Fernfachhochschule Schweiz
Zürich | Basel | Bern | Brig
Mitglied der SUPSI
Scuola universitaria professionale
della Svizzera italiana
SUPSI

Machine Learning (ML)

Geo.BigData(Science)

Dr. Joachim Steinwendner

Laboratory for Web Science

Menschen

Physik

Geo-
Health- Informatik

Data Science

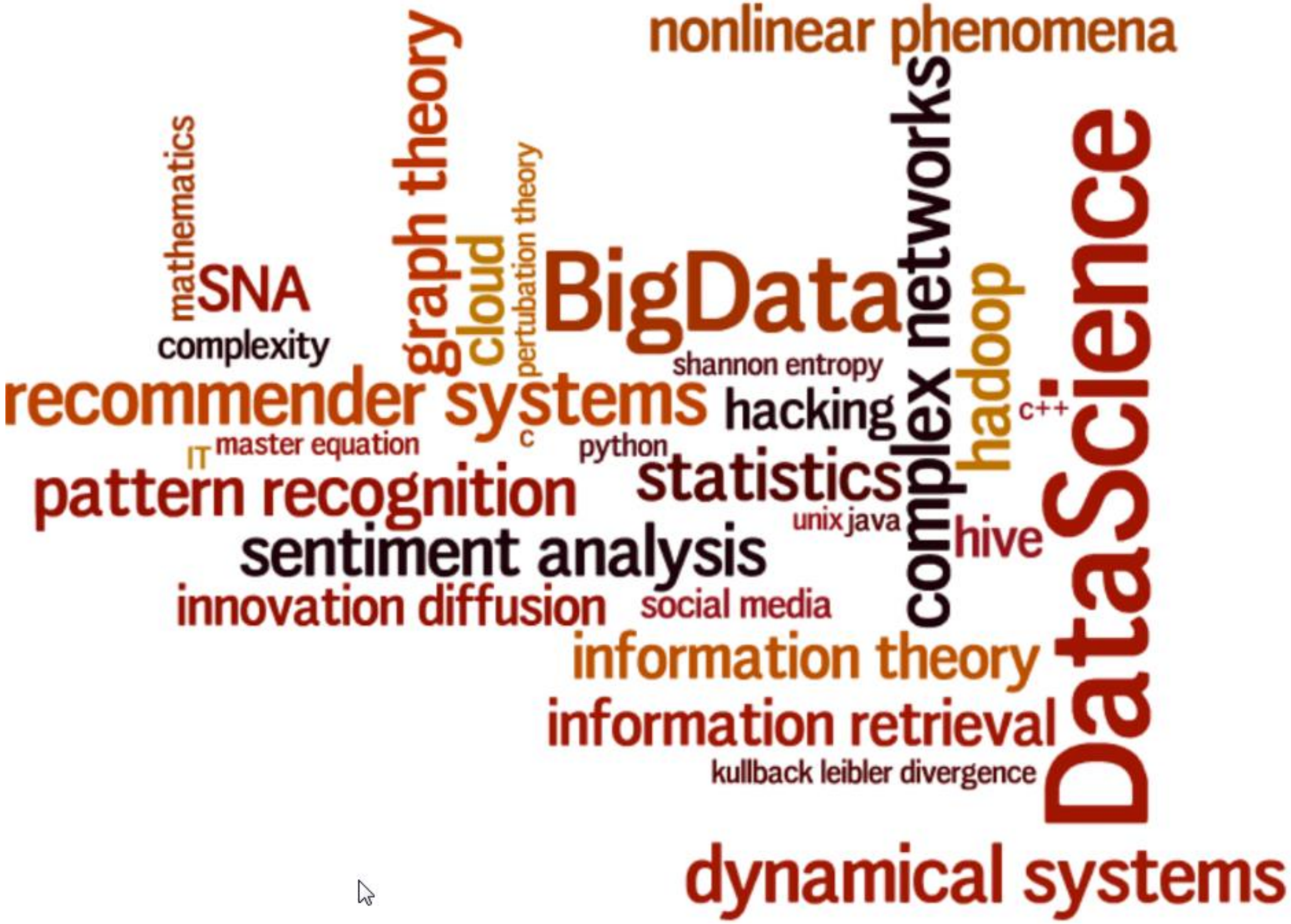
Mathematik

Laboratory for Web Science

Kompetenzen

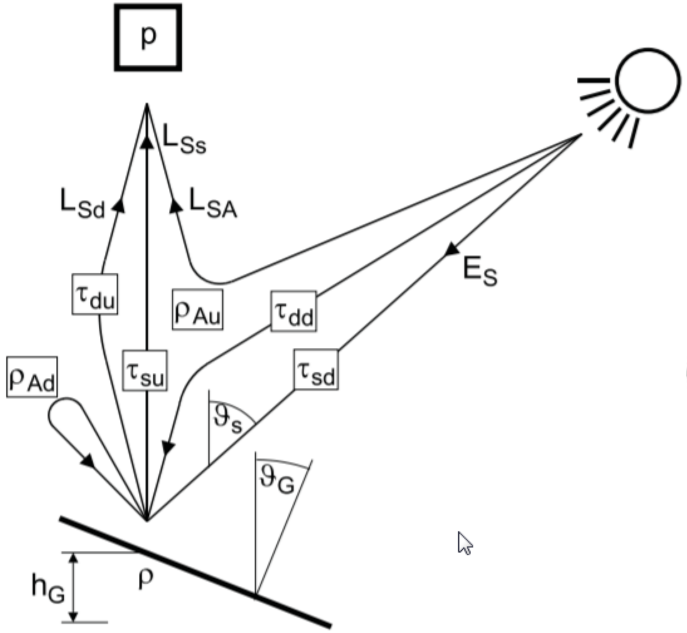
Applied Data Science:

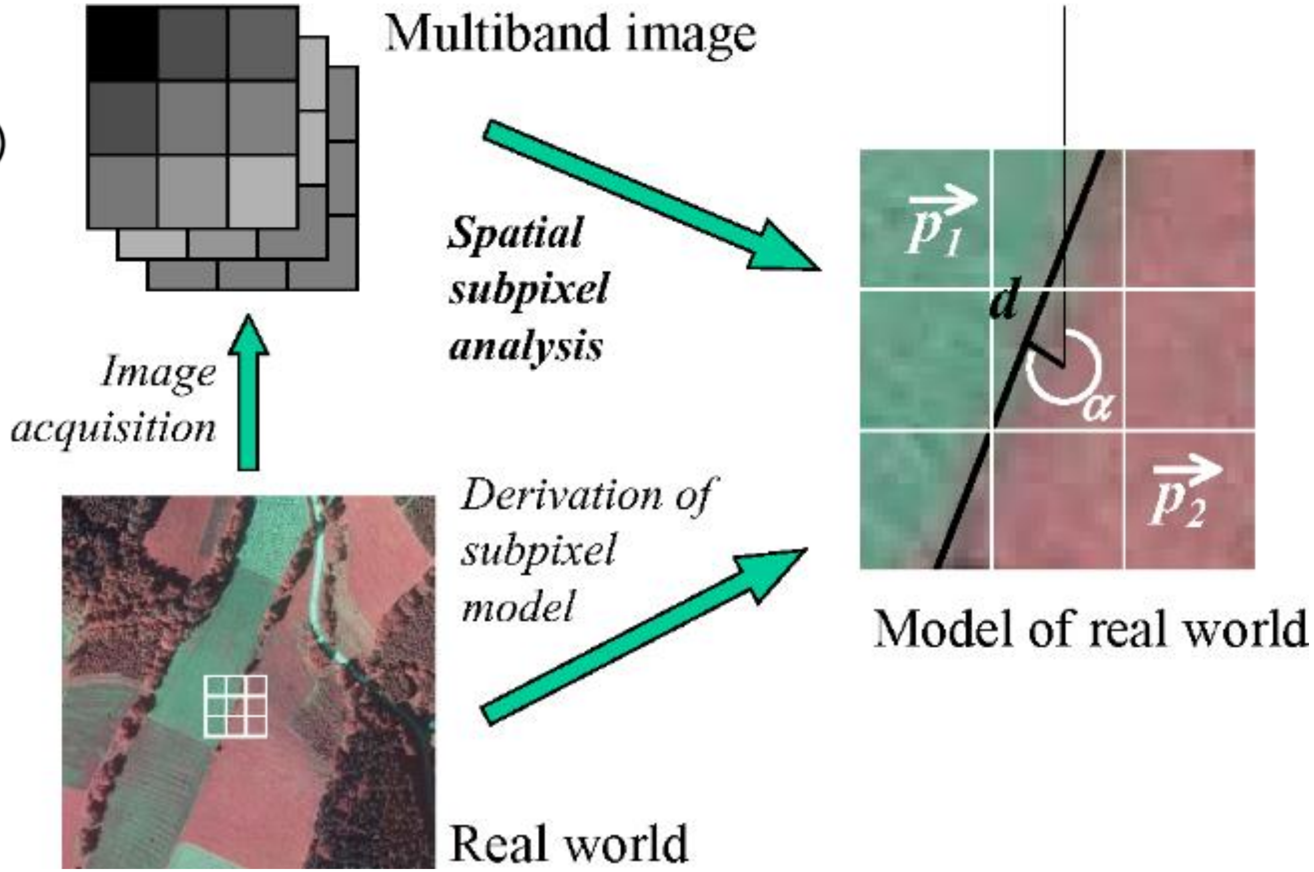
- Machine Learning
- Deep Learning
- Big Data
- Complex Networks
- Recommender Systems

