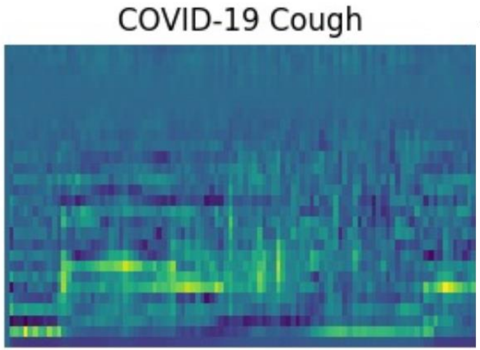
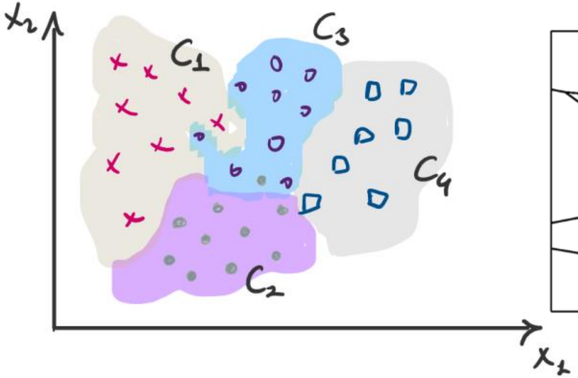
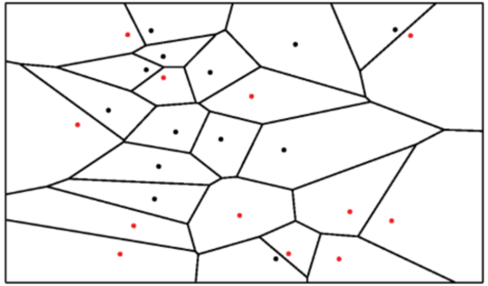
Data Driven Engineering I: Machine Learning for Dynamical Systems

Basics II: An Ode to Learning

Institute of Thermal Turbomachinery
Prof. Dr.-Ing. Hans-Jörg Bauer







Administrative Business

- Recorded lectures are online at ILIAS
- Lecture notes and active session notebooks
- Local installation guide available // Colab
- Project topics are uploaded >> check ILIAS

Today’s Agenda

① What is ML ?

② How does it work ?

③ What kinds of problems can be solved ?

④ What can go wrong here ?

Machine Learning

- ❑ **Definition**: automated **process** that extracts patterns from **data**

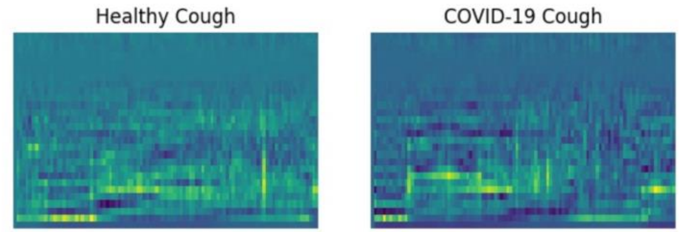
- ❑ AI >> ML := Data + Model

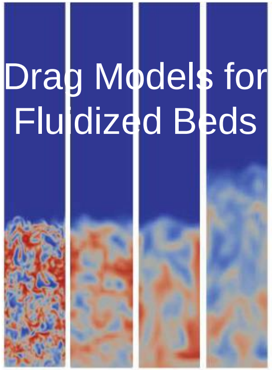
✓ Success of a learning algorithm depends on the data used
✓ Inherently related to data analysis and statistics
✓ Probability and optimization

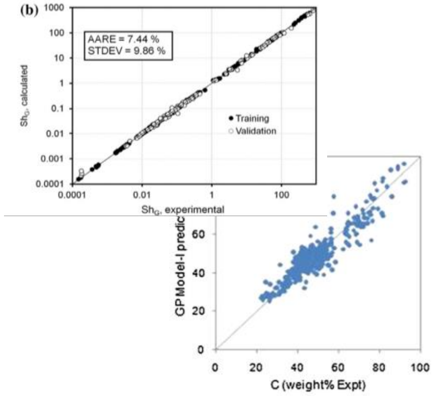
- ❑ **Model >>** Predictions **>>** help to make a decision
  - ✓ **Data analytics:** a prediction is the assignment of a value to any unknown variable
  - ✓ fundamentally about **generalization**
  - ✓ Temporal / static decision

"the field of study that gives computers the ability to learn without being explicitly programmed."
Arthur Samuel, 1959

Standard Learning Tasks-I

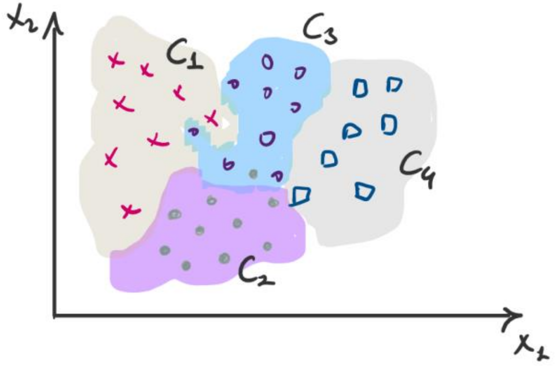


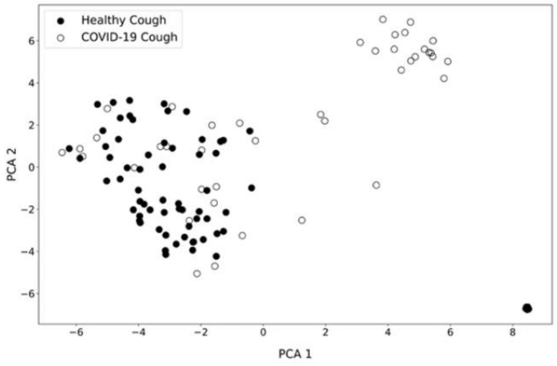




- **Classification:** problem of assigning a category to each item (discrete)
  - ✓ COVID-19 classifier of MIT
  - ✓ # Categories < 100
  - ✓ Unbounded classification: text classification, speech recognition
- **Regression**: problem of predicting a real value for each item (continuous)
  - ✓ Predicting the noise of an airfoil, turbofan predictive maintenance
  - ✓ Error estimation> difference between the true and predicted values

