Berner Fachhochschule
Haute école spécialisée bernoise
Bern University of Applied Sciences

CAS Practical Machine Learning
Introduction

Machine Learning Overview

Prof. Dr. Jürgen Vogel (juergen.vogel@bfh.ch)

Machine Learning

Learning

- observe the real world
- and derive knowledge
  - facts ("this image shows a person")
  - rules ("images show a person if there is a shape with the properties X, Y, and Z")
- that can be applied to new situations

Machine Learning

- informal

  an algorithm that implements learning
- formal

  an algorithm learns from experience `E` to solve some tasks `T` with performance `P` if `P` improves with `E`

Machine Learning vs. AI

Artificial Intelligence (AI)

= Systems that…

…act…

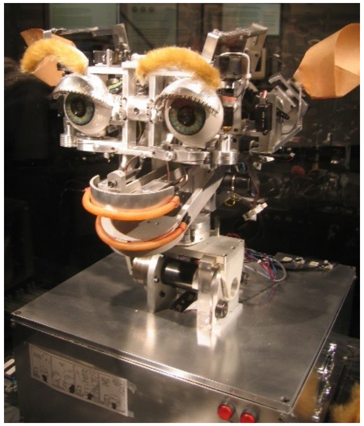
- …like a human (Turing test)
- …rationally (optimization)

…think…

- …like a human (cognitive science)
- …rationally (logic)

…can…

- …solve problems (search solutions)
- …derive new knowledge from prior one (reasoning)
- …**learn** from experience
- …interact with their environment (sense and manipulate)



Kismet @ MIT

Machine Learning Models

an algorithm learns from experience `E` to solve some tasks `T` with performance `P` if `P` improves with `E`

- Model
  - represents the solution to the tasks `T`
  - is learnt by the ML algorithm
  - is adapted by the ML algorithm based on `E`
  - can be evaluated with respect to `P`
  - can be stored
  - may be human-readable or not (white box vs. black box)
- Features
  - are the relevant part of the data `E` for creating the model
  - may have to be designed explicitly depending on the ML algorithm

Age	Spectacle prescription	Astigmatism	Tear production rate	Recommended lenses
Young	Myope	No	Reduced	None
Young	Hypermetrope	No	Normal	Soft
Pre-presbyopic	Hypermetrope	No	Reduced	None
Presbyopic	Myope	Yes	Normal	Hard
…	…	…	…	…

```
If tear production rate =
    reduced
    then recommendation = none
Otherwise, if age = young and
    astigmatic = no
    then recommendation = soft
```

Machine Learning vs. Traditional Programming

Traditional Programming

- transforms input data into output data via a series of fixed instructions (conditions, loops etc.)
- white box problem solving

Machine Learning

- maps input data to output data via an adaptive model
- black box problem solving

benefits

- deterministic and relatively easy to test in well-defined environments
- avoid errors introduced by random noise
- deterministic results

- encounter complex or even unknown environments or situations
- avoid misconceptions about the real world
- continuous result improvements

Machine Learning vs. Statistics

Statistics

- theoretical foundations for analyzing and interpreting data
- formalizes relationships between variables via mathematical equations

Machine Learning

- algorithms that improve automatically through experience
- facilitates statistical and mathematical methods
- sacrifices correctness for computability

some terms

Statistics:

- estimation
- hypothesis testing
- data point
- independent variable
- dependent variable

Machine Learning:

- learning
- classification
- instance/example
- feature
- class/label

Machine Learning Applications

- information retrieval
  - personalized search
- marketing and sales
  - product recommendation
- meteorology
  - weather forecast
- medicine
  - diagnosis
- finance
  - (stock) portfolio management – buy or sell
- industry
  - predictive maintenance
- transport
  - self-driving cars
- security and law enforcement
  - biometric authorization and identification
- …

Different Machine Learning Approaches (1)

What experience `E` can be exploited for learning?