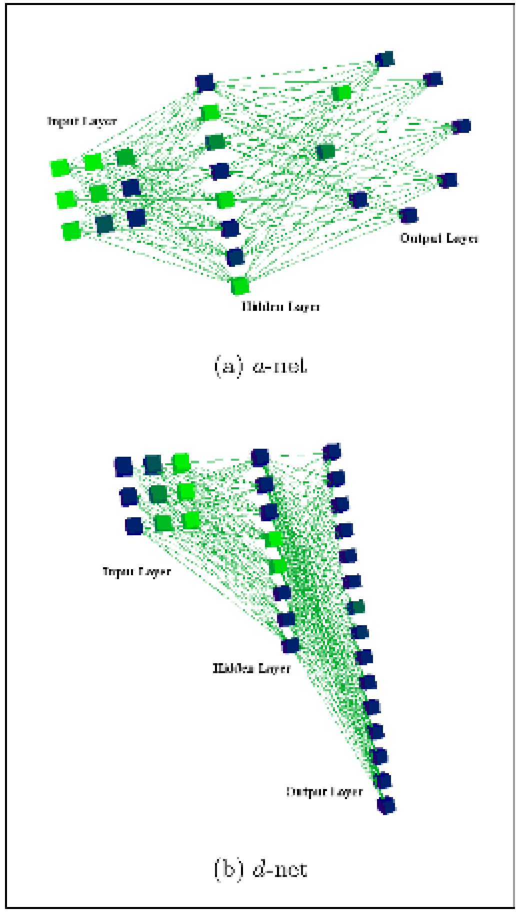


Funding: FWF – Österreichischer **F**onds zur Förderung der **W**issenschaftlichen **F**orschung

PhyRS - Physical models in remote sensing image understanding

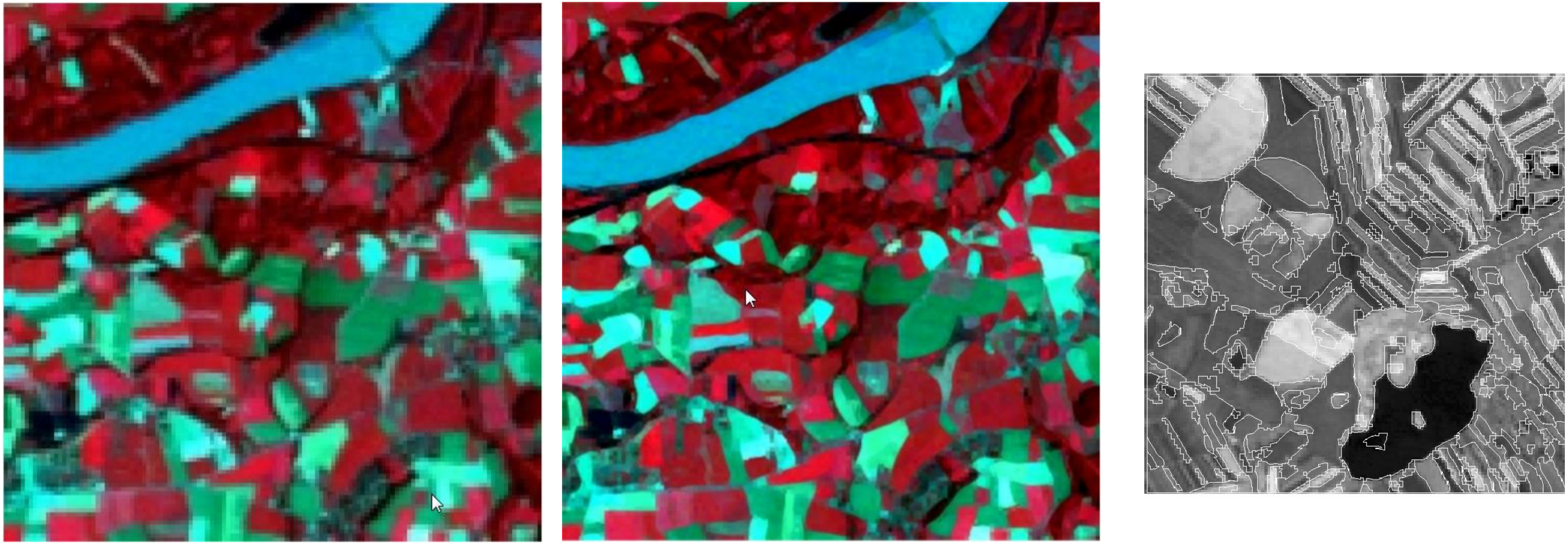






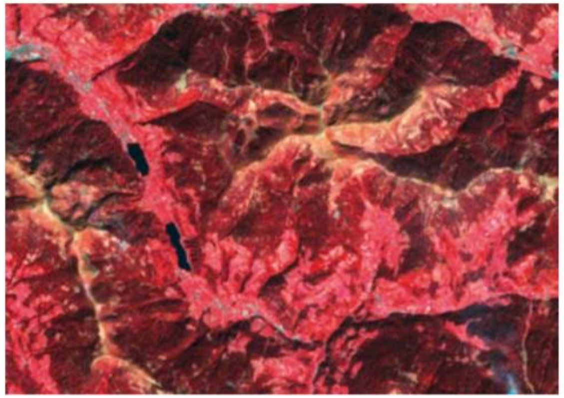
Funding: FWF – Österreichischer **F**onds zur Förderung der **W**issenschaftlichen **F**orschung

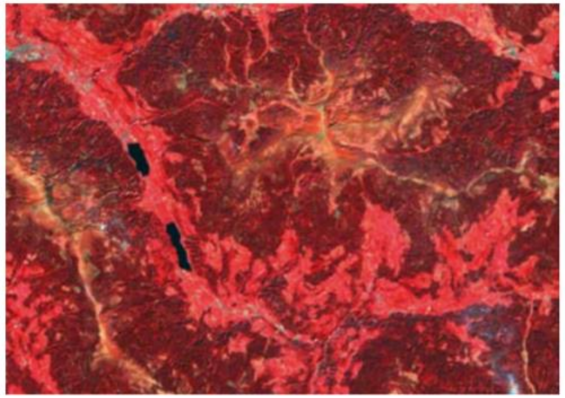
PhyRS - Physical models in remote sensing image understanding

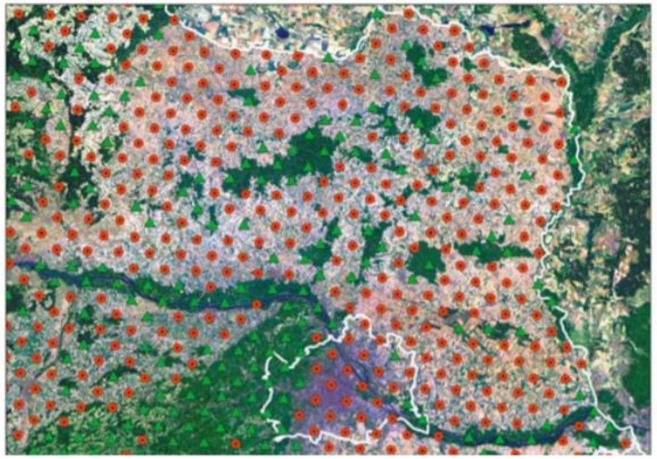


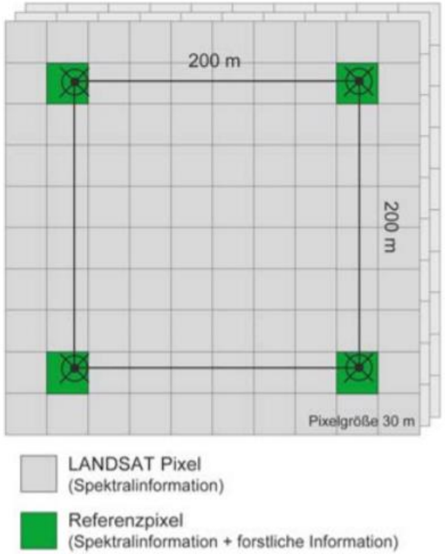
Funding: BFW - Bundesamt für Wald, Österreich

