Day 2: Self-guided practical - Fitting ARIMA models with `forecast`

Andrew Parnell

Introduction

In this practical you will:

- Fit a basic ARIMA model to the `airquality` data
- See if you can beat it with some of the other methods available through `forecast`
- Quantify the accuracy of each method

As yesterday, you should work your way through the questions and put your hand up if you get stuck. The answer script is also available in the practicals folder.

Tasks

1. Create a new data frame called `airquality2` from the `airquality` data set already loaded into R. Use the `as.Date` function to convert Month and Day into a proper time variable called `Date`.
2. Create some time series plots of Date vs Ozone and interpret them. Do you think it looks stationary?
3. Yesterday we found that `Ozone` was a bit better behaved when we log-transformed it. Forecast has a smarter method for doing transformations using the Box-Cox transformation. Use the function `BoxCox.lambda` to estimate the lambda transformation parameter for the Ozone variable. Re-plot the data with the Box Cox transformed series and see if it looks more stationary.
4. Hopefully you noticed that there are lots of missing values. If you try an ACF plot here using the standard `acf` function it will fail, but if you use the `forecast` functions `Acf` and `Pacf` it will work. Create and interpret the ACF and PACF plots. (You could also try running them on the transformed data using the function `BoxCox`)
5. It looks like some kind of AR model might work for these data. Use `Arima` to fit an AR(1) model and interpret the output. Don't forget to include the lambda argument
6. Try an `auto.arima` model and interpret your output
7. Use the `forecast` function to plot 10 steps into the future using the model you just created
8. Check the residuals of your `auto.arima` model using `hist` and QQ-plots (hint: see answers from yesterday for a reminder)
9. Unfortunately with missing values, many of the other time series methods won't work. However, we can impute (i.e. replace) the missing values using the `na.interp` function. Create a new variable `Ozone2` in your `airquality2` data frame which has no missing values. Plot this new series
10. Let's now use this complete data set to try others types of model. Run the `ets`, `nnetar`, and `tslm` functions used in the earlier lectures and tutorials today to create some different models. See if you can find ones that beat the ARIMA versions. Try and interpret the model output

Extra questions

If you finished all the above try picking another data set from the pile (not a seasonal one yet as we don't cover it till tomorrow!) and see if you can follow our standard steps:

1. Plot the data and the ACF/PACF
2. Decide if the data look stationary or not. If not, perform a suitable transformation and return to 1. If the data has a strong trend or there is a high degree of autocorrelation try 1 or 2 differences
3. Guess at a suitable p and q for an ARMA(p, q) model
4. Fit the model
5. Try a few models around it by increasing/decreasing p and q and checking the AIC (or others)
6. Check the residuals
7. Forecast into the future

In step 3 you can also try some of the other standard models in `forecast`, such as `naive`, `ets`, `nnetar`, and `tslm`.

If you're feeling really brave see if you can get your own data set into a format suitable for ARIMA modelling. Even if it involves throwing away or approximating some data it will be a useful start