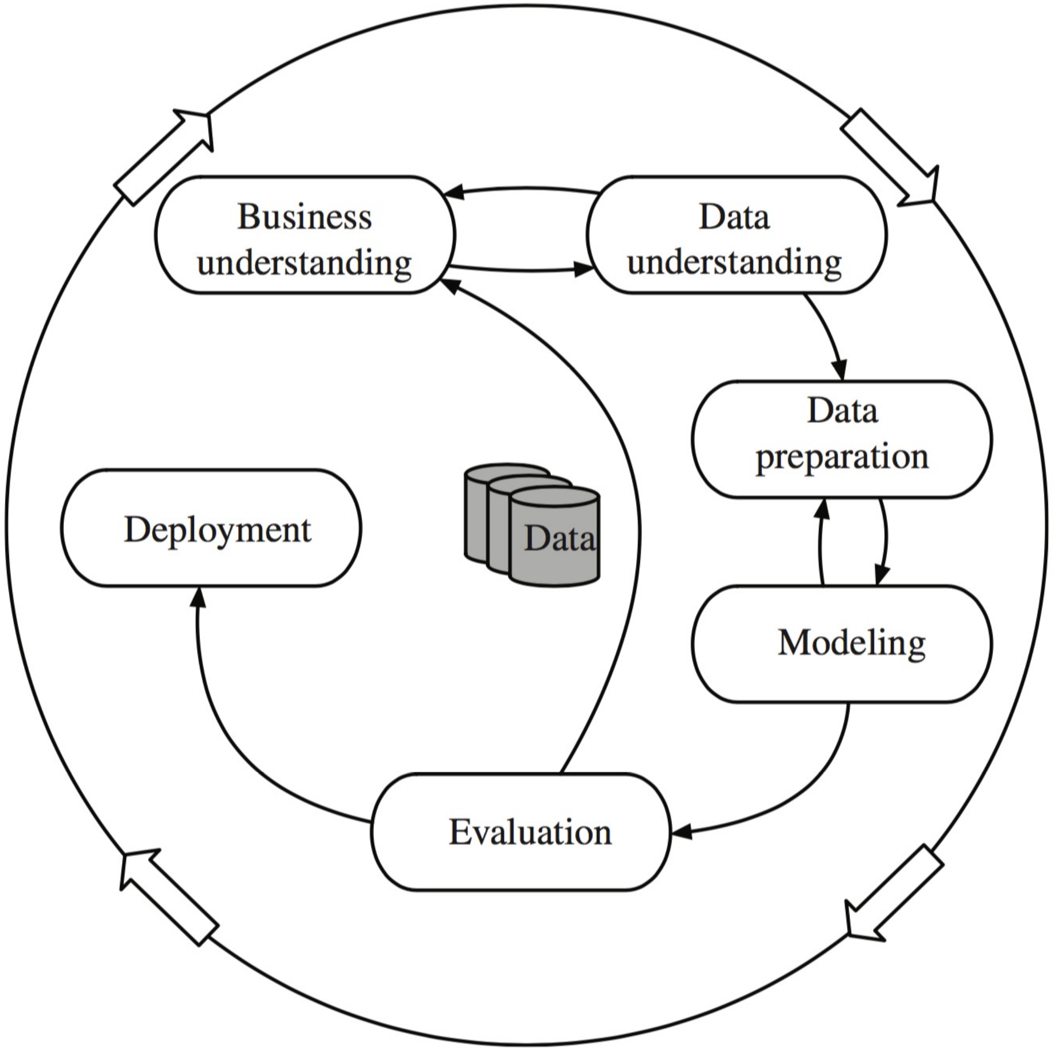
1. Supervised Learning
   - the ML algorithm infers the model from sample data `E` for which the task `T` has been solved with optimal performance `P`
   - the algorithm is taught by example
2. Unsupervised Learning
   - the ML algorithm infers the model from data `E` based on some adaptation to distinctive features of `E`
   - the algorithm does not have access to `P`
3. Reinforcement Learning
   - the ML algorithm starts with a baseline model to solve `T` and continuously improves it based on feedback `P`
   - feedback may be in the form of (short-term/long-term) rewards or punishments
   - the algorithm interacts with its environment and explores the consequences of its actions

Different Machine Learning Approaches (2)

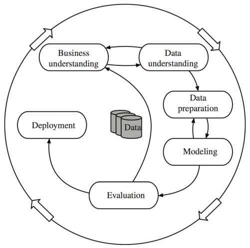
What type of task T is to be solved?