Standard Learning Tasks-II



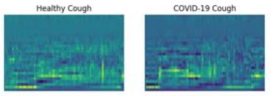


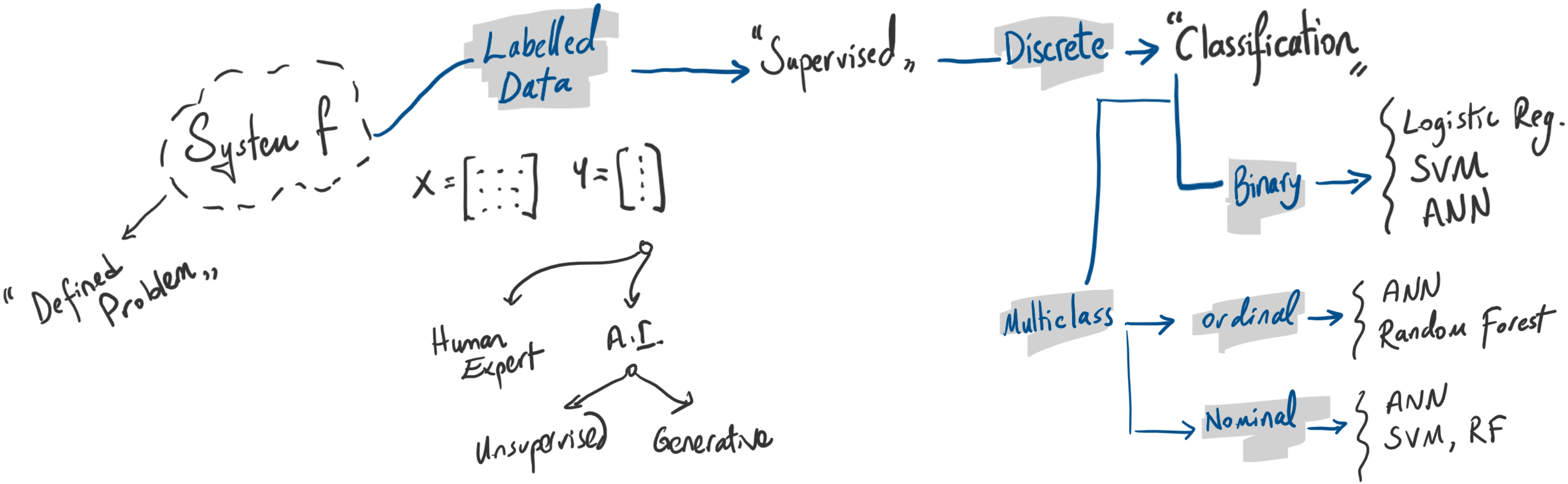
- ❑ **Clustering**: problem of partitioning a set of items into homogeneous subsets
  - ✓ Manufacturing error analysis
  - ✓ Very large data sets

- ❑ **Dimensionality reduction:** problem of transforming an initial representation of items into a lower-dimensional space
  - ✓ Manufacturing error analysis, image compression
  - ✓ Preserving the properties of the initial representation

How does it work?

"Different recipes for different problems"





How does it work?

"Different recipes for different problems"

System f

"Defined Problem"

Labelled Data

X = [ ] Y = [ ]

Human Expert

A.I.

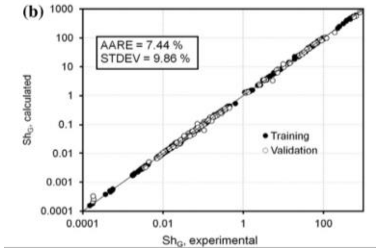
Unsupervised

Generative

"Supervised" — Discrete

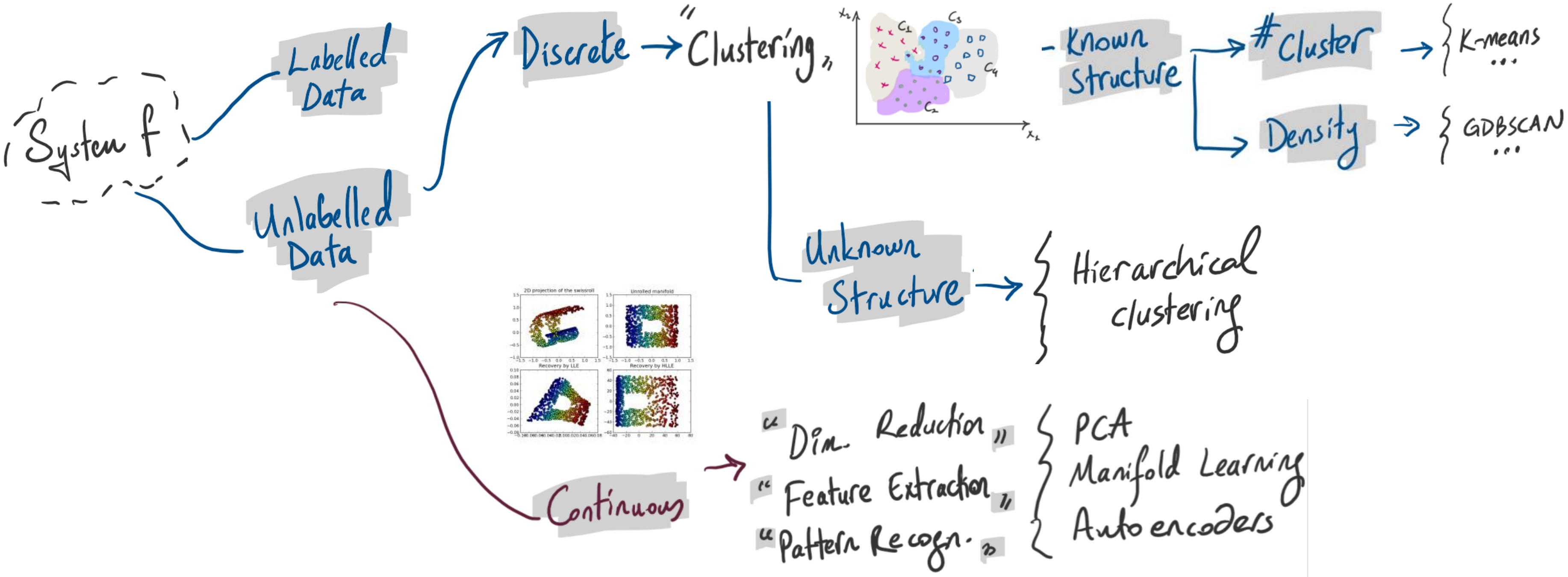
Continuous → "Regression"

