Spring 2015 Statistics 153 (Time Series) : Lecture Nineteen

Aditya Guntuboyina

07 April 2015

1 Time Series Cross Validation

Read the two articles on Rob Hyndman's blog: `http://robjhyndman.com/hyndsight/crossvalidation/` for a simple introduction to cross validation in general and `http://robjhyndman.com/hyndsight/tscvexample/` for cross validation specific to time series.

There are many ways to do cross validation for time series. Suppose we have monthly data for m years $x_1, \dots, x_n$ where $n = 12m$ and the objective is to predict the data for the next year (This is similar to the midterm problem which has weekly data instead of monthly). Suppose we have ℓ competing models $M_1, \dots, M_\ell$ for the dataset. We can use cross-validation in order to pick one of these models in the following way:

1. Fix a model M_i. Fix $k < m$.
2. Fit the model M_i to the data from the first k years.
3. Using the fitted model, predict the data for the $(k+1)$st year.
4. Calculate the sum of squares of errors of prediction for the $(k+1)$st year.
5. Repeat these steps for $k = k_0, \dots, m-1$ where k_0 is an arbitrary value of your choice.
6. Average the sum of squares of errors of prediction over $k = k_0, \dots, m-1$. Denote this value by CV_i and call it the Cross Validation score of model M_i.
7. Calculate CV_i for each $i = 1, \dots, \ell$ and choose the model with the smallest Cross-Validation score.

2 Overfitting as a Diagnostic Tool

After fitting an adequate model to the data, fit a slightly more general model. For example, if an AR(2) model seems appropriate, overfit with an AR(3) model. The original AR(2) model can be confirmed if while fitting the AR(3) model:

1. The estimate of the additional ϕ_3 parameter is not significantly different from zero.
2. The estimates of the common parameters, ϕ_1 and ϕ_2, do not change significantly from their original estimates.

How does one choose this general model to overfit? While fitting a more general model, one should not increase the order of both the AR and MA models. Because it leads to lack of identifiability issues. For example: consider the MA(1) model: $X_t = (1+\theta B)Z_t$. Then by multiplying by the polynomial $1-\phi z$ on both sides: we see that X_t also satisfies the ARMA(1, 2) model: $X_t - \phi X_{t-1} = Z_t + (\theta - \phi)Z_{t-1} + \phi\theta Z_{t-2}$. But note that the parameter ϕ is not unique and thus if we fit an ARMA(1, 2) model to a dataset that is from MA(1), we might just get an arbitrary estimate for ϕ.

In general, it is a good idea to find the general overfitting model based on the analysis of the residuals. For example, if after fitting an MA(1) model, a not too small correlation remains at lag 2 in the residuals, then overfit with an MA(2) and not ARMA(1, 1) model.