1. Clustering
   - input data should be divided into distinctive groups
   - model often based on similarity: members in one group are similar
   - e.g., identifying similar genes in a genome
   - usually solved via unsupervised learning
2. Classification
   - input data should be assigned a certain class/category/label
   - model often based on distinctive features
   - e.g., an email is classified as *spam* or *not-spam*
   - usually solved via supervised learning
3. Regression
   - estimates the relationship between input data (independent variables) and the output we are interested in (dependent variable)
   - model is a regression function, often for continuous variables
   - e.g., prediction of birth rate
   - usually solved via supervised learning / statistical methods

Machine Learning Workflow (CRISP-DM)



Business Understanding



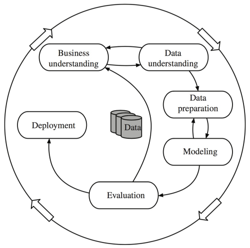
Defining the Problem

- machine learning is about solving a very distinctive problem
  - predict the remaining lifetime of a machine
  - classify an email as spam or not spam
  - detect a human face in an image
- unfortunately, we often hear something like: *we have a ton of XYZ data – can you make something interesting out of it?*
  - data science: iteratively addressing the information need
- machine learning not necessarily means automation
  - machine learning generates information
    - e.g., in a supermarket many shoppers buy beer and chips
  - different actions and decisions may be taken based on this
    - e.g., place beer and chips together (“easy shopping”) or further apart (“maximize store time”)

Example: To Wait or Not to Wait

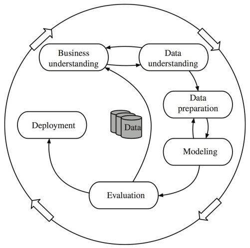
Will people wait in line for eating in a restaurant?

Understanding and Preparing Data (1)



- collecting data
  - internal or external sources
  - from target or related domain
  - need to be labeled for supervised training
- understanding data
  - which aspects of available data can be utilized as features
  - feature types: Boolean, nominal (categorial), or numeric
  - often helpful to look at basic statistics to develop a “feeling”
    - histograms and frequency distributions
    - graphs
  - many classification problems are unbalanced
    - classes are not evenly distributed but a class may be very rare or very frequent
      - e.g., detecting rare diseases: only x cases in a million
      - may not even occur in dataset at hand

Understanding and Preparing Data (2)



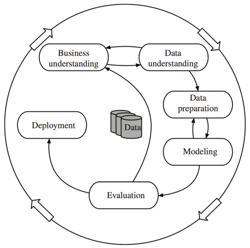
- preparing (cleaning) data
  - data integration: data from different sources may use different schema (syntax) or semantics
  - data transformation: e.g., turn numeric feature ("age") into nominal ("age category")
  - data may be missing: e.g., some features have not been recorded for some instances
  - data may be wrong: typos, measurement errors, duplicates, deliberate errors, …
  - data may be obsolete: e.g., too old
  - feature generation: processing low-level data into higher-level features, e.g., from pixels to edges in image analysis
  - feature selection: selecting the best features from a (too) large set

Example: To Wait or Not to Wait

Sample	Alternative	Bar	Fri/Sat	Hungry	Patrons	Price	Rain	Reservation	Type	Estimated Waiting Time	Wait
1	T	F	F	T	Some	$$$	F	T	French	0–10	T
2	T	F	F	T	Full	$	F	F	Thai	30–60	F
3	F	T	F	F	Some	$	F	F	Burger	0–10	T
4	T	F	T	T	Full	$	F	F	Thai	10–30	T
5	T	F	T	F	Full	$$$	F	T	French	>60	F
6	F	T	F	T	Some	$$	T	T	Italian	0–10	T
7	F	T	F	F	None	$	T	F	Burger	0–10	F
8	F	F	F	T	Some	$$	T	T	Thai	0–10	T
9	F	T	T	F	Full	$	T	F	Burger	>60	F
10	T	T	T	T	Full	$$$	F	T	Italian	10–30	F
11	F	F	F	F	None	$	F	F	Thai	0–10	F
12	T	T	T	T	Full	$	F	F	Burger	30–60	T