Linear Reg. Methods
SVM, ANN
Bayesian Reg.



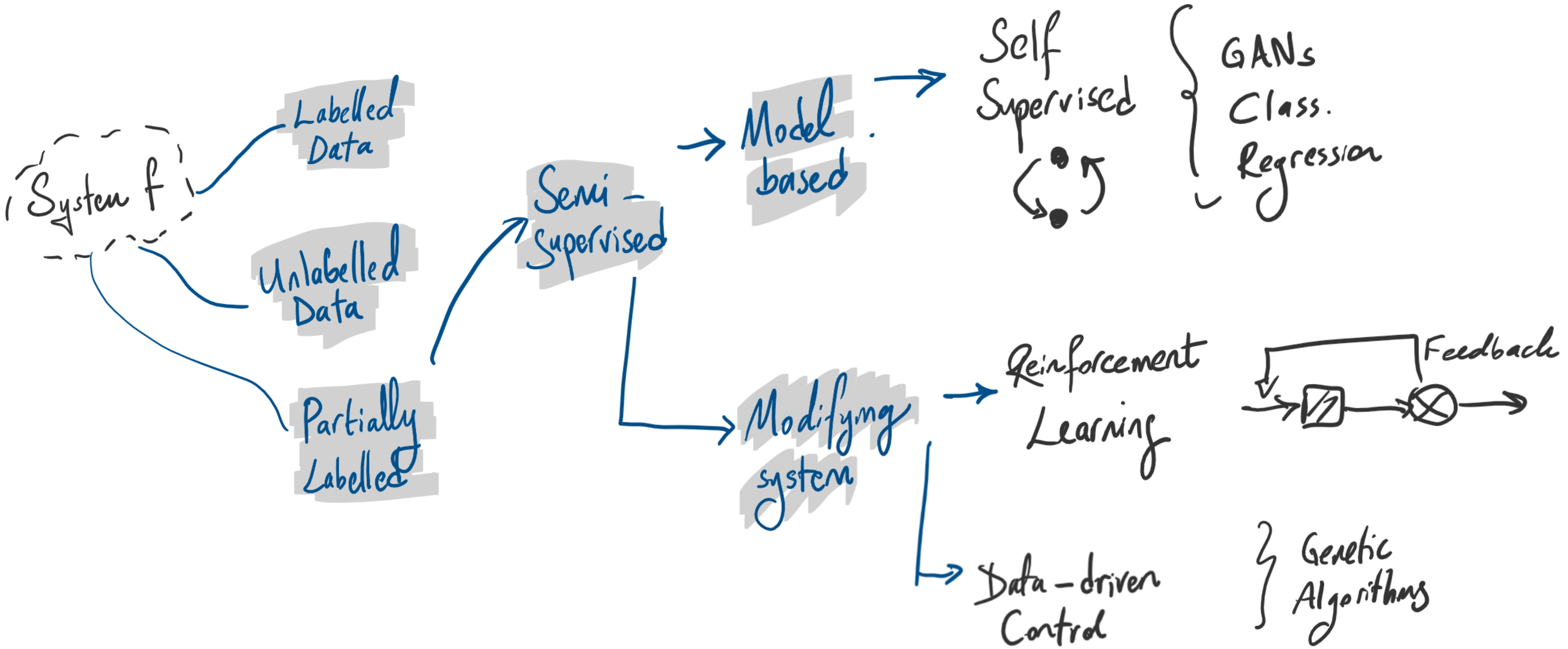
How does it work?

"Different recipes for different problems"



How does it work?

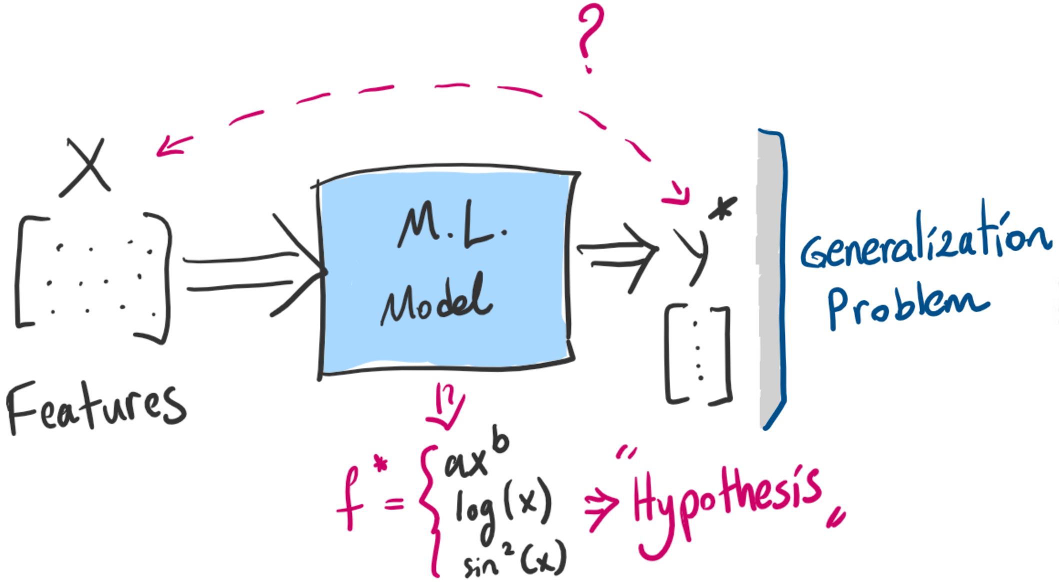
"Different recipes for different problems"



