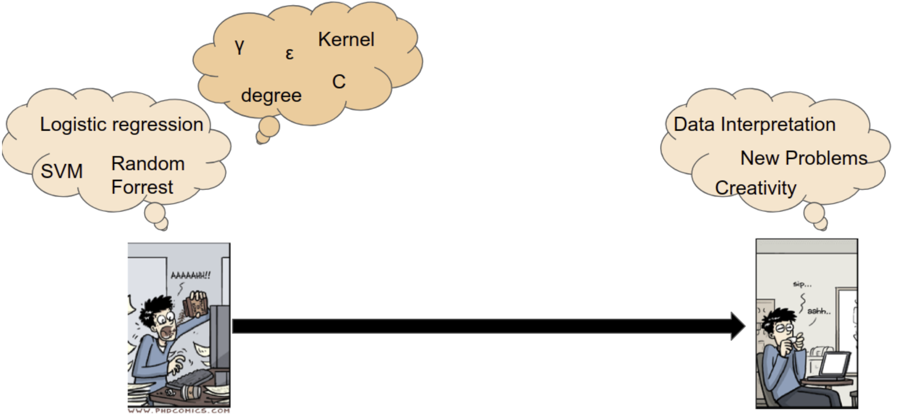
Automated Machine Learning
(AutoML)

M. Lindauer   F. Hutter

University of Freiburg

In a Nutshell



What are your expectations?

Lecture 1:
Overview and Motivation

Overview

What do we learn today?

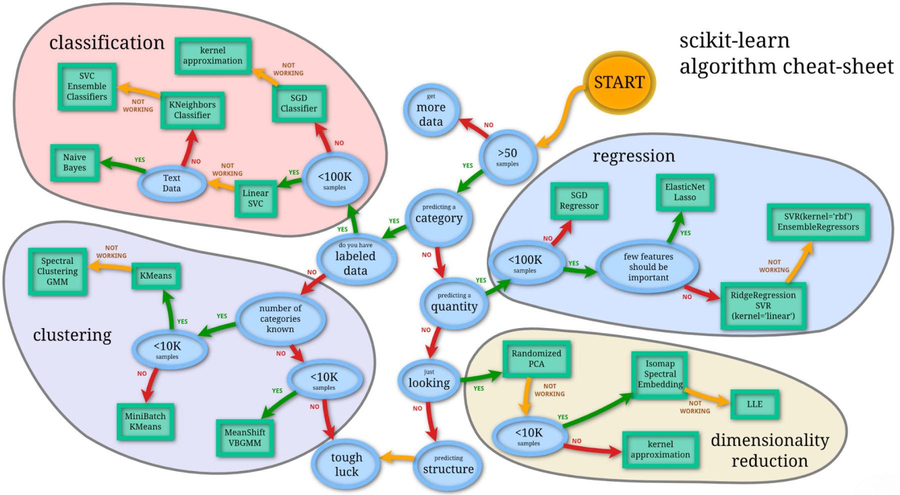
- Why ML does not scale up
- Design decisions in ML
- What is AutoML?
- Challenges in AutoML
- Risks of AutoML
- Meta-algorithmic hierarchy
- Organization of the course

Machine Learning

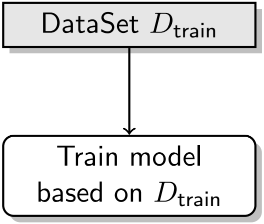
"Machine learning is the science of getting computers to act without being explicitly programmed."

by Andrew Ng

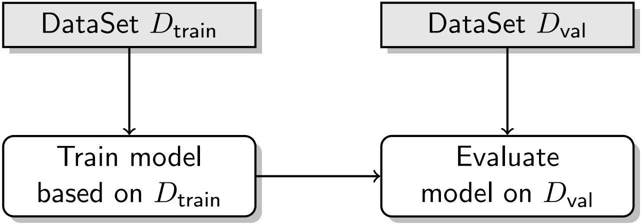
Machine Learning requires many design decisions