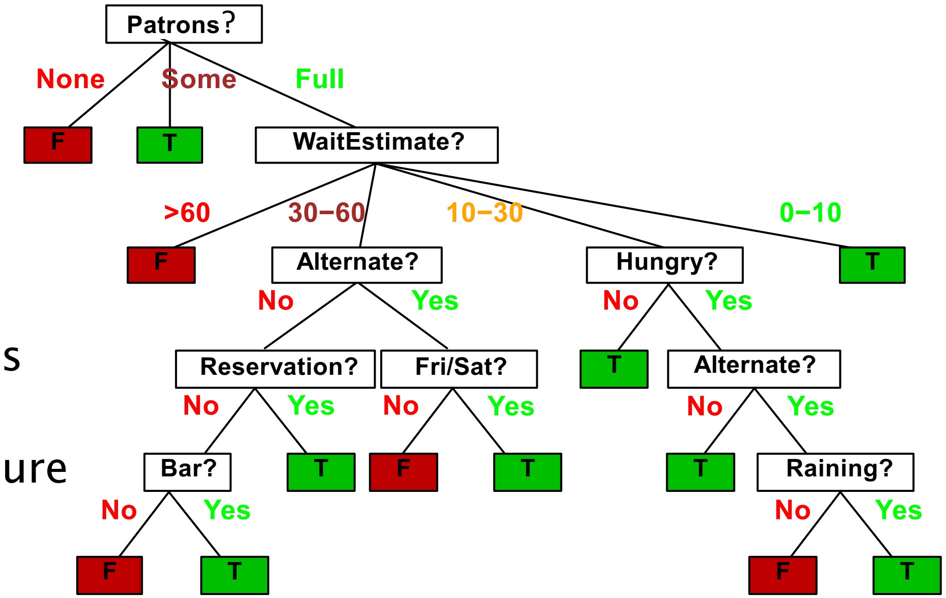
Boolean features: (T)rue or (F)alse

Modeling



1. Selecting an appropriate ML algorithm
   - black box or white box
   - …
2. Building the Model
   - via training (=automatically)
   - via configuration (=manually)

Example: To Wait or Not to Wait



Decision Tree (DT)

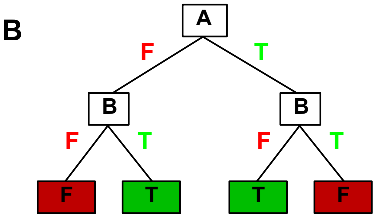
- supervised classification algorithm
  - here: binary classification (= 2 classes: true/false)
- core idea
  - tree leaves represent class labels
  - all other nodes represent a decision based on a certain feature
    - decision: feature take(s) certain value(s) or not
  - path between parent to child node represents next relevant decision
  - path from root to leaf node represents conjunctive decisions
- white box
  - DT can be visualized and is easy to understand
  - the tree can be easily transformed into logical rules
    - here: e.g., if there are no patrons in the restaurant we do not (need to) wait

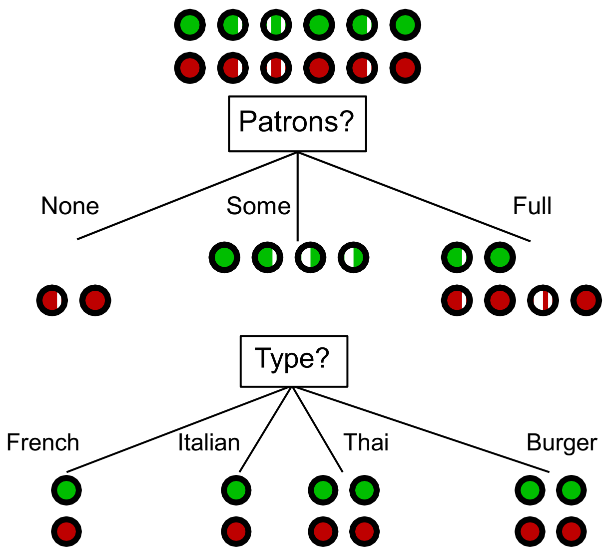
Example: To Wait or Not to Wait

DT construction

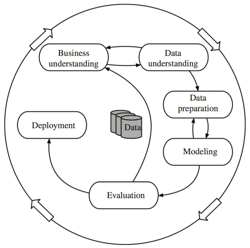
- challenge: many alternative and equivalent DT exist
- correct but inefficient DT: all samples are leaf nodes with path utilizing all present features
- instead: smallest tree possible
- core idea
  - for the next node, choose the feature that splits the sample data into the most homogeneous subsets
  - formally: the feature that minimizes the remaining entropy
- different DT algorithms around this core idea: ID3, C4.5, C5, …