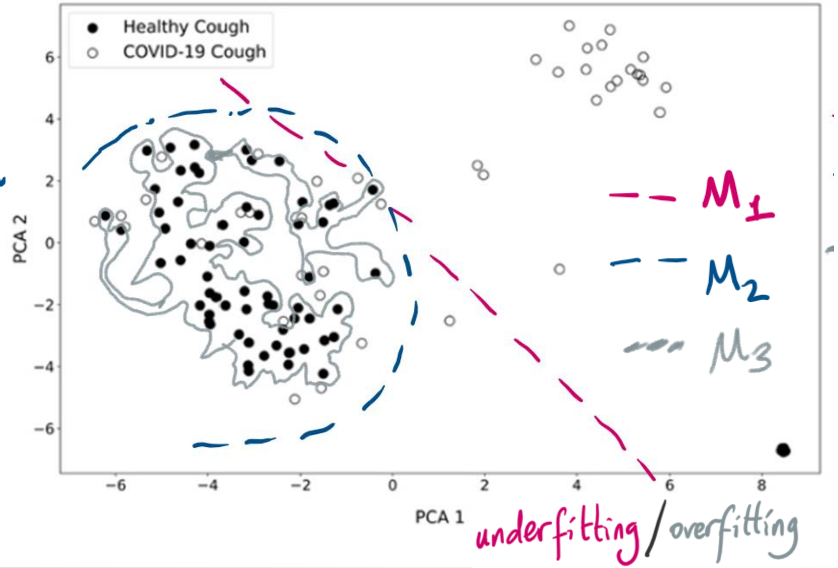
How can I select a model?

"Many recipes for the same problem"

ML is an "ill-posed problem"



How can a hypothesis be chosen?



How can I select a model?

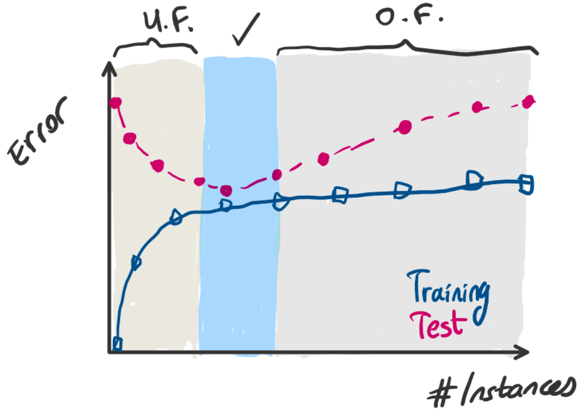
"Many recipes for the same data"

How can we pick the right model?

We can calculate the error ?

↳ Prediction error ⇒ "overfitting

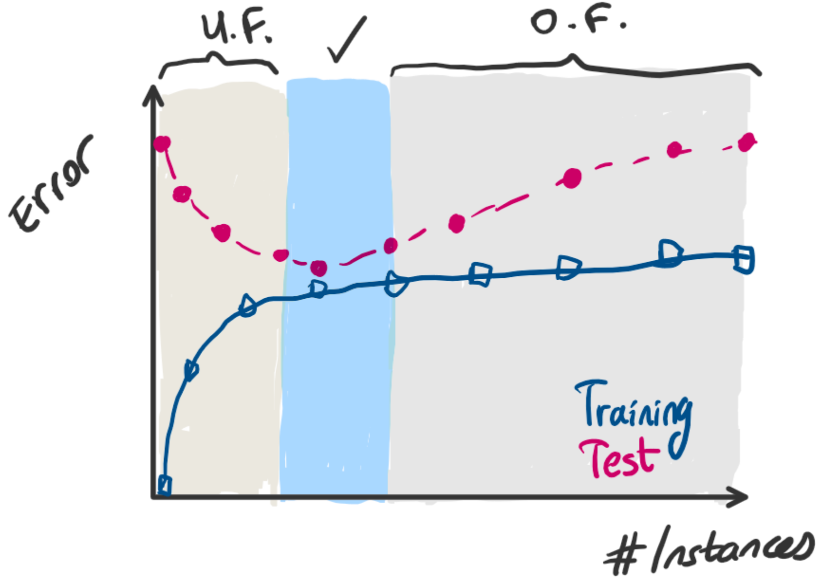
↳ Objective : Minimize the generalization error



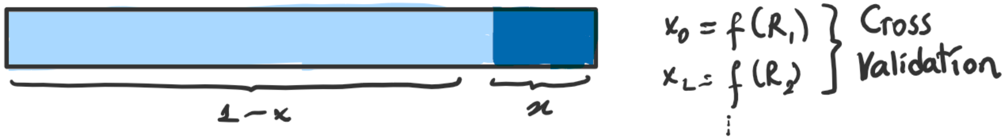
How can I select a model?

"Many recipes for the same data"

How can we pick the right model?