WIS - Untersuchung zur flächigen Verdichtung der Waldinformation aus der österreichischen Waldinventur mit Satellitenbilddaten









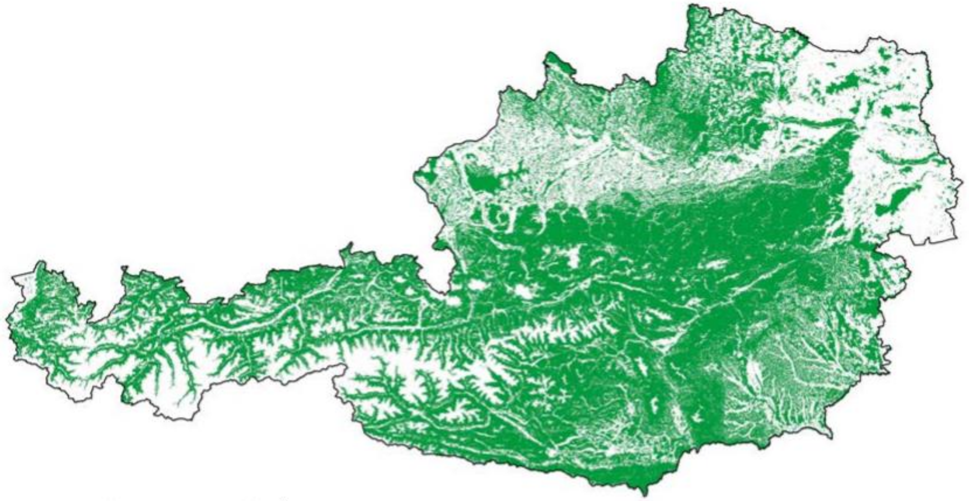


Abb. 5: Wald-/Nichtwaldkarte für Österreich

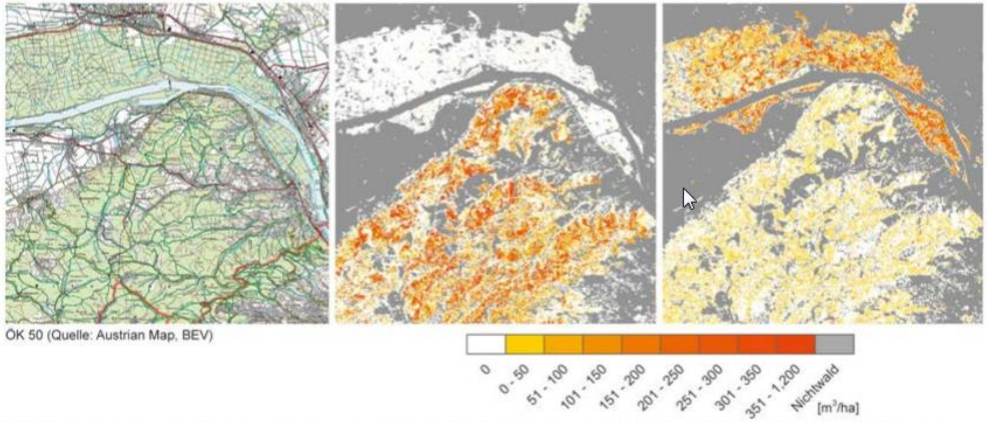
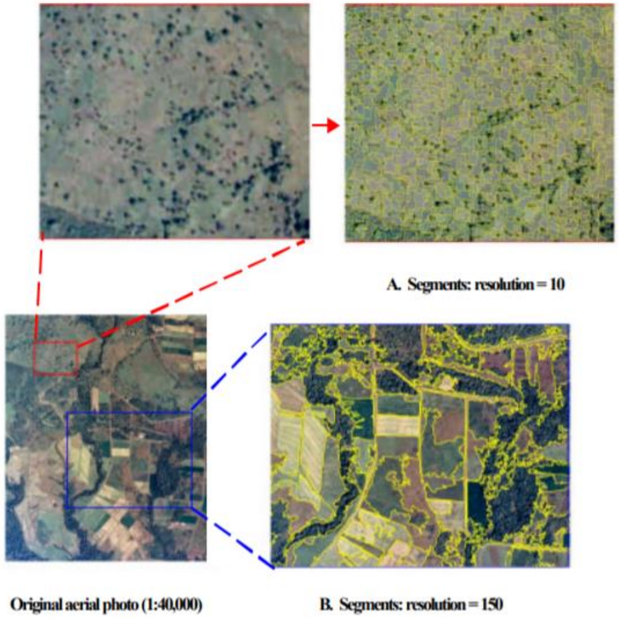
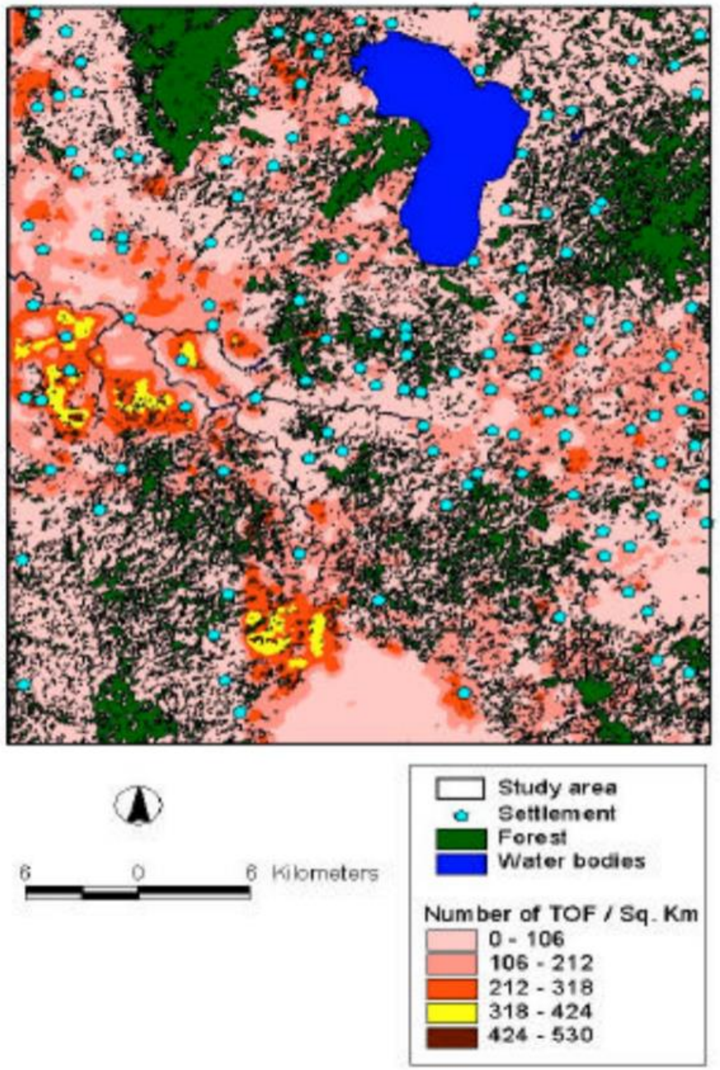


Abb. 6: Schätzung des Holzvorrats am Beispiel von Buche (Mitte) und Weichlaubhölzern (Pappel, Weide, Erle, …) im Wienerwald und in den Donauauen bei Klosterneuburg

Funding: EU INCO/DC - **IN**ternational **CO**operation with **D**eveloping **C**ountries Programme

