TECHNISCHE UNIVERSITÄT DRESDEN

Dresden Database Systems Group

Time Series Forecasting

Programming for Data Science

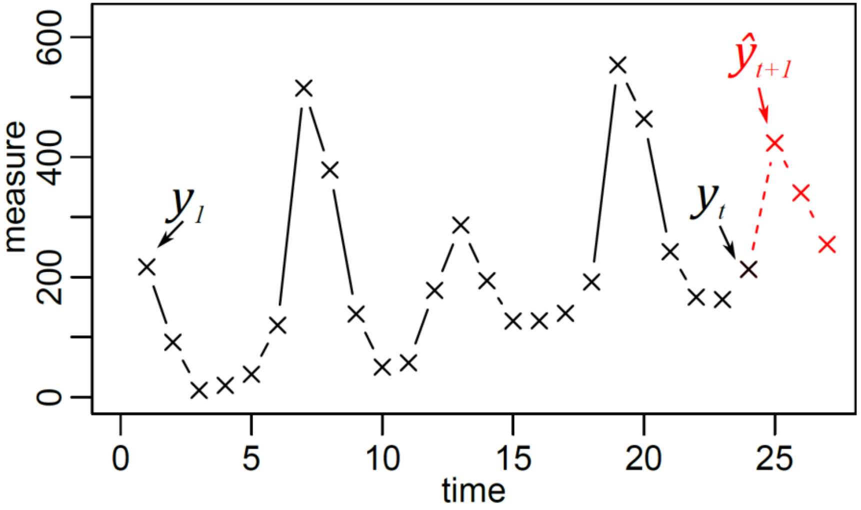
Classical Time Series Forecasting

Time Series

- Sequence of measure values
- Equidistant
- Complete

Forecast Models

- One model represents one time series
  - Trend (long-term changes)
  - Season (regular reoccurring changes)
- Accuracy is the only requirement



ARMA Models

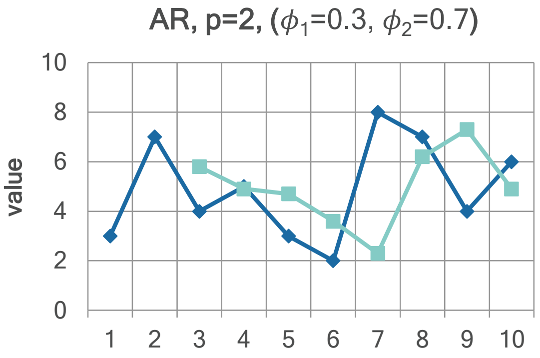
Dresden Database Systems Group

Box and Jenkins Methodology

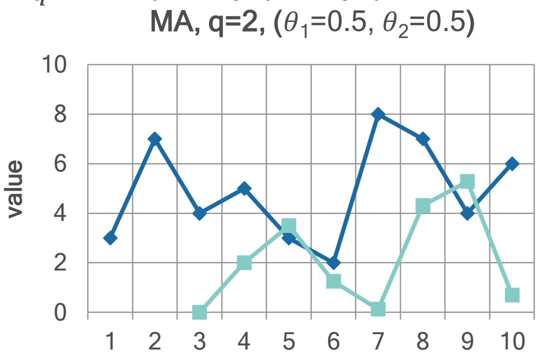
- Modeling time series with AutoRegression

Classification and Model Hierarchy

- AR(p): $\hat{x}_t = c + \sum_{i=1}^{p} \phi_i \cdot x_{t-i}$
  AutoRegression



- MA(q): $\hat{x}_t = \sum_{i=1}^{q} \theta_i \cdot \varepsilon_{t-i}$
  Moving Average
- $\varepsilon_0, \cdots, \varepsilon_q = 0, \varepsilon_i = x_{t-i} - \hat{x}_{t-i}$



- Combine AR and MA by addition: $\hat{x}_t = \sum_{i=1}^{p} \phi_i \cdot x_{t-i} + \sum_{i=1}^{q} \theta_i \cdot \varepsilon_{t-i} + c$