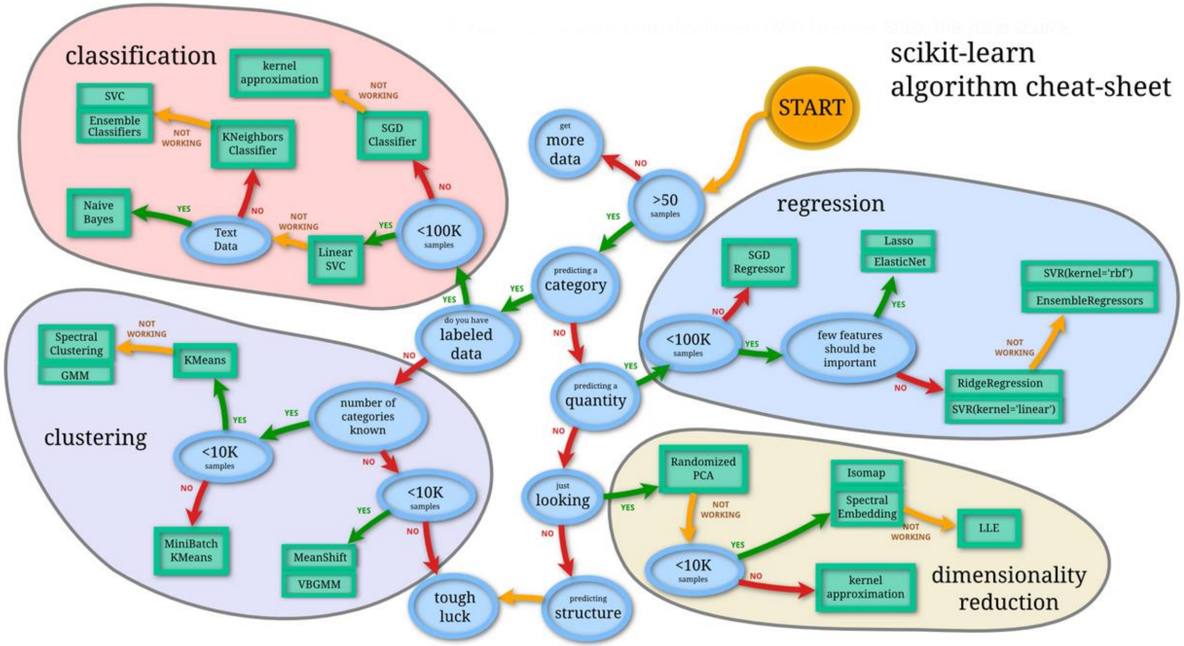


* Problem: We can not use test set to pick the right degree of complexity.



How does it work?

"Different recipes for different problems"



No free lunch theorem

Wolpert, 1996

There is no universally best model

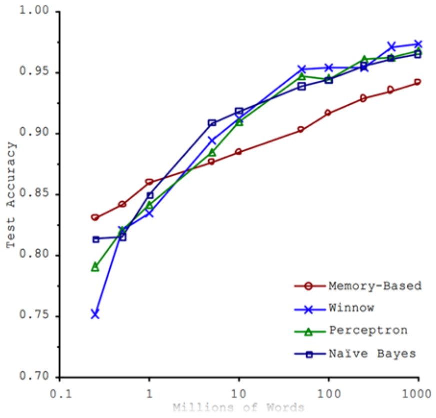
Assumptions that works well in one domain may work poorly in another.

How can I select a model?

“Insufficient **Data** for the Training”

- ❑ **High accuracy:** high volumes of data required
  - ✓ very simple problems ~ 1000s examples
  - ✓ Complex problems ~ millions of examples
- ❑ **Large datasets** >> computational burden
  - ❑ SoA Deep Learning training: Energy eq. of the electricity consumption of a city for a few days
- ❑ **Why bother with larger datasets?**

“The Unreasonable Effectiveness of Data”



“Scaling to Very Very Large Corpora for Natural Language Disambiguation”

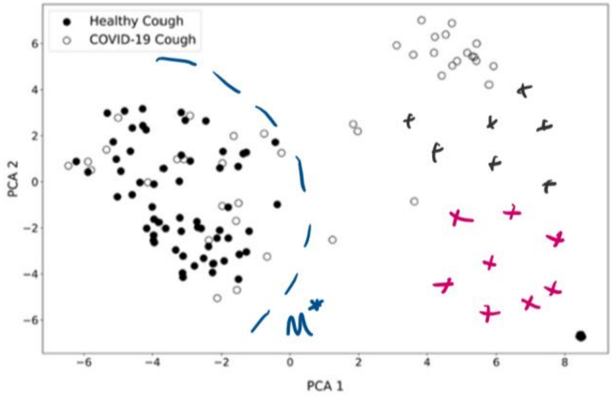
Michele Banko and Eric Brill
Microsoft Research
2001

- ✓ Size of the data mattered far more than the choice of ML approach
- ✓ Differences became very small as the data grew large

How can I select a model?

"Insufficient **Quality** for the Training"

- ❑ **Representative data**
  - ❑ training data ~ new cases
  - ❑ Sample is too small >> "sampling noise"
  - ❑ Sampling method flawed >> "sampling bias"



- ❑ **Data quality**
  - ❑ errors, outliers, and noise (sensor, model)
  - ❑ spend time cleaning up (outliers)
  - ❑ missing features (ignore / guess / omit)
  - ❑ Dimen. Reduction
  - ❑ "Feature engineering": select, extract, combine

$Re = \frac{L_c \cdot v \cdot \rho}{\mu}$

$Pr = \nu/\alpha$

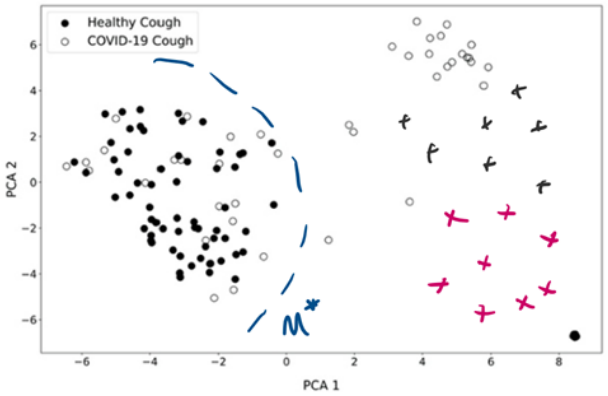
$Nu = f(Re, Pr)$

$L_c := h_d$?

How can I select a model?

