Key Factors for Success in Data Science

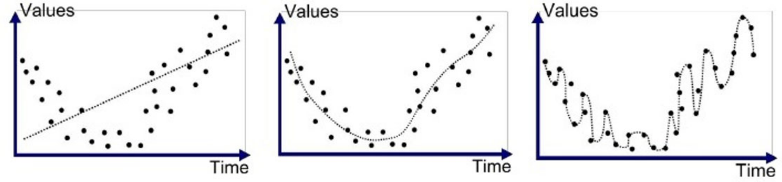
1. Unambiguous Objective
- In Forecasting: Well-selected forecasting horizon/interval, data & target feature(s) and algorithms (interpretable ML vs ANN)

2. Proper Data Reprensentation
- Choice of appropriate resolution and units/objects
  - For example is there a systematic change between weekdays (e.g. [Taylor/Letham, 2018]) or day and night (e.g. [Locarek-Junge, 2019])
  - => Disaggregate one time series into components
- Does the regional location of time series has structural differences?
  - Perform cluster analysis (e.g. [Thrun, 2019b])
- Can problems be summarized? ->Aggregate multivariate problem to univariate time series (e.g. [Thrun et al., 2019]))
- Preprocessing (e.g. detrending, standardization,…)

3. Limited Uncertainty
- Overfitting is accounted for well if time interval of test set is at least one season with a statistically large enough sample (>100 cases)

4. Choice of Similarity
- Quality measure with an appropriate bias is chosen
- Algorithm with appropriate bias is selected

Effects of Uncertainty



Underfitted | Good Fit/Robust | Overfitted

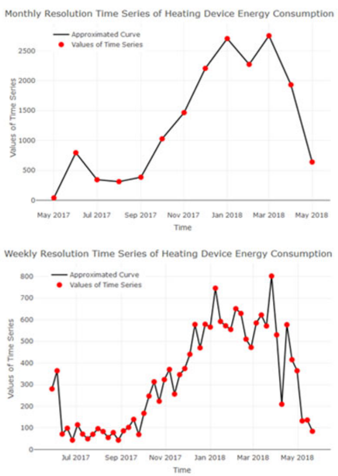
- Underfitting refers to a model that can neither model the training data nor generalize to new data
  - -> Knowledge Discovery approaches on residuals and temporal structures
- Overfitting refers to a model that learns also noise instead of only learning the signal
  - -> Should be investigated with statistical approaches using specific cross-validation prozedure
  - -> More likely with nonlinear models like neural networks

Recapitulation: Forecasting

Y_t is equidistant and a series of measurements visualized as curve with a specific resolution in time

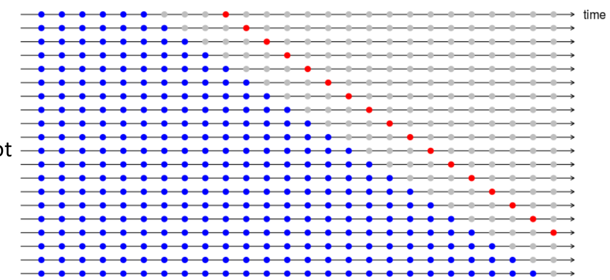
Given Y_f, an h step forecasts is a prediction of the values of $Y_{f+1}, \dots, Y_{f+h}$, if it is based only on the information available at time t=f

- Forecast origin f
- Forecast horizon h
- Forecasted time series $\widehat{Y_i}$ in $f+1, \dots, f+h$



Best Practice: Cross-Validation

- Cross-validation
  - Use more than one round of out-of-sample forecasting
  - Account for temporal structures by rolling forecast
  - Multi-step forecasting horizon with unit point h in red
  - See coding in Rmarkdown for details



- Blue: training set
- Grey: test set (sample not used in model building)
- Red: forecast horizon

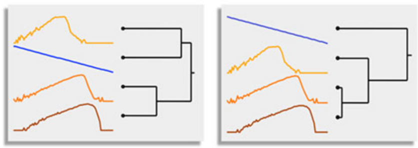
The Ugly Duckling Theorem (UDT) [Watanabe, 1969]

- One of the key question in Data Science is similarity
  - Relevant for every pattern recognition algorithm
  - (Dis-)similarity is most often Euclidean distance, but thats most often incorrect
- UDT states: **classification is impossible without some sort of bias**
  - Depends on the features chosen

⇒Similarity depends on the representation of data

=> Quality Measurement can be regarded as a similarity measure [Thrun et al., 2019]

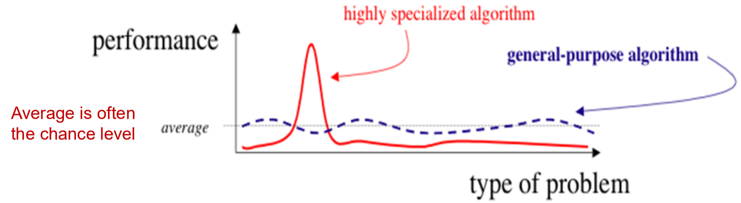
Are this two pictures is similar to each other?



The definition of similarity depends on the user, the domain and the task at hand. We need to be able to handle this subjectivity.

No Free Lunch Theorem (NFL) - [Wolpert, 1996]

- Let X Time series of electricity prices should be forecasted, one per region, then
- **No solution given by an algorithm can be better than any other if the number of problems is high enough**



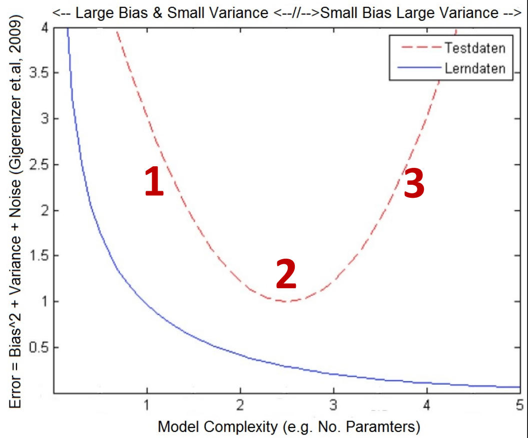
⇒Choose the "right" representation of data meaning that you answer the question: „What are my objects/smallest units in which I operate"?

⇒Select algorithm(s) problem-specific and adapted to your data representation(e.g. [Thrun et al., 2019])

⇒Requires discussion with domain expert

Pitfall of the Learning Behavior of ML Models

1. Bias of algorithm is reduced and variance is increased in relation to model complexity
2. As more and more parameters are added to a model, the complexity of the model rises and variance becomes our primary concern while bias steadily falls
3. However, above a certain generally unknown limit the underlying principles are not well mapped, i.e. the test data set is forecasted with large errors



Sources: [Ultsch Lectures; Geman et al., 1992; Gigerenzer/Brighton, 2009; Ben-David et al., 2018]

Implicit Assumptions for Working with ANNs

I. Less complex but „understandable" ML approaches approaches failed
   - E.g. Facebook's prophet, [Taylor/Letham, 2018]

II. Knowledge discovery was performed extensively
   1. Structure and quality of the data were statistically clarified PRIOR to coding
   2. Problem is multivariate
      - Relevant information is contained in more than one time series
      - Information is nonlinear and complex interrelated
   3. Causality (cause -> effect)
      - No other predictors majorly influences the price market significantly (e.g. energy production, hydro reservoirs in Scandinavia, …)
      - Used predictors are available BEFORE period of forecasting

=> Then, and only then we should use artificial neural networks