Lec 12

Thursday, 21 November 2024 19:03

1. Time series analysis
2. Time series is a sequence of some value through time
3. Must be equally spaced. Data points are took in equal periods of time.
4. We can't really gather complex data for this model. The only data we know is historical data.
5. Time series examples: stocks prices, temperature, humidity through time.
6. The main goal is to crack the hidden law of changing value.
7. The only data we have are measurements in the past
8. Accuracy of such models is a little cooler than throwing a coin, but they can be used in more complicated predictive models - ensemble models
9. Main goal to predict random part of the value
10. The further in future the prediction, the less accurate it is.
11. Also does not take in account some global events like covid
12. Box-Jenkins method:
    - Arima or arma or box-jernkind is working when the random component of sequence is stationary. Otherwise we can' be sure in predictions
    - Historical value consists of trending term, random term and seasonal term
    - De-trending - approximating with linear(but not always, sometimes trending component can be quadratic or exponential function) regression on trending component and substracting from core value
    - If we plot box-plot via year or week or month we can explore if sequence have seasonality
    - Than we substract this seasonal term from sequence
    - What have remained is random term
    - ARMA(p, q) - autoregressive moving average
    - $$Y_t = \boxed{\delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p}} + \boxed{\epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q}}$$
    - Linear function from previous moments in time with addition of white dampened noise part or our model error in previous time moments predictions
    - ARIMA(p, d, q) - autoregressive integrated moving average
    - Here we are using differences and not values themselves. It helps us to get rid of trending term wether its linear or quadratic or power of -> **d**
    - There are no strict rules to choose p,d,q params. Only some heurestic methods
    - ACF and PACF - correlation functions that can help us to choose p, q and d for ARMA and ARIMA. ACF helps us to choose q and PACF helps with choice of p

ACF & PACF

- Auto Correlation Function (ACF)
  - Correlation of the values of the time series with itself
  - Autocorrelation "carries over"
  - Helps to determine the order, q, of a MA model
    - Where does ACF go to zero?
- Partial Auto Correlation Function (PACF)
  - An autocorrelation calculated after removing the linear dependence of the previous terms
  - Helps to determine the order, *p*, of an AR model
    - Where does PACF go to zero?

- We'll see some pattern in these functions and can determine values p and q
- Advantages: no serious data retrievement required(but usually we don't use them straightforward without some other regression model), also accounts trends and accountability
- Cautions: no explanatory value, only prediction, hard to choose params
- Useful funcs for R:

> - The function "*ts*" is used to create time series objects
>   - **mydata<- ts(mydata,start=c(1999,1),frequency=12)**
> - Visualize data
>   - **plot(mydata)**
> - De-trend using differencing
>   - **diff(mydata)**
> - Examine ACF and PACF
>   - **acf(mydata)**: It computes and plots estimates of the autocorrelations
>   - **pacf(mydata)**: It computes and plots estimates of the partial autocorrelations

- **ar()**: Fit an autoregressive time series model to the data
- **arima()**: Fit an ARIMA model
- **predict()**: Makes predictions
  - *"predict"* is a generic function for predictions from the results of various model fitting functions. The function invokes particular methods which depend on the *class* of the first argument
- **arima.sim()**: Simulate a time series from an ARIMA model
- **decompose()**: Decompose a time series into seasonal, trend and irregular components using moving averages
  - Deals with additive or multiplicative seasonal component
- **stl()**: Decompose a time series into seasonal, trend and irregular components using loess