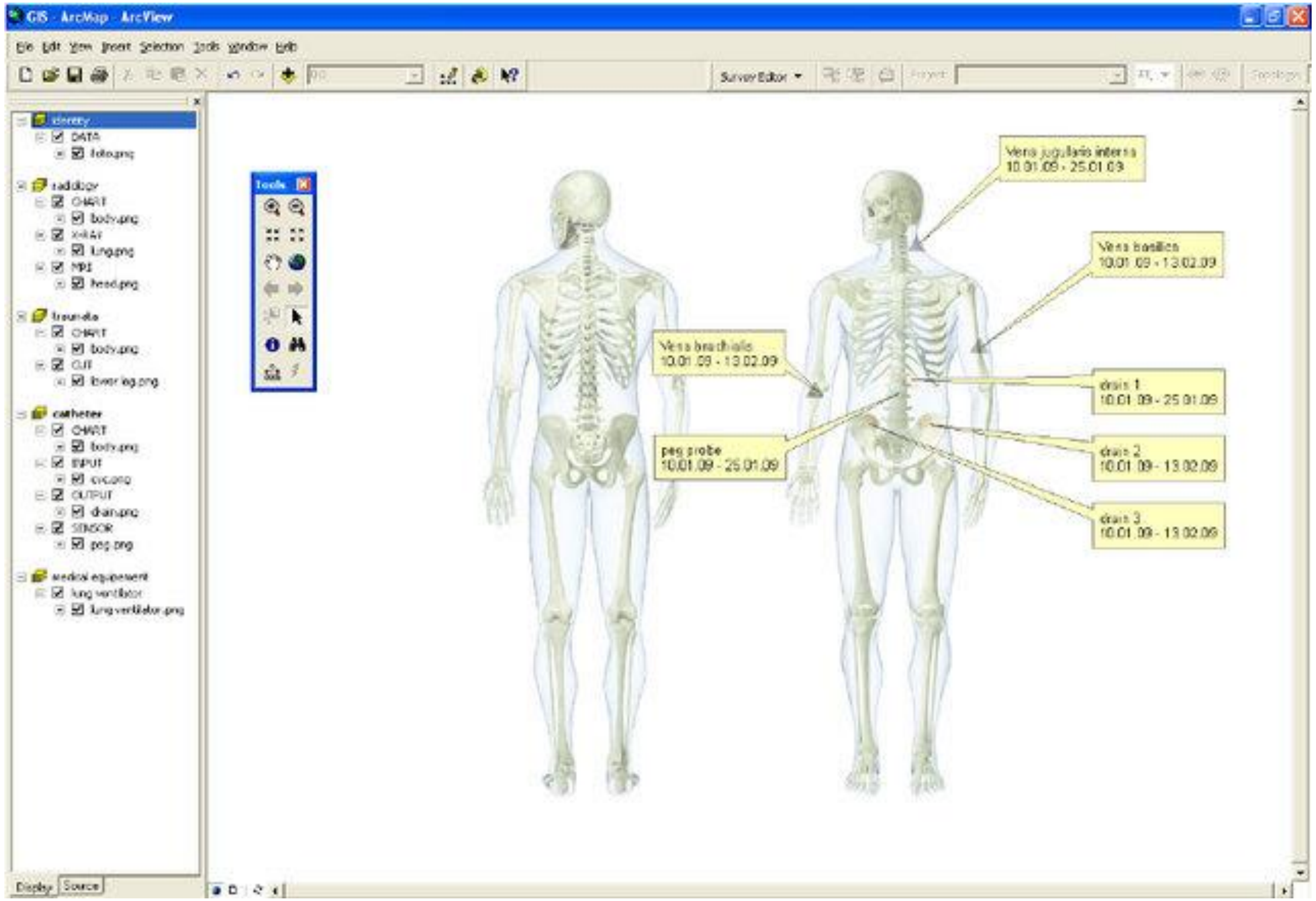
TROF – Tree resources outside the forest
assessment and monitoring of natural resources in Central America





Funding: PMU - Paracelsus Medizinische Universität Salzburg – UK Anästhesie und Intensivmedizin

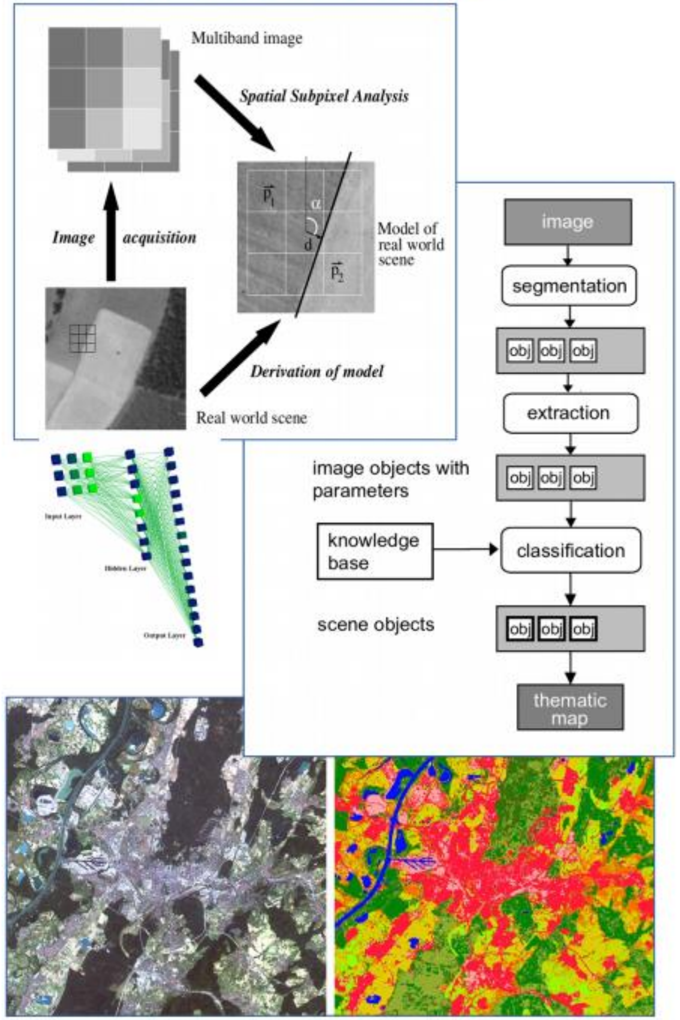
GeoHealth – GIS in der Intensivmedizin



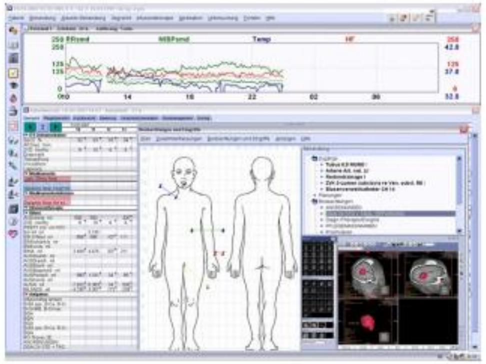
Funding: PMU und FH Salzburg - Study Project

GeoHealth – Entwicklung einer semi-automatischen Wunddokumentationsmethode

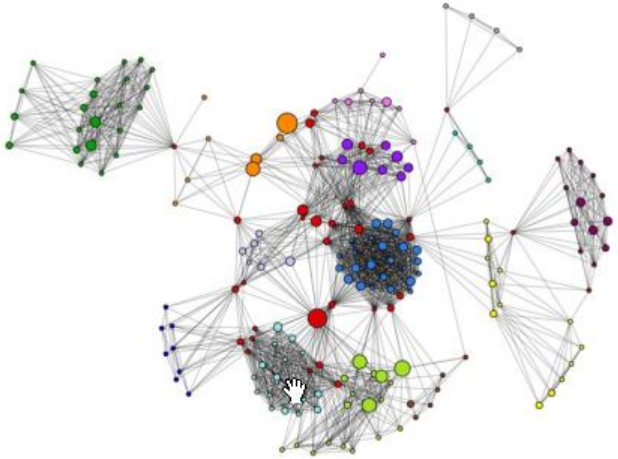
Ihr Dozent



Bildverarbeitung für erdbeobachtende Satelliten



Medizinische Informationssysteme für Anästhesie, Intensivmedizin und Neonatologie



Laboratory for WebScience
Forschungsinstitut der FFHS
Data Science,
Recommendersysteme
Machine Learning

Machine Learning
"a little history"

- A. H. (ante Hinton)
  - Statistik
  - Pattern Recognition
  - Computer Vision
  - Spracherkennung
  - ...
- 2006 Geoffrey Hinton: "Deep Learning", Erkennung handschriftlicher Ziffern mit >98% Genauigkeit
- P. H. (post Hinton)
  - "Machine Learning" Tsunami
  - Machine Learning erobert die Industrie

Machine Learning
Positionierung

- Was ist "Machine Learning"?

  „...is the field of study that gives computers the ability to learn without being explicitedly programmed.“
  (Arthur Samuel, 1959)

