

Dengue Exercise (Part 2)

In the previous exercise you have fit a static model to predict dengue cases using training data from 2004-2006. This means that you learn a single, unchanging model which you use to predict future cases. In this exercise, you will fit a dynamic model that is retrained for each prediction step. You will then add more features that will turn your model to a basic ARGO model.

- (a) In a *sliding window* approach, the original training period (Jan. 2004 - Dec. 2006) is used to fit a least squares model and predict the next time step (January 2007). Then for the next prediction, the training period shifts by one month (Feb. 2004 - Jan. 2007), and the model is retrained. This process is continued for each prediction from 2007-2011. Implement this approach and store the predictions in a vector.
- (b) Plot the number of cases predicted by this method and compare with the predictions from the static model.
- (c) Compute the mean square error (MSE) of both methods compared with the ground truth. Which method is better?
- (d) *Adding covariates.* Instead of using just “dengue” as the predictor variable, add the remaining Google search terms as predictors in the dynamic model. How does the MSE change?
- (e) *Adding autoregression.* You can make use of the fact that the model changes over time to improve predictions by adding past observations of the ground truth as additional features. For example, for the Jan. 2007 prediction, you can add the observed dengue case counts from Nov. 2006 and Dec. 2006 as features (this is a Google+AR2 model). Implement a dynamic model with autoregression and compare the MSE of this model with the others.

Hint: Shift the case count column and stack it to the original data matrix. Your sliding training window for the first two predictions can have fewer than 36 observations.