A	B	A xor B
F	F	F
F	T	T
T	F	T
T	T	F





Evaluating a Model

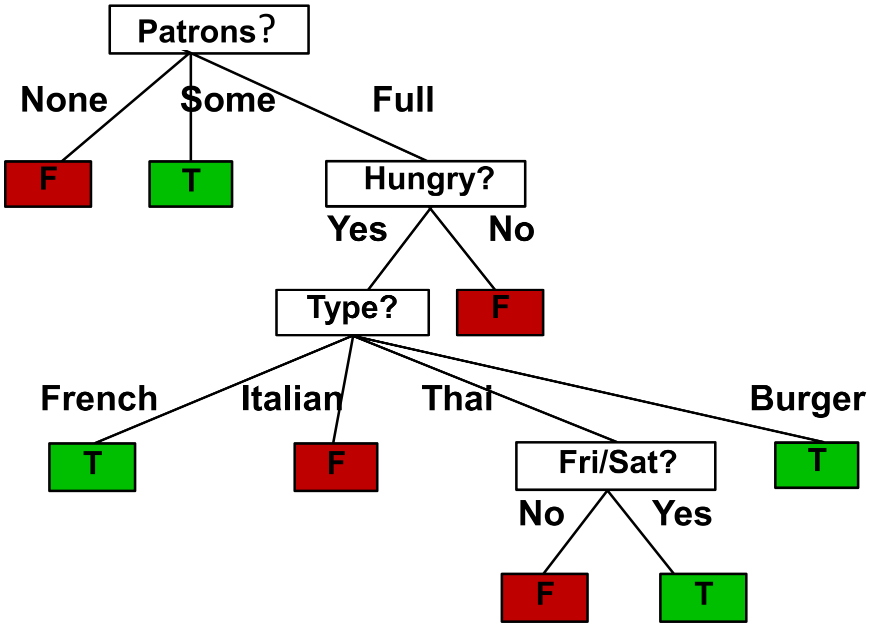


Is the model good (enough)?

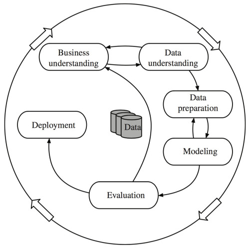
- good may relate to many criteria (metrics)
  - average quality of results
    - error rate
  - best worst case
    - worst task result is still acceptable
  - processing time and scalability
  - offers a human-understandable explanation
- in real-world scenarios (virtually) impossible to have error-free model
  - missing or noisy data
  - model tries to generalize but there may be exceptions

Example: To Wait or Not to Wait

DT fits all samples
➔ no errors ☺



Productizing a Model



Some aspects to consider

- (human and computational) resources
  - while building the model may be costly in labor and processing time/storage, executing it on new data usually is not
  - unless applied to big streaming data
  - unless costly feature generation
- updatability
  - some models may be incrementally updated
    - reinforcement learning
  - others have to be retrained
- technical and project considerations (ML Ops)
  - runtime environment
  - interface
  - versioning
- application development
  - model provides useful feature(s) of a larger application
  - human factors: trust, bias, accountability, security, …

Summary: Machine Learning Challenges

For ML to work, we need

- a defined task with a quantifiable result
    - many tasks have highly subjective and qualitative results
- large amounts of high quality data
    - data may be garbled or missing
    - dataset may be too small for the task at hand
    - dataset may not be representative or biased
- distinctive features (if needed)
    - which ones in which combination
- decisive pattern(s)
    - patterns may be inexact (or spurious)
    - most patterns are not interesting
- appropriate ML algorithm
    - which one in which configuration
    - needs to scale to the amount of data
- to accept a certain amount of errors
    - basically no model is free of errors
    - max. acceptable error rate depends on application