



Katja Glass
Consulting



Machine Learning for Humans

Hands-on Workshop

PhUSE SDE Heidelberg 2019

Katja Glass

Eike Nicklas

Agenda



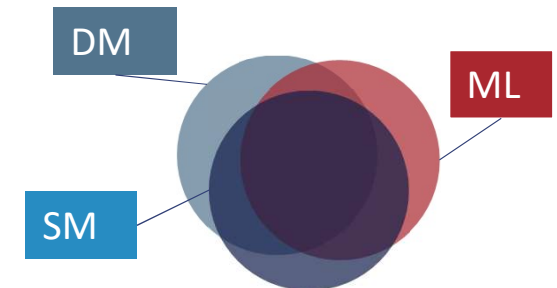
- Short Introduction
 - Hands-on: Breast Cancer
Sample Analysis
-



Statistical Modeling, Data Mining, Machine Learning



- Statistical Modeling:
 - Small data, apply stats to derive information
- Data Mining:
 - Exploratory analysis, find new information
- Machine Learning:
 - Big data, find information by “brute-force analysis”, often black-box



Machine Learning Areas



Supervised

- Learning by “labeled” data
- Approaches:
 - Classification
 - Regression
- Methods:
 - Linear Regression
 - Decision Trees
 - Neural Networks
 - ...

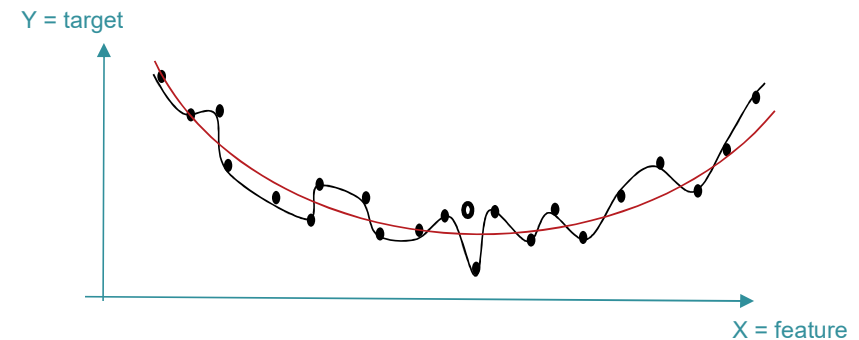
Unsupervised

- Learning by data
- Approaches:
 - Clustering
 - Association
 - Dimension Reduction

Supervised Learning - Concepts & Terms



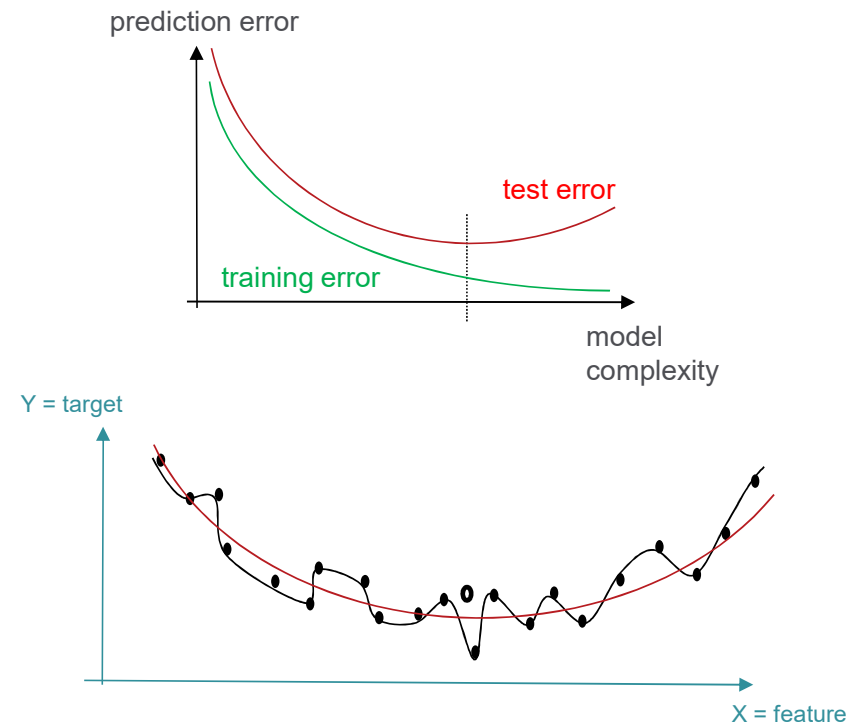
- Goal: Approximate a function
- Train model on data
- Minimize deviation from training data
→ Overfitting!
- Train/Test cycle



Supervised Learning - Concepts & Terms



- Tuning
- Feature Engineering
- Balancing / Overbalancing



Hands-on: Breast Cancer Sample Analysis



- Activate WLAN access on your laptop
 - Connect to the hands-on working environment (Chrome or Firefox)
 - <https://phusemlworkshop2019.westeurope.cloudapp.azure.com:8000>
 - Ignore the warning about a potentially unsafe connection and proceed
 - Log-in with the credentials provided (user<xy>, PhUSE2019)
 - Open the notebook ML_Hands_On_Workshop_Phuse2019.ipynb
 - Work in teams, discuss, explore
 - Raise your hand, if you run into any issues or have questions
-

Summary (1/2)

- Predicted whether tumor is benign or malignant (Supervised Learning)
- Algorithms used: Decision Tree, Random Forest, Boosted Trees, Neural Nets
 - How to choose an algorithm?
- Concepts used:
 - Training / Test Cycle
 - Cross Validation
 - Model quality evaluation – What is the best quality metric for the task?

Summary (2/2)

- Quality assurance of Machine Learning models
 - What are the consequences if a prediction is wrong?
 - How can I understand the black box? Or at least describe it?

Golden rules of Data Science (in the real world) ...

It is simple to get started, but...

- If results are too good or too bad to be true, you have made an easy fault.
- “WITT”: What is the target? I.e. the target variable as well as the overall objective of the model.
- Your data is always bad: Missings, Outliers, False content, Unknown content,)
- **KNOW YOUR DATA!**
 - -> Each variable: What, when, where, who, ..
 - Explore, explore, explore = plot, plot, plot
- Always try to evaluate the robustness/variance/variability of your results.
- **DO NOT SWITCH OFF YOUR MIND!** ... because the algorithm cannot think.

Thank you for your attention



HMS Analytical Software
GmbH
Rohrbacher Str. 26
69115 Heidelberg
www.analytical-software.de
info@analytical-software.de
+49-6221-6051-0

Katja Glass

katja.glass@glacon.eu

Eike Nicklas

eike.nicklas@analytical-software.de