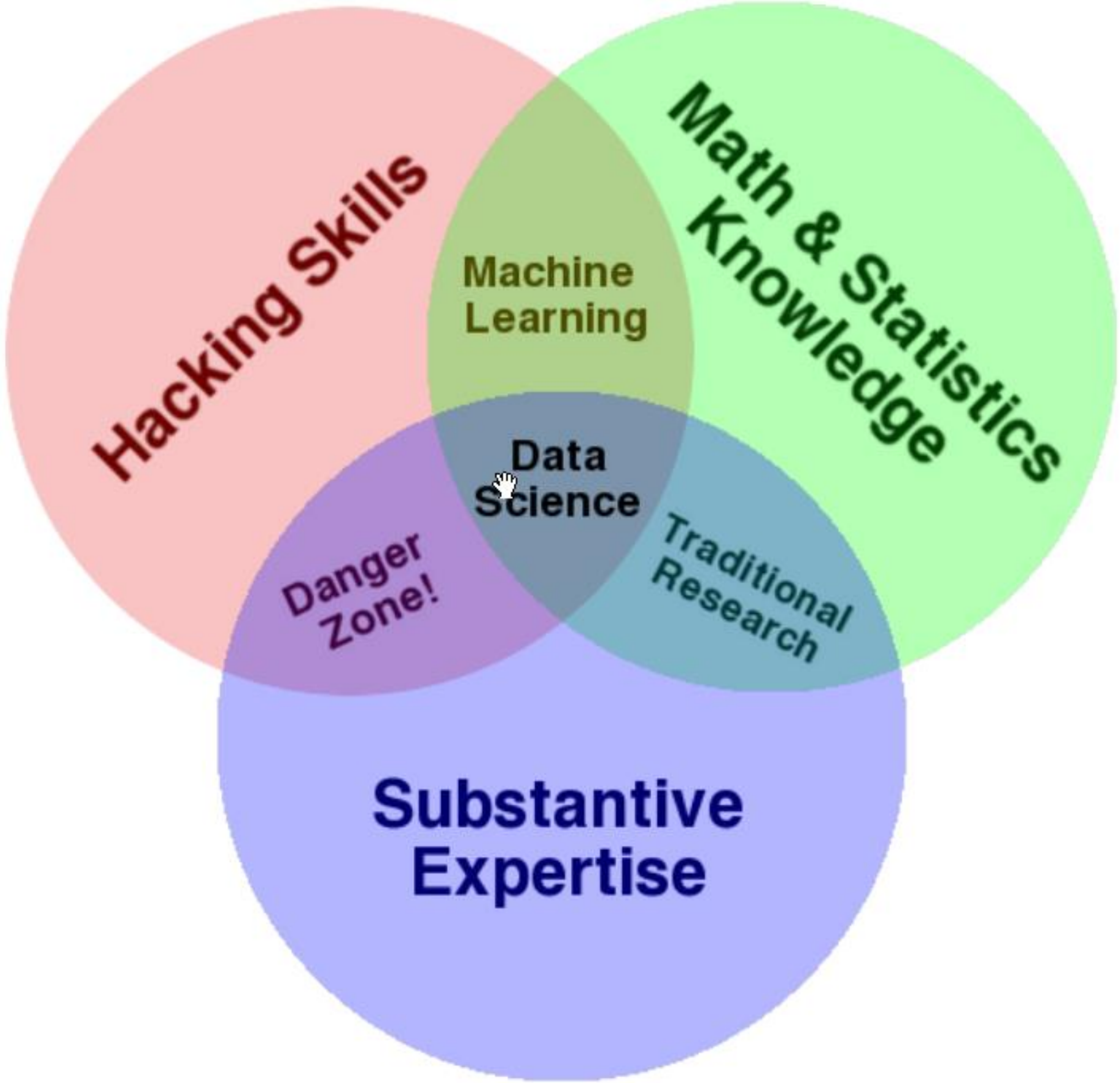
  „...is a computer program that is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.“
  (Tom Mitchell, 1997)

  „...is about extracting knowledge from data.“
  (Aurelion Geron, 2017)

  „...ist ein Oberbegriff für die „künstliche“ Generierung von Wissen aus Erfahrung.“
  (wikipedia, 22.8.2017)

Machine Learning
Positionierung

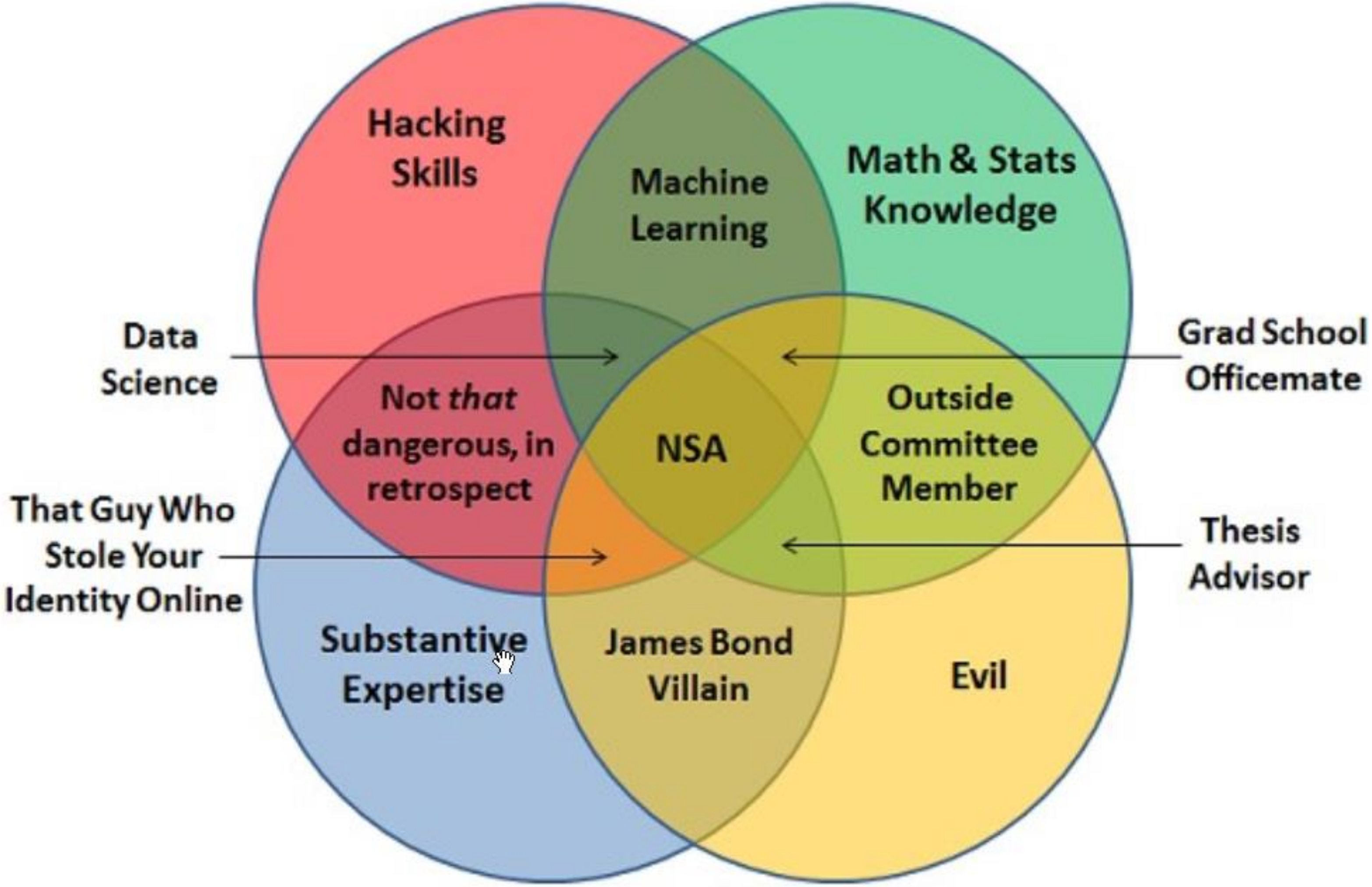
- Was ist "Machine Learning"?



