Problem and Goals

- Facts: The COVID-19 pandemic has been wreaking havoc in the US since early 2020. Most epidemiologists agree that it is likely to become endemic.

- Problem: In most cases, when COVID-19 simply behaves like a cold or flu, it doesn't cause much disruption. However each time COVID-19 spikes there is a risk of the healthcare system becoming overwhelmed if more patients need hospitalisation than the number of beds available.

- Goal: Create a model to predict the number of hospital beds needed by COVID-19 patients up to 14 days in advance.

Our Approach

- Data gathering and cleaning: We downloaded data from the CDC website https://data.cdc.gov/ where we combined the datasets giving cases, hospitalizations and vaccine rates by state. We then cleaned our data to remove territories outside the fifty states, combine multiple data from the same day and remove dates from before vaccinations started.

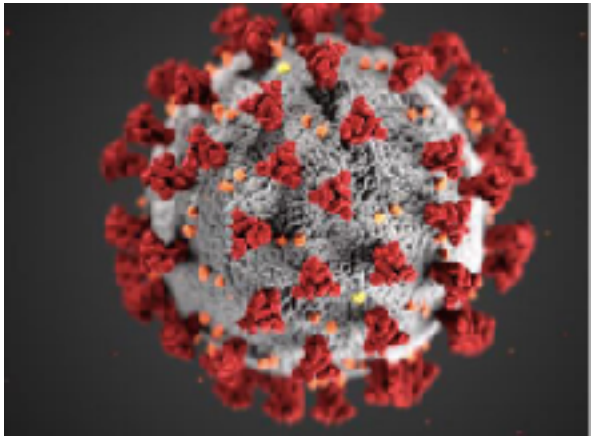
- Supervised learning: We tried linear regression, ridge regression, vector autoregression, XG Boost and GreyKite. The model that performed the best combined ridge regression with an ARIMA model.

Our optimal model

- The baseline model predicts that the number of hospital beds used on any future day is the same as the number of hospital beds today.
- The ARIMA model predicts a time series. It takes in a series of values and predicts the continuation of that series.
- In the first stage of our model we trained an ARIMA model to predict the number of future hospital beds used in each state purely based on the historical number of beds used.
- We also trained a ridge regression model, using the ARIMA prediction, the number of current beds used and the daily COVID-19 case rate as features.
- We found our optimal model to be the average of the baseline, the ARIMA and the ridge model.

Methods/Packages Used

- sklearn (statistical analysis)
- statsmodels (ARIMA and VAR)
- Greykite



Our approach to training and evaluating models

- We tuned hyperparameters using only data through 2021-07-31.
- We isolated data for each state.
- For each day between 2021-08-14 and 2021-09-30, we made 14 predictions coming from each of the previous 14 days. Each prediction for a given day was for the number of hospital beds used per 100,000 residents.
- We computed the mean squared error for all predictions, averaged over all states.
- The smaller this average was, the better we deemed the model to be.

Results

- We first ran a simple linear model to predict the number of COVID-19 hospital beds. We predicted the number of beds needed on a given day in the future as a linear function of the number of cases and beds today. The error in this model was 16.16.
- Our baseline model gave an error of 11.59.
- The ARIMA model gave an error of 7.45.
- The ridge model using ARIMA as a feature gave an error of 9.54.
- Our optimal model that averaged over the baseline, ARIMA and ridge models gave an optimal error of 6.13.

Challenges

- The trajectory of the COVID-19 pandemic is constantly changing. A model that worked well during one time period often worked poorly a few months later.
- There was only a modest amount of data available: for each state, there are only a few hundred daily data points since vaccinations began.
- Several models that we looked at were very inaccurate:
    - Vector autoregression: it did just marginally better than the baseline but worse than ridge regression
    - Greykite: it underperformed the baseline model (worse mean squared error)
    - LSTM: it gave a prediction which was hard to interpret. It always underestimated the true number, predicting very few hospital beds used each day.

Next Steps

- Find better models. Investigate fitting models against additional features given in the CDC data
  - Medicine/drug supplies
  - COVID-19 deaths
  - Vaccine data broken down by brand

- Update our webpage
  - Click on states to see predictions for hospital bed usage by county within that state
  - Add graphs that show historical predictions by our model vs actual data as time progresses
  - Add new models with graphs so that users may compare and average the different models

Thank You!

Team Octopus: Predicting hospital beds for COVID-19 patients

https://github.com/czkaiweb/ErdosProject-COVID/

Lara Du
du290@purdue.edu

Purdue University

Julian Rosen
julianrosen@gmail.com

University of Maine

Kai Wei
wei.856@osu.edu

The Ohio State University