Introduction to time series modelling

Dr Gianluca Campanella

21st June 2016

Contents

Statistics for time series

Time series

Any data that change **over time**

Time series

Any data that change **over time**

Seasonality

- Cyclic pattern(s) repeated over time
- E.g. peak of sales in December

Trend

- Change in ‘baseline’ levels over time
- E.g. linear increase in sales over last 5 years

Rolling (or moving) statistics

Each observation is replaced with some statistic (e.g. mean) of k consecutive time points:

- k preceding points
- $k/2$ points prior to and following a given time point

Usage

- Reduce influence of outliers
- Smooth time series to identify patterns

Exponentially weighted averages

- Rolling statistics weigh the k time points equally
- Often, points closer in time are more important

$\rightarrow$ Weighting

Exponentially weighted averages

- Rolling statistics weigh the k time points equally
- Often, points closer in time are more important

$\rightarrow$ Weighting

Exponential weighting

- $\text{EWMA}_1 = y_1$
- $\text{EWMA}_t = \alpha\, y_t + (1 - \alpha)\, \text{EWMA}_{t-1},\ t > 1$

$\rightarrow$ α controls the **decay**

Expanding statistics

Each observation is replaced with some statistic (e.g. sum) of all points prior to the given time point

Usage

- Visualise cumulative distribution over time
- Identify trends

Autocorrelation

Correlation of the time series with itself at different lags:

- At lag 1, dependency on ‘yesterday’
- At lag 7, dependency on ‘last week’
- At lag 30, dependency on ‘last month’...

Usage

- Identify trends
- Identify period of seasonal cycles

Forecasting

Forecasting

Prediction

- Value of y given values for the predictors X
- Does not depend on time
  (or temporal effect is negligible)

Forecasting

Prediction

- Value of y given values for the predictors X
- Does not depend on time
  (or temporal effect is negligible)

Forecasting

- Value of y given **previous values** of y
- Captures autocorrelation to ‘project forward’
- (Some models can also incorporate predictors)

Model evaluation

- Same metrics used for regression (e.g. MSE)
- Standard cross-validation doesn’t apply

Model evaluation

- Same metrics used for regression (e.g. MSE)
- Standard cross-validation doesn't apply

Solution

- We can only forecast **based on past values**

$\rightarrow$ Split into training/test sets
(e.g. before/after some point in time)

Stationarity

Many models require time series to be **stationary**:

- Mean and variance constant over time

$\rightarrow$ Seasonality and trend must be removed

Stationarity

Many models require time series to be **stationary**:

- Mean and variance constant over time

$\rightarrow$ Seasonality and trend must be removed

Solutions

- Detrending (estimate and subtract ‘baseline’)
- Differencing (predict change or ‘change in changes’)

Time series models

ARMA models

Auto**R**egressive

- y_t depends on y_{t-1}, ...
- Regression on past values
- Captures (slow) changes in trend

Moving **A**verage

- y_t depends on ε_{t-1}, ...
- Smoothing of past errors
- Captures sudden changes (e.g. spikes)

$$y_t = \alpha + \beta_1 y_{t-1} + \ldots + \beta_p y_{t-p} + \gamma_1 \varepsilon_{t-1} + \ldots + \gamma_q \varepsilon_{t-q}$$

ARIMA models

- ARMA models on differentiated time series
- Typically $d = 1$ or 2

ARIMA models

- ARMA models on differentiated time series
- Typically $d = 1$ or 2

First-order differences $y_t - y_{t-1}$

- Predict change
- Corresponds to velocity in physics

Second-order differences $(y_t - y_{t-1}) - (y_{t-1} - y_{t-2})$

- Predict ‘change in changes’
- Corresponds to acceleration in physics