Drew Conways Venn Diagram

Machine Learning

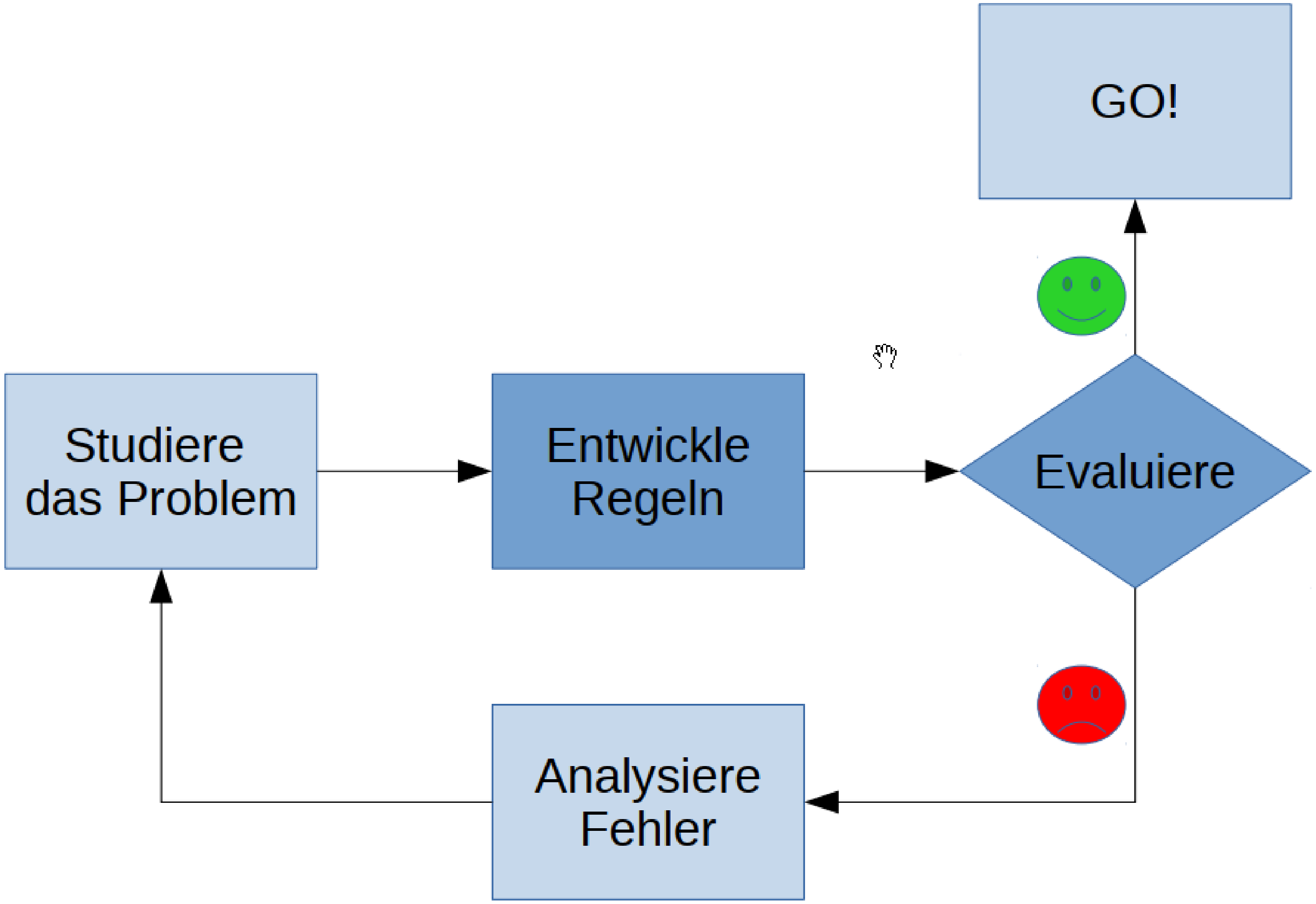
Positionierung



Joel Grues updated Venn Diagram

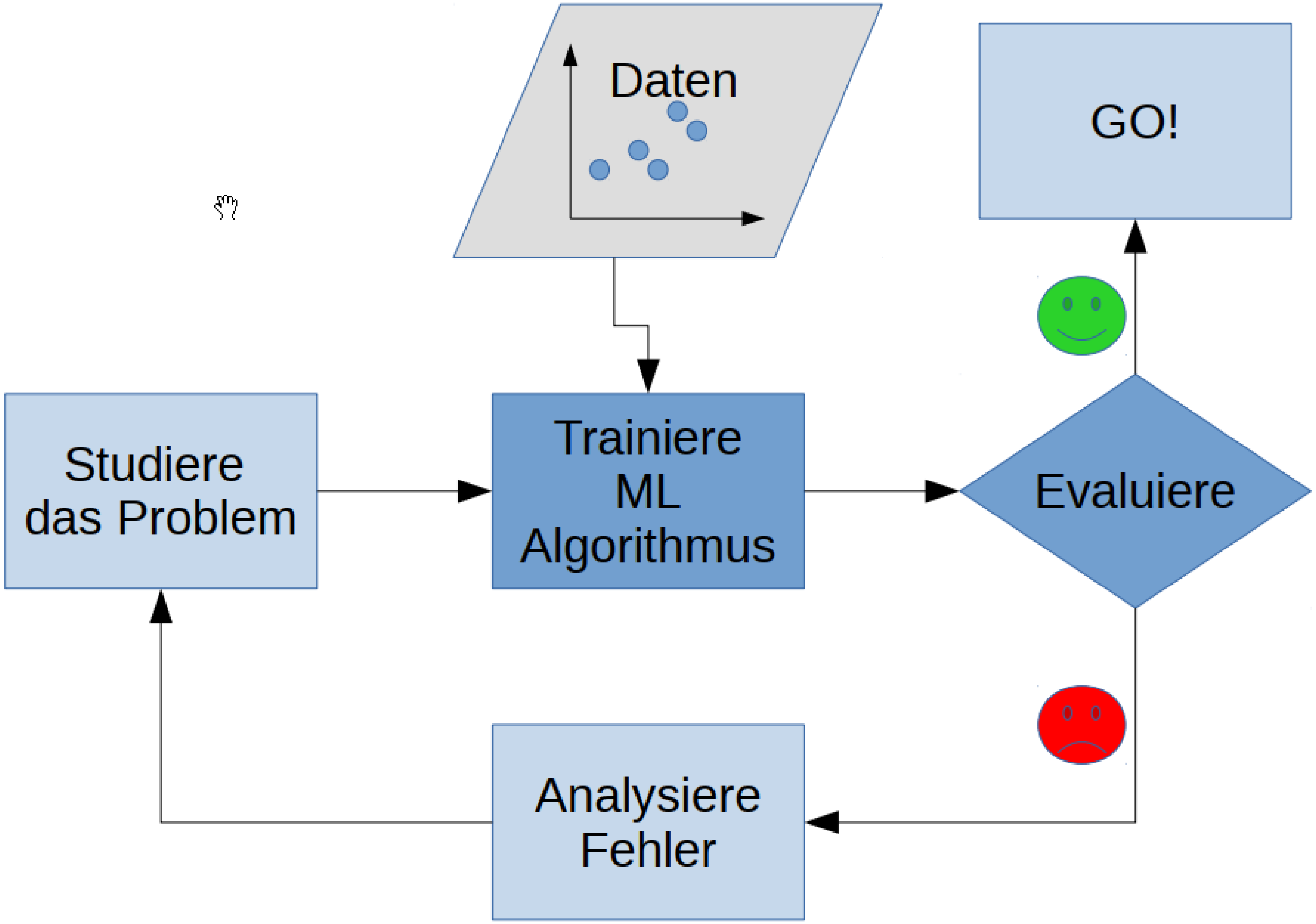
Machine Learning

- Traditionelle Methode



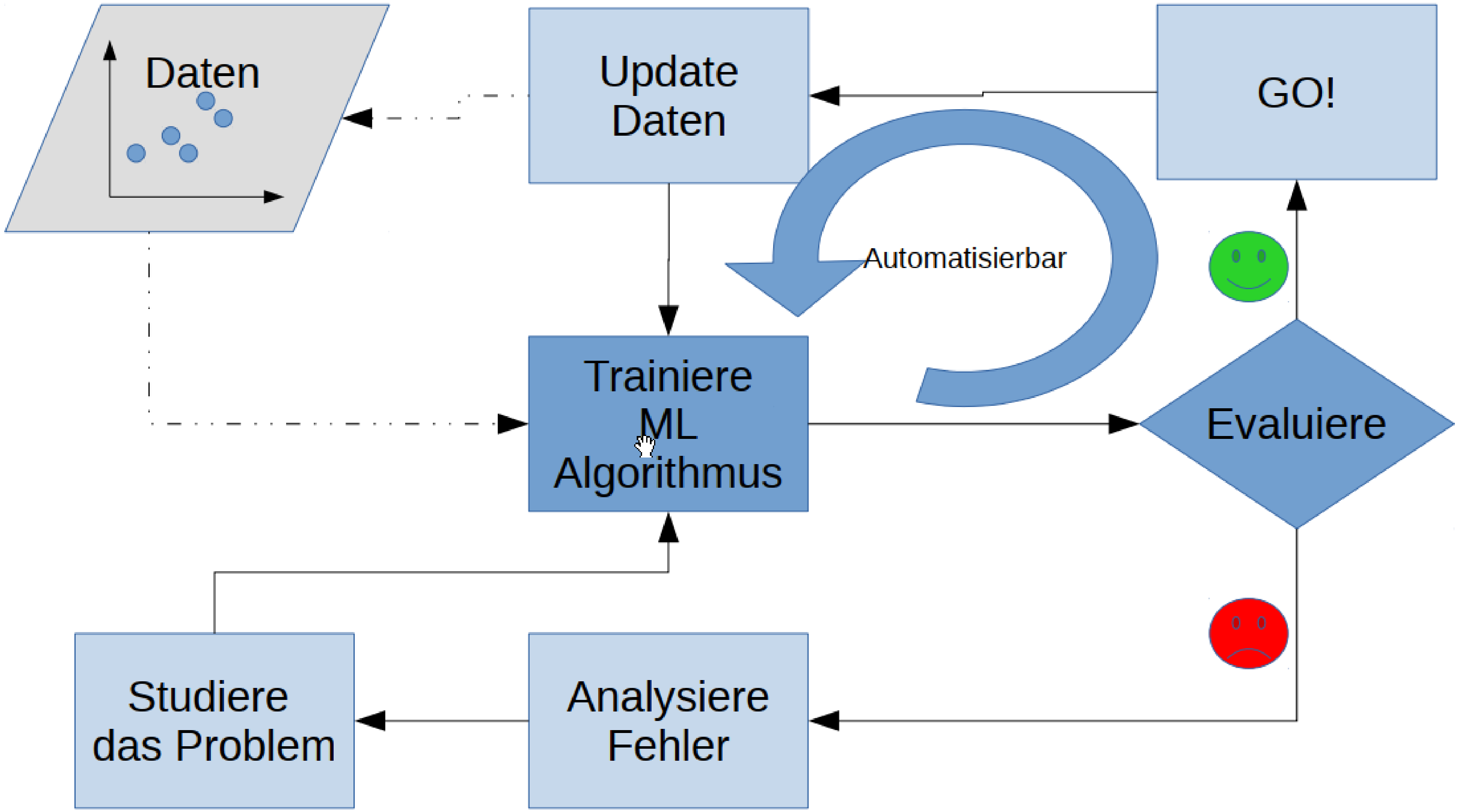
Machine Learning

- Machine Learning Methode



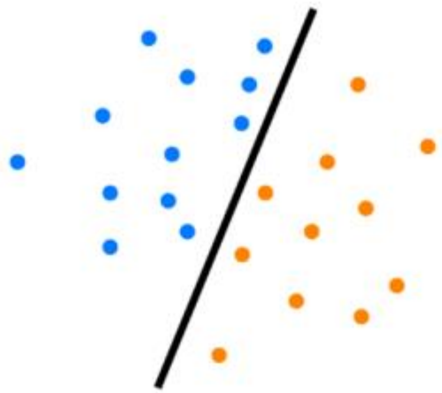
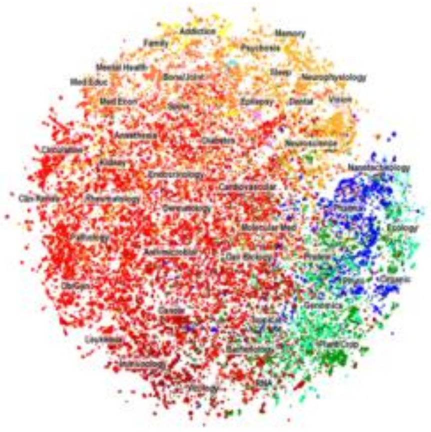
Machine Learning

- Machine Learning
  Methode
  (automatische Adaption auf Änderungen)



Machine Learning
Grundsätzliche Typen

Supervised	Unsupervised	Reinforcement
Learning known patterns	Learning unknown patterns	Generating data Learning patterns


Supervised Learning - Überwachtes Lernen

Zielfunktion

$f(x)$?

Trainingsdaten

$(x_1, y_1) \dots (x_N, y_N)$

Lernalgorithmus

Finale Hypothese

Hypothesenmenge

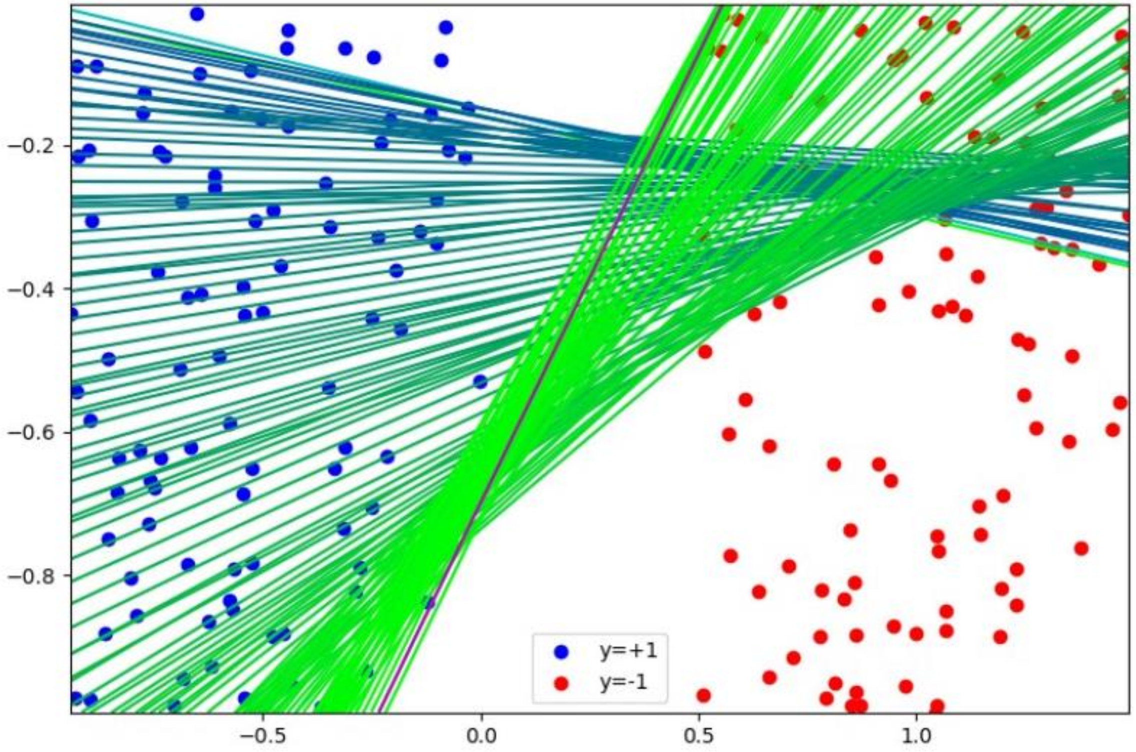
$\mathcal{H} = \{h_1, \dots\}$

Supervised Learning - Überwachtes Lernen
Hypothesen

- Hypothesen sind Abbildungen von den Trainingsdaten in die Zielgrösse

$$\hat{f}(x) \mapsto \hat{y}$$