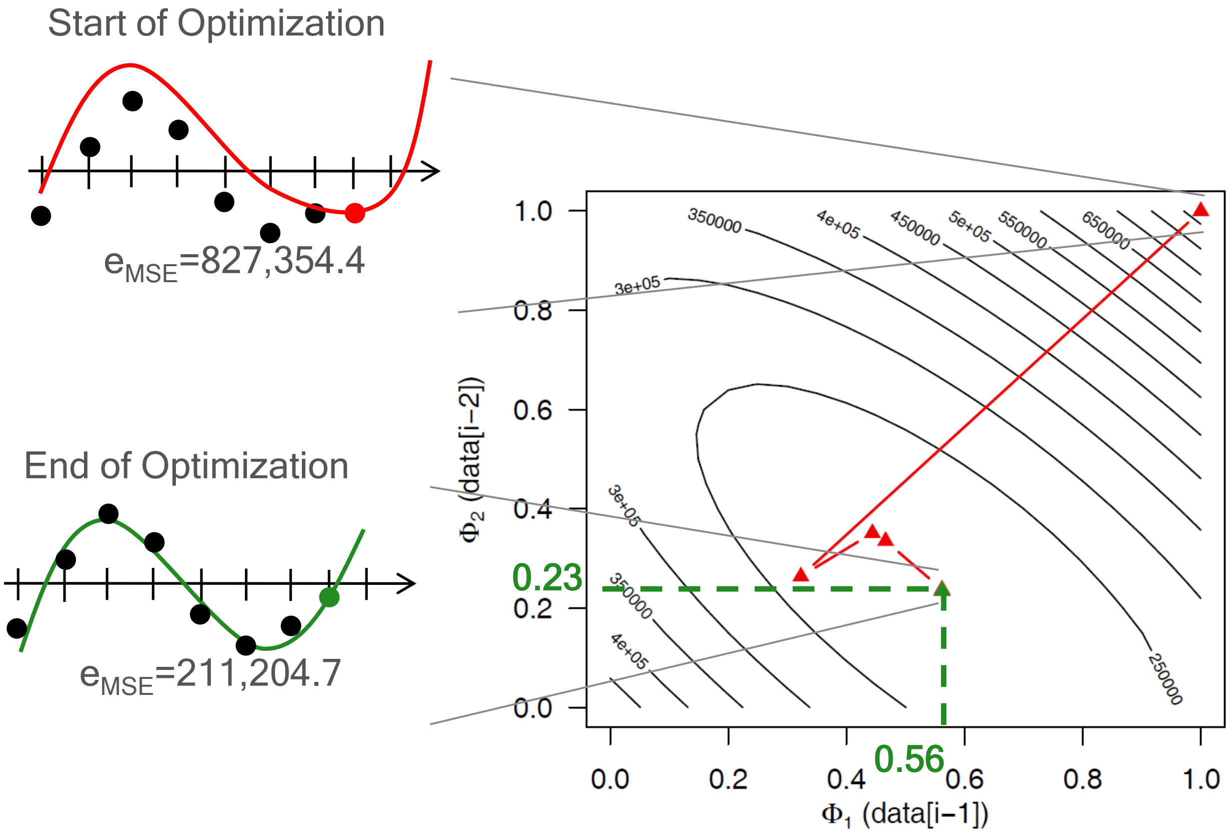
Model Estimation (Math Part)

Problem

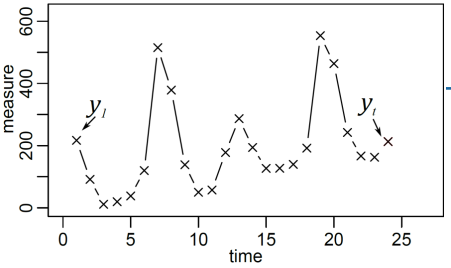
- Find the model parameters that minimize the error on the training data