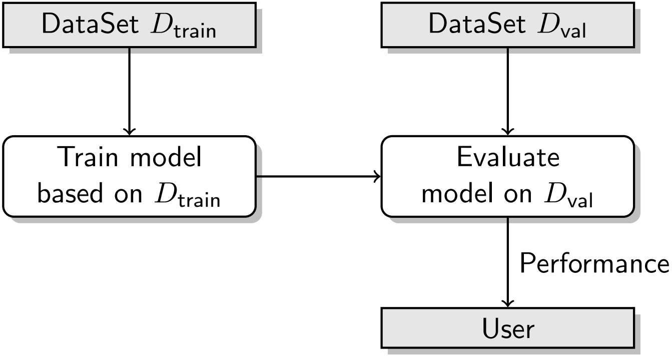
Machine Learning Workflow



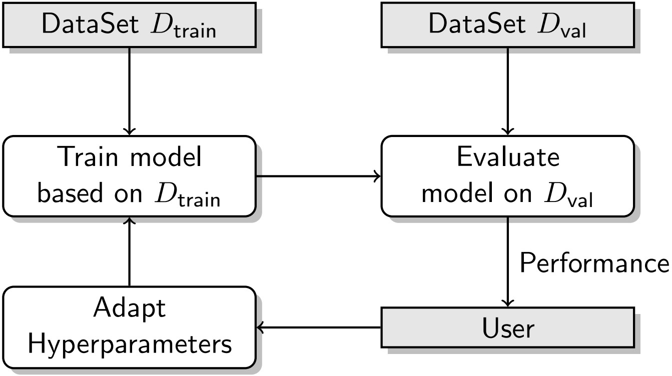
Machine Learning Workflow



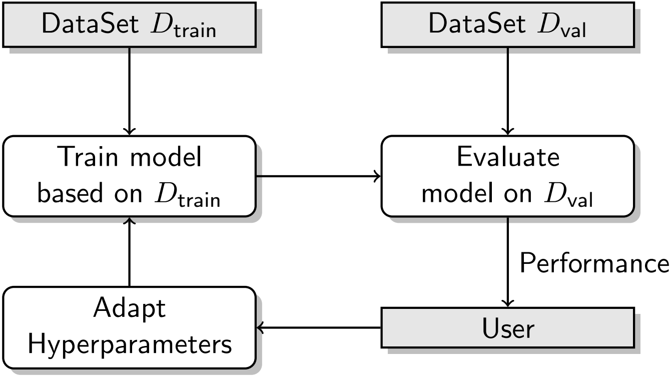
Machine Learning Workflow



Machine Learning Workflow



Machine Learning Workflow



$\rightsquigarrow$ Users indirectly teach machines how to learn.

Machine Learning does not scale up

- Basics in machine learning are not hard to grasp

Machine Learning does not scale up

- Basics in machine learning are not hard to grasp
- Achieving state-of-the-art performance is quite hard

Machine Learning does not scale up

- Basics in machine learning are not hard to grasp
- Achieving state-of-the-art performance is quite hard
- Design decisions are (sometimes) not intuitive and require a lot of expertise
  - making these design decisions is a tedious and error-prone task

Machine Learning does not scale up

- Basics in machine learning are not hard to grasp
- Achieving state-of-the-art performance is quite hard
- Design decisions are (sometimes) not intuitive and require a lot of expertise
  - making these design decisions is a tedious and error-prone task
- Many experts are employed in ML these days

Machine Learning does not scale up

- Basics in machine learning are not hard to grasp
- Achieving state-of-the-art performance is quite hard
- Design decisions are (sometimes) not intuitive and require a lot of expertise
    - making these design decisions is a tedious and error-prone task
- Many experts are employed in ML these days
- Nevertheless, developing new ML-applications takes time
- The job market for ML experts is nearly empty

"I'd like to use machine learning, but I can't invest much time"

Zoubin Ghahramani said that he often heard that

AutoML Tools Demo

Auto-Sklearn:
https://colab.research.google.com/drive/11UcQQ_dymL5spF8o56qgSRZpMC1GKag9

Auto-PyTorch:
https://colab.research.google.com/drive/14G5wvbqBkJ-SQJOdJsE_G8swq0JaFk6_

Design Decisions in Machine Learning

What could be design decisions for applying ML to a new dataset?