"The **curse** of **dimensionality**"

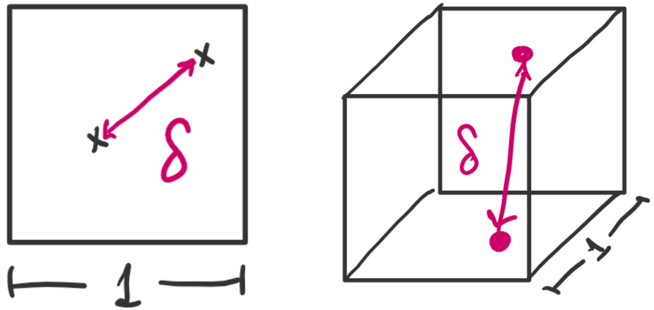
- ❑Typical ML feature space ~ millions
  >> higher dimensional space
  >> How can I "**draw**" separation "**curve**"



✓Pick two points randomly in a unit square:
>> the distance ~ 0.52

✓Pick two random points in a unit 3D cube:
>> the distance ~ 0.66

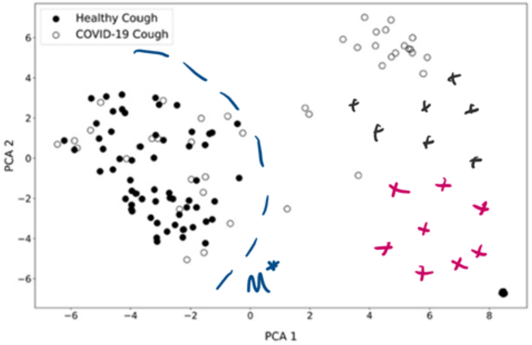
✓Pick two random points in a unit 1M-D hypercube:
>> the distance ~ 408



How can I select a model?

"The **curse** of **dimensionality**"

- ❑Typical ML feature space ~ millions
  - >> higher dimensional space
  - >> How can I "**draw**" separation "**curve**"



Why this is a problem?

- ❑New instance will be far away from any training example
- ❑Higher the dimension ~ greater the risk of overfitting
  - >> Solution 1: increase the size of the training set
  - >> Solution 2: dimensionality reduction

Today's Agenda

- ✓ What is ML ?
- ✓ How does it work ?
- ✓ What kinds of problems can be solved ?
- ✓ What can go wrong here ?

Then how should I approach the problem ?

The Style

How does it work?

"PFD of a ML project"

Task I

* A new item on the menu

"Local Ingredients"

#0: Understand the business/problem/task.

How does it work?

"PFD of a ML project"

What ingredients do I have?

Types?

Properties?

Task

Recipe

(ü)

(i)

Ingredients (good)

#1: Understand the data: The Sources available & the type.

How does it work?

"PFD of a ML project"

Remove the spoiled food

The form: pealed; chopped boiled; frozen

size, shape, texture

"Recipe"

↓ Data: Type, Scale Structure

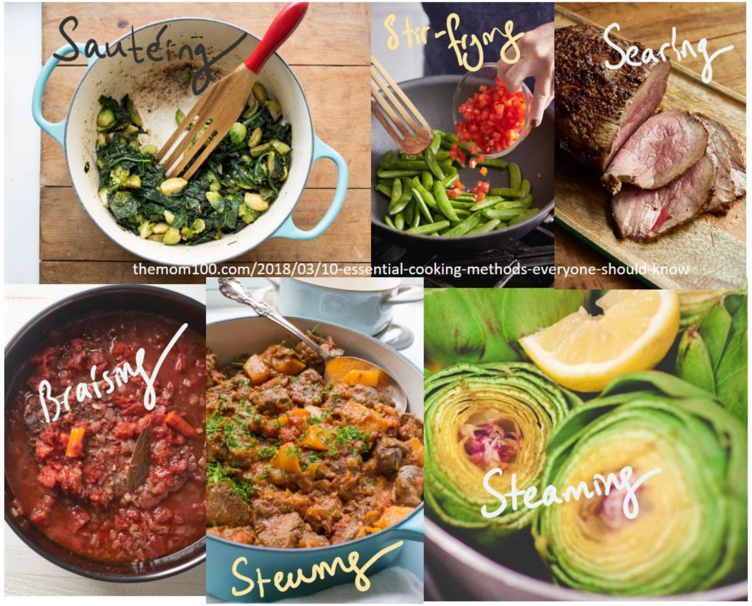
ML

↓ Predictions

#2 Data Preparation / Exploratory Data Analysis

How does it work?

"PFD of a ML project"



⇒ "Chosing the right cooking method"

Data

⇓

Model → $\begin{bmatrix} M_1 \\ M_2 \\ M_3 \\ \vdots \end{bmatrix}$

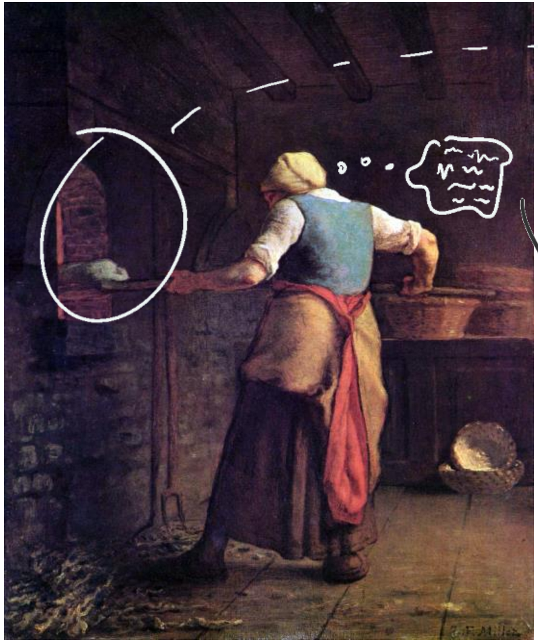
⇓

Prediction

#3 Modeling: Try different ML approaches

How does it work?