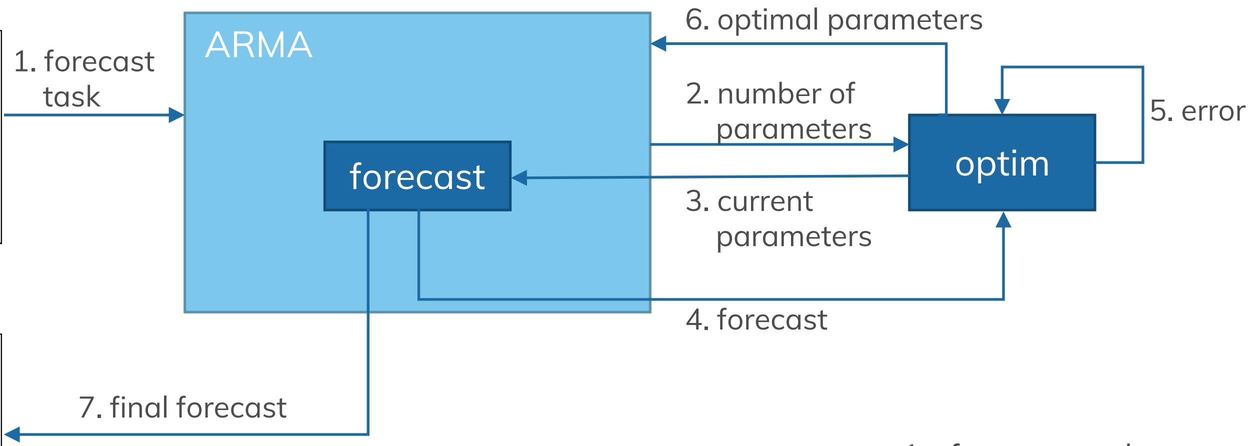
Example

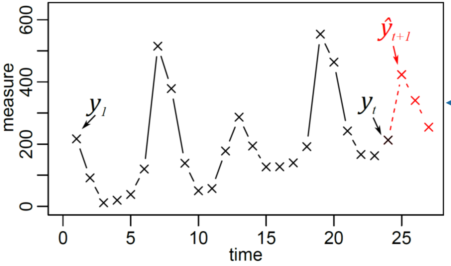
- Forecast Model Type AR(2):
  - $\hat{x}_t = \phi_1 \cdot x_{t-1} + \phi_2 \cdot x_{t-2}$
- Error Metric: MSE
  - $\frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2$



Model Estimation (Technical Part)





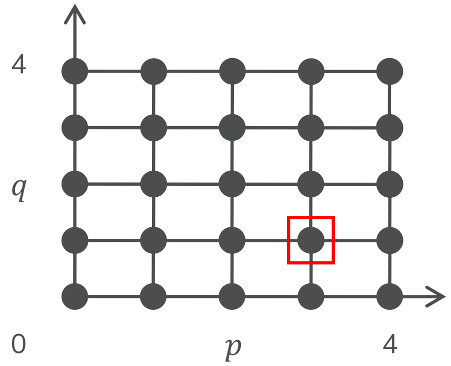


1. forecast task
2. number of parameters
3. current parameters
4. forecast
5. error
6. optimal parameters
7. final forecast

Global Optimization Methods

Grid Search (Derivation Free)

- Simple but robust method: no local suboptimal
- First define a step width to sample the parameter space (Grid)
- Exponential number of combinations X^D (granularity X and dimensionality D)
- Model evaluation for every combination



Task

Dresden Database Systems Group

Step 0

- Load the given csv file in your language/environment.
- Explore the data and its structure.

Step 1

- Implement the AR and MA parts of the ARMA model*.

Step 2

- Use a general purpose optimizer (Nelder-Mead) to fit the model parameters to a given time series.

Step 3

- Search for the optimal combination of AR (p) and MA (q) components of the model for each given time series*. Perform a grid search* for $p \in [0,3]$ and $q \in [0,3]$.
- Report the forecast error.

*use your own implementation

Package suggestions

R

- (data.table)
- stats

python3

- pandas
- numpy
- scipy.optimize
- (matplotlib)

Modeling Results

Result

- Mean Squared Error of all time series using the best combination of *p* and *q* and a constant c

Comparison

- The values for p and q can vary based on your implementation and parameter initialization
- Only compare if your MSE correlate with our solution and are in a similar range

shop_id	Mean Squared Error
145	210
260	345
315	303
375	265
548	403
560	1272
750	539
897	9458
1332	3476
1750	377

Exercise Appointment

Dresden Database Systems Group

We compare and discuss the results

- Tuesday, 17.12.2019,
- Consultation: 12.12.2019,
- Please prepare your solutions! Send us your code!

If you have questions, please mail us:

claudio.hartmann@tu-dresden.de Orga + Code
lucas.woltmann@tu-dresden.de Tasks + Python
lars.kegel@tu-dresden.de Tasks + R