Design Decisions in Machine Learning

What could be design decisions for applying ML to a new dataset?

- Choice of machine learning algorithm
  - SVM, random forest, deep neural network?

Design Decisions in Machine Learning

What could be design decisions for applying ML to a new dataset?

- Choice of machine learning algorithm
  - SVM, random forest, deep neural network?
- Hyperparameters of machine learning algorithms
  - SVM: C, gamma, kernel, . . . ?

Design Decisions in Machine Learning

What could be design decisions for applying ML to a new dataset?

- Choice of machine learning algorithm
  - SVM, random forest, deep neural network?
- Hyperparameters of machine learning algorithms
  - SVM: C, gamma, kernel, ...?
- Architecture of a neural network
  - #layers, #neurons, activation function, ...

Design Decisions in Machine Learning

What could be design decisions for applying ML to a new dataset?

- Choice of machine learning algorithm
  - SVM, random forest, deep neural network?
- Hyperparameters of machine learning algorithms
  - SVM: C, gamma, kernel, ...?
- Architecture of a neural network
  - #layers, #neurons, activation function, ...
- Data preprocessing
  - data cleanup, missing data imputation, feature selection, ...

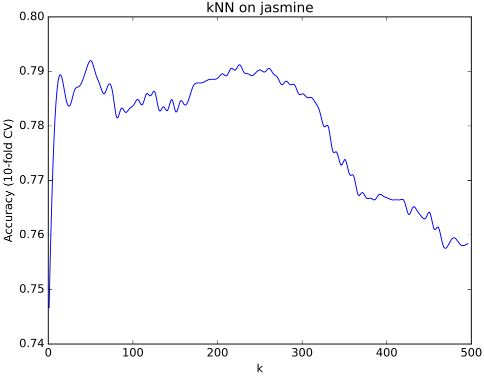
Design Decisions in Machine Learning

What could be design decisions for applying ML to a new dataset?

- Choice of machine learning algorithm
  - SVM, random forest, deep neural network?
- Hyperparameters of machine learning algorithms
  - SVM: C, gamma, kernel, ...?
- Architecture of a neural network
  - #layers, #neurons, activation function, ...
- Data preprocessing
  - data cleanup, missing data imputation, feature selection, ...
- anomaly detection
- allocation of computational resources
- ...

$\rightsquigarrow$ To achieve state-of-the-art performance,
all these design decisions have to be made for each new dataset.

A simple Example with k-NN



- k-nearest neighbors is one of the simplest ML algorithms
- Size of neighbourhood (k) is very important for its performance
- The performance function depending on k is quite complex (not at all convex)

Goal of AutoML

AutoML

The goal of AutoML is to automate all parts of machine learning (as needed) to *support* users efficiently building their machine learning applications.

Goal of AutoML

AutoML

The goal of AutoML is to automate all parts of machine learning (as needed) to *support* users efficiently building their machine learning applications.

AutoML System

Given

- a dataset
- a task (e.g., regression or classification)
- a performance metric (e.g., accuracy or RMSE)

an AutoML system automatically determines the approach that performs best for this particular application.