"PFD of a ML project"



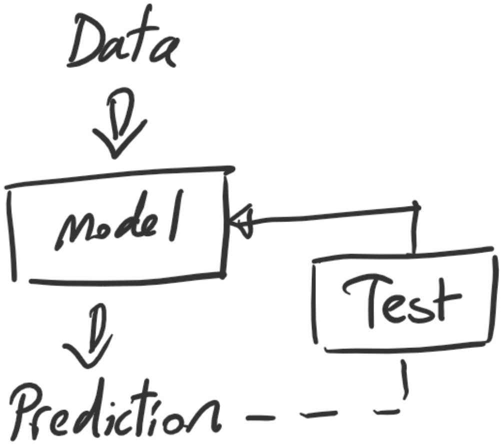
Cook the dish by using the recipe

Compare its taste to the dishes tasted before.

How much creative you can be?

↳ Pure randomness

↳ "Regularization"



#4 Training the ML model

How does it work?

"PFD of a ML project"

⇒ how satisfied are you with the taste?

↳ Adjust ingredients

↳ Add more "complexity"

↳ Change the cooking method

"Feature Engineering"

Data → model → Predictions

$\begin{bmatrix} M_1 \\ M_2 \\ M_3 \\ \vdots \end{bmatrix}$ models

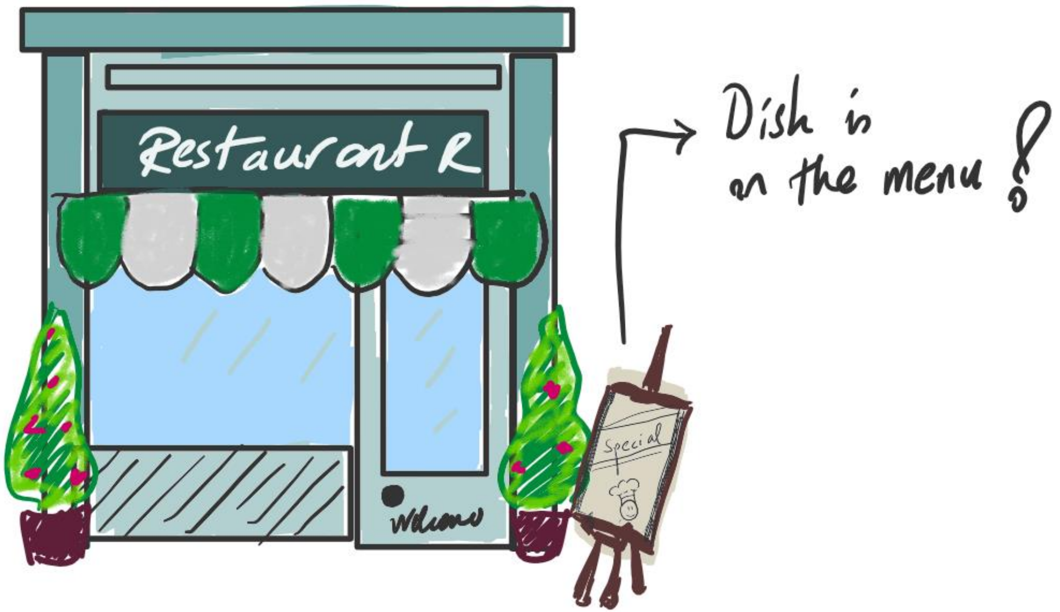
$\begin{bmatrix} \alpha \\ \beta \\ \gamma \\ \vdots \end{bmatrix}$ hyperparameters

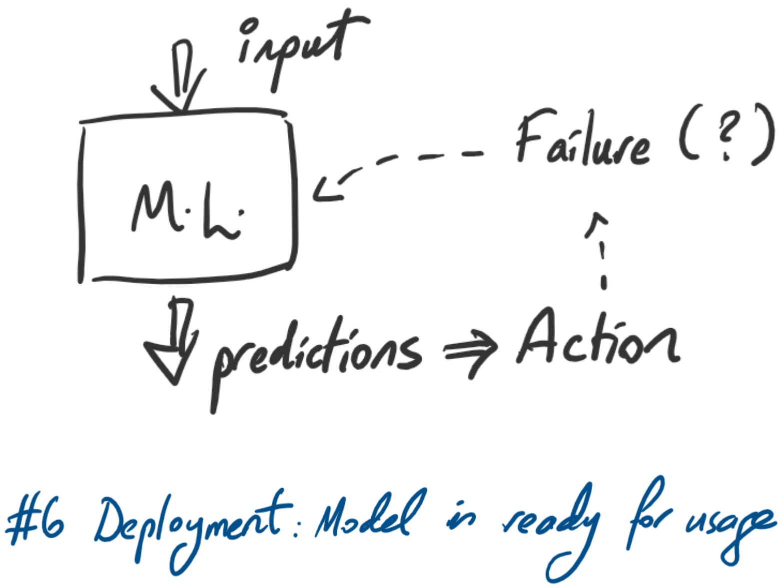
$\begin{bmatrix} l_1 \\ l_2 \\ C \\ \vdots \end{bmatrix}$ Regulation

#5 Evaluation: capable of making accurate predictions

How does it work?

"PFD of a ML project"





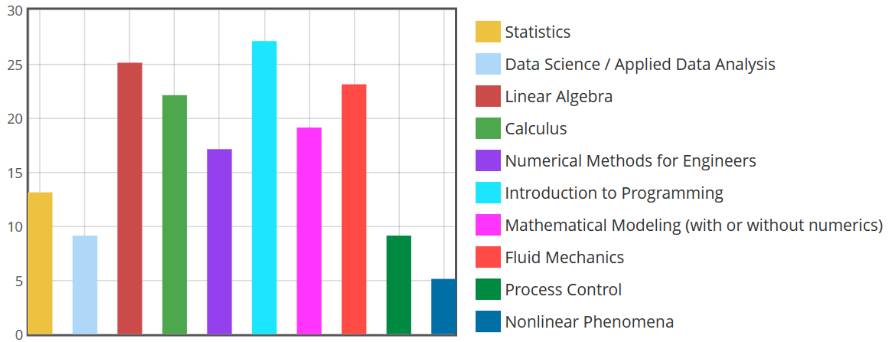
Colab: An Introduction

- ❑ Introduction to Python with Colab

Objective:

