STAT 390 Weekly Report 5

Nov 6-10

Group 2: Cindy Ha, Willie Xie, Erica Zhang

Table of contents

1 Progress/Accomplishments

- Each member worked on model building and hyperparameter tuning:
  - **arima** & **auto-arima**:
    - Checked countries that have stationary data (17 countries) vs. non-stationary data (6 countries) [`Models/erica/lm_stationary_check.R`]
    - Used `arima_reg()` initially to tune the model for stationary rep (U.S.) and non-stationary rep (Germany) but prediction is bad (a straight line) even after using first-difference to remove trend;
    - Did a manual grid search –> gives p, d, p order combination of (0, 0, 0), suggesting there's no pattern; automatic grid search using auto.arima is giving better combination in terms of lower AIC
    - Also looked at ACF and PACF plots –> still gave a white noise model of (0, 0, 0)
    - Pivoted by using linear model first to model the trend and arima to model the predicted error
  - **prophet single** & **prophet multiple**:
    - both works well! tune the model using U.S. data and apply the model to rest of 22 countries
  - **xgboost**:

    - the model works well but just need to further improve the hyperparameters and maybe reduce the tree to reduce running time
  - **lstm**:
    - Have a rough outline of model spec, recipe, and tuning but still exploring how the model works

2 Challenges

- Long run times when tuning hyperparameters
  - Even with parallel processing, it can take anywhere from 1-12 hours depending on model and cores on device
- Trying to troubleshoot and improve arima model, as explained above
- Difference in LSTM model (adding layers) compared to the other models is a bit confusing so we need to explore further

3 Next Steps

- Continue working on hyperparameter tuning and model building for LSTM.

We are making steady progress and will be able to present our 5 models: Arima, Auto Arima, Prophet Single, Prophet Multiple, and XGBoost.