Lecture 09

Case Studies: Time Series Forecasting

COMP3162 – Data Science Principles

Dr. Ricardo Anderson

Time Series Data

Components of Time Series Data

- Trend: increase or decrease in series over a period
  - The value of the dollar is trending up
- Seasonality: short term variation that results from seasonal factors (fluctuation)
  - eg. Sweater sales increase during winter months
- Cyclicity: variations caused by circumstances that repeat at IRREGULAR intervals
  - Interval is not fixed (not yearly)
- Irregularity
  - Variations due to unpredictable factors (does not repeat in any pattern)

Time Series Forecasting

Predictive analytics technique

Predict future values over time based on historical data

Popular Applications of Time Series Forecasting

- Predicting sales of products
- Predicting the weather
- Predicting population growth
- Predicting stock prices

Time Series Data Patterns

Trend - Gradual changes in the data, usually long-term growth or decline.

Level - Baseline values for the series data if it were a straight line.

Seasonality - Short-term patterns that occur within a single unit of time and repeats indefinitely.

Noise

- Random/irregular variations in the data.
- Usually not predictable or unexplained

Forecasting Methods/Models

Autoregressive Integrated Moving Average (ARIMA)

- Uses Autoregressive (AR) and Moving Average (MR) model.
  - AR model - Linear combination of past variable values
  - MR model - Linear combination of past forecast errors

The Vector Autoregression (VAR)

- Models the next step in each time series using an AR model
- Useful for predicting multiple time series variables using a single model

Seasonal Autoregressive Integrated Moving Average (SARIMA)

- Uses ARIMA along with past seasonal forecast errors

The Long Short Term Memory network (LSTM)

- Uses neural network to map long term dependencies in dataset by learning dependencies and ordering sequences

Relative vs. Linear Scale of Changes

Log-scale informs on relative changes (multiplicative), while linear-scale informs on absolute changes (additive).

When do you use each?

- When you care about relative changes, use the log-scale;
- When you care about absolute changes, use linear-scale.

More info:

- *https://stats.stackexchange.com/questions/18844/when-and-why-should-you-take-the-log-of-a-distribution-of-numbers*

Case Study: Forecasting Stock Prices

Date codes

Symbol	Meaning	Example
%d	day as a number (0-31)	01-31
%a %A	abbreviated weekday unabbreviated weekday	Mon Monday
%m	month (00-12)	00-12
%b %B	abbreviated month unabbreviated month	Jan January
%y %Y	2-digit year 4-digit year	07 2007

CODE

```
rm(list=ls())

install.packages('forecast')
install.packages('tseries')
install.packages('tidyquant')

## load packages
library(forecast)
library(tseries)
library(tidyquant)

##using CSCo Data
stock.data <- read.csv(file.choose())

## review data
str(stock.data)
summary(stock.data)

## covert thedate to a R native date
temp.date <- as.Date(stock.data$thedate,format="%m/%d/%Y")
##temp.date['2014-01-13':'2014-01-31',]

##convert to xts/zoo using tidyquant library
stock.data.xts <- xts(stock.data$open,temp.date)
colnames(stock.data.xts) <- "open"

## convert open to logarithmic values since stock prices are focused on day to day change relative values
stock.data.xts$open <- log10(stock.data.xts$open)

## check if data is stationary (constant mean) - makes the model more predicatble | gradual descent
acf(stock.data.xts$open)

##plot time series
plot.ts(stock.data.xts)

## build model
stk.model <- auto.arima(stock.data.xts)
stk.model ##view

stk.forecast <- forecast(stk.model,h=300) ## use the model to predict 300 points/date intervals in the future
plot(stk.forecast)
```

END

Case Study – Time Series Forecasting