- Beispiel binäre Klassifikation:
  Die Hypothesenmenge =
  {Alle möglichen Hyperebenen,
  die Positive und Negative
  Beispiele trennen}

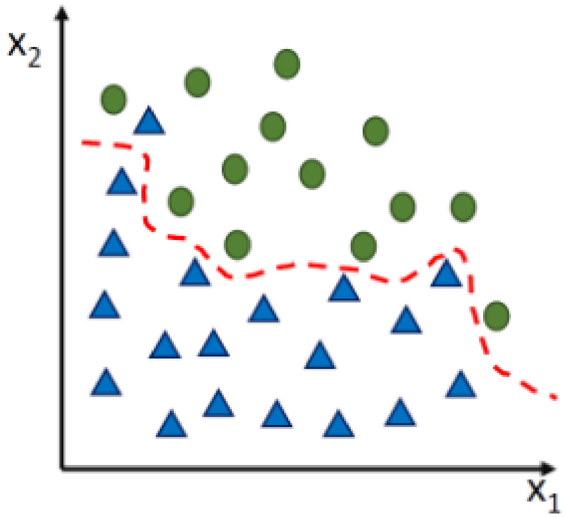


Supervised Learning - Überwachtes Lernen
Klassifikation vs. Regression

- Klassfikation:

  Zielvariable: Diskretes, qualitatives Label

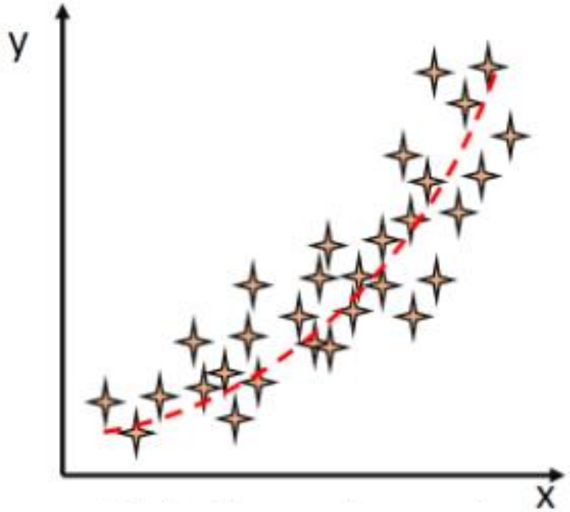
  Hypothesen haben eine endliche Wertemenge



- Regression:

  Zielvariable: Reelle Zahl

  Hypothesen haben eine reelle Wertemenge



Modelle

- Parametrische Modelle (z.B. $y=f(x)=x^3*a+x^2*b+x*c+d$):
  Beschreibe die Trainingsdaten durch eine fixe Anzahl an Parametern, brauchen daher im allgemeinen weniger Parameter.

  z.B.

  - Lineare und nicht-lineare Regression
  - Logistische Regression
  - Naive Bayes Klassifikation

Modelle

- Nicht-Parametrische Modelle (z.B.: KNN):
  Diese Modelle entwickeln eine Funktion, die im Zusammenhang zwischen Eingabe und Ausgabe stehen. Die Struktur der Funktion ist vor dem Training nicht über eine bestimmte Anzahl an Parametern festgelegt.

  z.B.