✓ Introduce Colab environment

✓ Introduce some basics



Today's Agenda

- ✓ Machine Learning: Overview, Means and Goals
- ✓ Problem Solving and Reasoning
- ✓ Planning: How a ML project is organized
- ✓ Theory of Learning and Learning Types
- ✓ Decision Theory

- ❑ Project datasets >> ILIAS
- ❑ Local installation guide >> ILIAS
- ❑ Next Week: Classification Methods in ML **with active session!**

Additional Notes

Some important keywords to know…

- **Examples:** Items or instances of data used for learning or evaluation
- **Features:** The set of attributes, often represented as a vector, associated to an
- example.
- **Labels:** Values or categories assigned to examples. In classification problems,
- examples are assigned specific categories, (e.g. healthy // sick)
- **Hyper-parameters:** Free parameters that are not determined by the learning algorithm, but rather specified as inputs to the learning algorithm.
- **Training sample:** Examples used to train a learning algorithm.

Some important keywords to know…

- **Validation sample:** Examples used to tune the parameters of a learning algorithm when working with labelled data.
- **Test sample:** Examples used to evaluate the performance of a learning algorithm. The test sample is separate from the training and validation data and is not made available in the learning stage.
- **Loss function:** A function that measures the difference, or loss, between a predicted label and a true label.
- **Hypothesis set:** A set of functions that maps the features (feature vectors) to the set of labels (assumed relationship possibilities between the features and labels).

Learning Theory 1: Information-based Learning

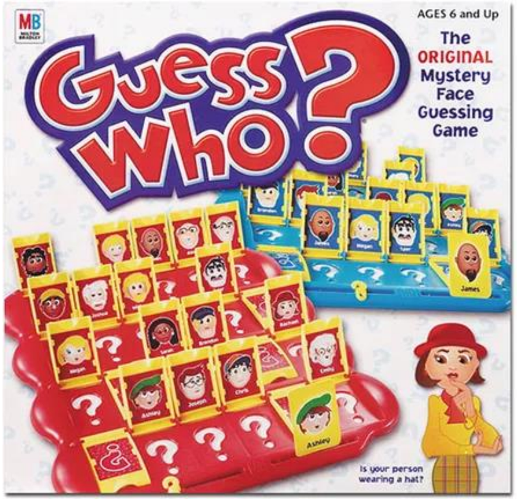
Information-based algorithms determine which descriptive features provide the most information and make predictions by sequentially testing the features in order of their informativeness

Example: Decision trees
Key concept: entropy

a. Figure out which features are the most informative ones

b. Ask questions about by considering the effects of the different answers

c. how the domain is split up after the answer is received and the likelihood of each of the answers.



boardgamegeek.com

Learning Theory 2: Similarity-based Learning

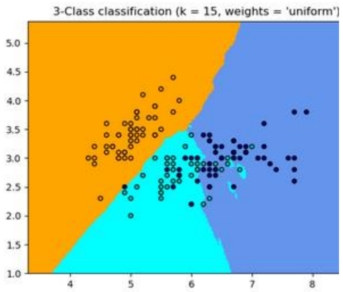
- ❑ Idea: look at what has worked well in the past and predict the same
- ❑ Method: build a feature spaces and measure the similarity

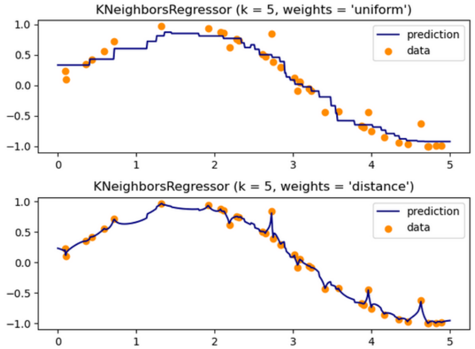
“When I see a bird that walks like a duck and swims like a duck and quacks like a duck, I call that bird a duck.” J. Raley

Example: nearest neighbor algorithm

Key concept: the mean to measure distance in many dimensions (feature space)

- ✓ *Euclidean distance*
- ✓ *Manhattan distance*
- ✓ *Minkowski distance*





scikit-learn.org/stable/modules/neighbors.html#

Learning Theory 3: Probability-based Learning

- Heavily based on Bayes' Theorem
- Estimates of likelihoods to determine the most likely predictions
- Revise these predictions based on data /extra evidence when available

Key concept: Bayes' Theorem

Example: The Naive Bayes Model, Bayesian networks

"Find the lady"

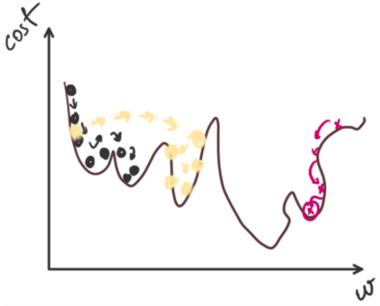
Check: https://seeing-theory.brown.edu/bayesian-inference/index.html#section1

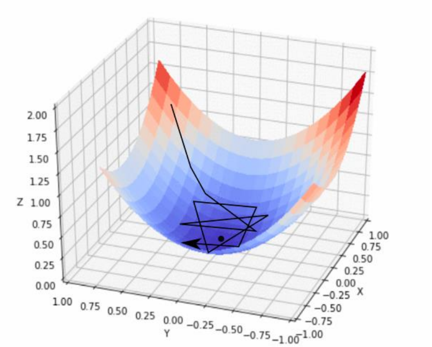
Learning Theory 4: Error-based Learning

- ❑ Search for a set of parameters that minimizes the total error across the predictions
- ❑ Need a set of training instances for the optimization process

Key concepts: measuring the error
navigating on the error surface

Methods: sum of squared errors, MSE, MAE, R^2 …
gradient decent





blog.paperspace.com/intro-to-optimization-in-deep-learning-gradient-descent/