  - k-nächste Nachbarn Klassifikation/Regression (KNN)
  - Entscheidungsbaum
  - Nicht-lineare Support Vector Machines (RBF Kernel SVM)

Kommentare zu Modellkomplexität

- Berechnungs"komplexität"

Assuming k is fixed (as both of the linked lectures do), then your algorithmic choices will determine whether your computation takes $O(nd + kn)$ runtime or $O(ndk)$ runtime.

First, let's consider a $O(nd + kn)$ runtime algorithm:

- Initialize $selected_i = 0$ for all observations i in the training set
- For each training set observation i, compute $dist_i$, the distance from the new observation to training set observation i
- For $j = 1$ to k: Loop through all training set observations, selecting the index i with the smallest $dist_i$ value and for which $selected_i = 0$. Select this observation by setting $selected_i = 1$.
- Return the k selected indices

Each distance computation requires $O(d)$ runtime, so the second step requires $O(nd)$ runtime. For each iterate in the third step, we perform $O(n)$ work by looping through the training set observations, so the step overall requires $O(nk)$ work. The first and fourth steps only require $O(n)$ work, so we get a $O(nd + kn)$ runtime.

Now, let's consider a $O(ndk)$ runtime algorithm:

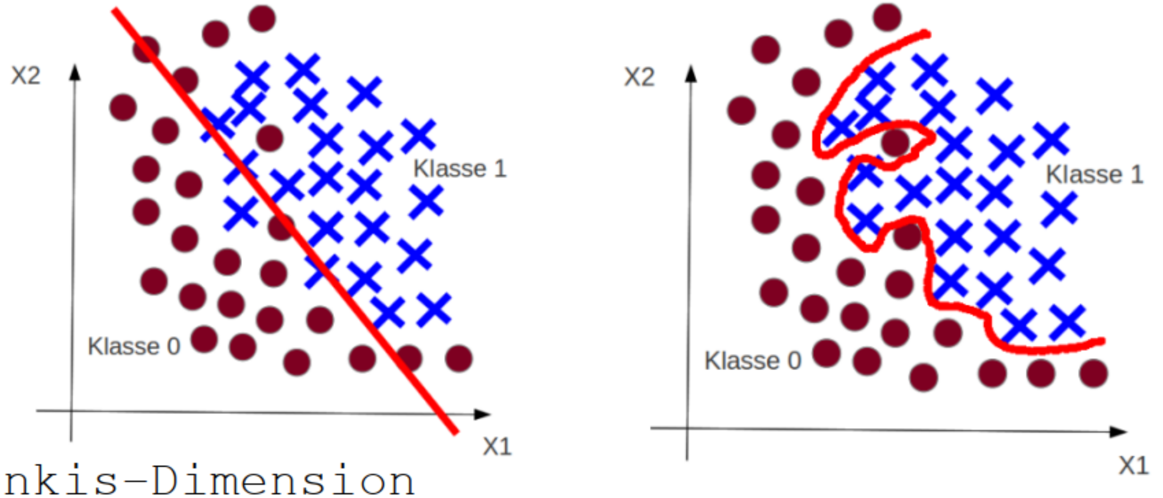
- Initialize $selected_i = 0$ for all observations i in the training set
- For $j = 1$ to k: Loop through all training set observations and compute the distance d between the selected training set observation and the new observation. Select the index i with the smallest d value for which $selected_i = 0$. Select this observation by setting $selected_i = 1$.
- Return the k selected indices

For each iterate in the second step, we compute the distance between the new observation and each training set observation, requiring $O(nd)$ work for an iteration and therefore $O(ndk)$ work overall.

Kommentare zu Modellkomplexität

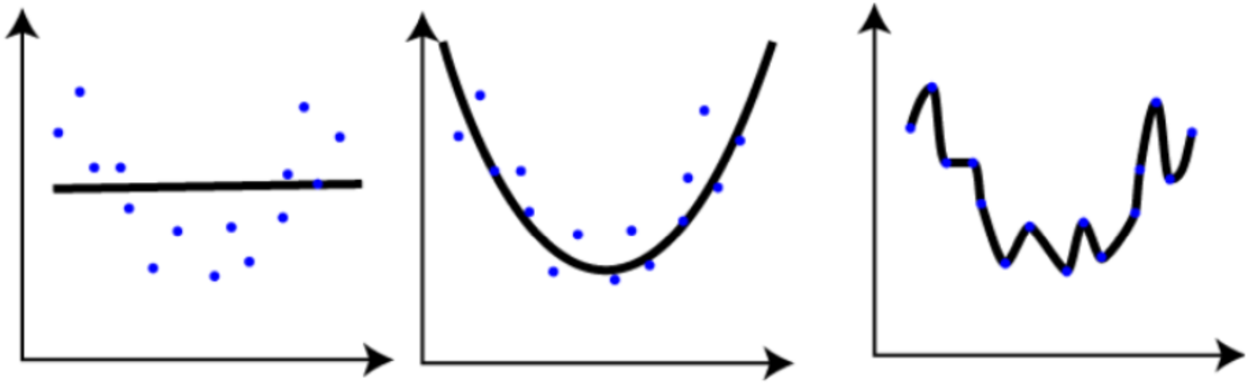
- Modellierungs"komplexität"

Klassifikation

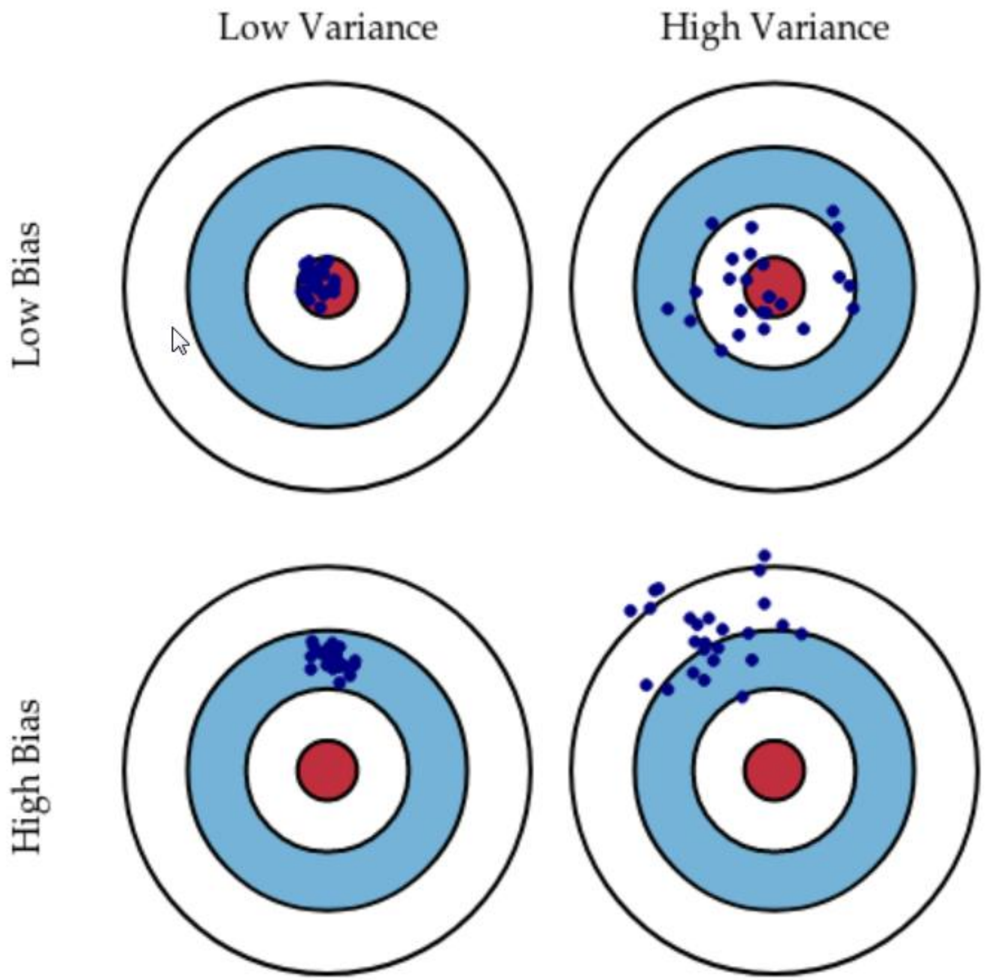


Vapnik-Chervonenkis-Dimension

Regression



Bias Variance <-> Overfitting



Bias Variance <-> Overfitting

$$Err(x) = E\left[(Y - \hat{f}(x))^2\right]$$

$$Err(x) = \left(E[\hat{f}(x)] - f(x)\right)^2 + E\left[\left(\hat{f}(x) - E[\hat{f}(x)]\right)^2\right] + \sigma_e^2$$

$$Err(x) = \mathrm{Bias}^2 + \mathrm{Variance} + \mathrm{Irreducible\ Error}$$