AutoML in Research

AutoML enables:

1. more efficient research
   - AutoML has shown on subproblems to outperform human experts

AutoML in Research

AutoML enables:

1. more efficient research
   - AutoML has shown on subproblems to outperform human experts
2. more systematic research
   - humans tend to be unsystematic which leads to errors

AutoML in Research

AutoML enables:

1. more efficient research
   - AutoML has shown on subproblems to outperform human experts
2. more systematic research
   - humans tend to be unsystematic which leads to errors
3. more reproducible research
   - since AutoML is systematic and human's unsystematic approaches cannot be reproduced

AutoML in Research

AutoML enables:

1. more efficient research
   - AutoML has shown on subproblems to outperform human experts
2. more systematic research
   - humans tend to be unsystematic which leads to errors
3. more reproducible research
   - since AutoML is systematic and human's unsystematic approaches cannot be reproduced
4. broader use of ML also in other disciplines
   - ML should not be limited to computer scientists;
   - the most amazing applications of ML are often done by either interdisciplinary teams or even non-computer scientists

Challenges in AutoML

1. Design decisions have to be made for each dataset again

Challenges in AutoML

1. Design decisions have to be made for each dataset again
2. Training of a single ML model can be quite expensive (e.g., hours, days or weeks)
   - $\rightsquigarrow$ often, we cannot try many design decisions

Challenges in AutoML

1. Design decisions have to be made for each dataset again
2. Training of a single ML model can be quite expensive (e.g., hours, days or weeks)
   - $\rightsquigarrow$ often, we cannot try many design decisions
3. the mathematical relation between design decisions and performance is (often) unknown
   - $\rightsquigarrow$ gradient-based optimization is not directly possible

Challenges in AutoML

1. Design decisions have to be made for each dataset again
2. Training of a single ML model can be quite expensive (e.g., hours, days or weeks)
   - ⇝ often, we cannot try many design decisions
3. the mathematical relation between design decisions and performance is (often) unknown
   - ⇝ gradient-based optimization is not directly possible
4. optimization in highly complex spaces
   - incl. categorical choices, continuous parameters, conditional dependencies

Risks of AutoML

What could be risks of AutoML?

Risks of AutoML

What could be risks of AutoML?

1. Users apply AutoML without understanding anything.
   - Users might wonder why (Auto-)ML does not perform well after they passed in poor data.

Risks of AutoML

What could be risks of AutoML?

1. Users apply AutoML without understanding anything.
   - Users might wonder why (Auto-)ML does not perform well after they passed in poor data.
2. Users over-trust AutoML too much.
   - humans might not use human reasoning skills and do not second guess machine decisions

Risks of AutoML

What could be risks of AutoML?

1. Users apply AutoML without understanding anything.
    - Users might wonder why (Auto-)ML does not perform well after they passed in poor data.
2. Users over-trust AutoML too much.
    - humans might not use human reasoning skills and do not second guess machine decisions
3. We enable non-ML experts to use ML without knowing the risks and consequences of ML.

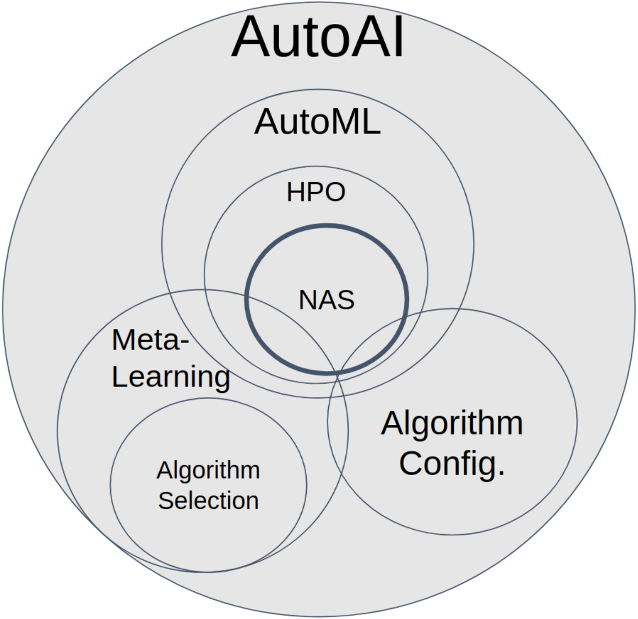
Risks of AutoML

What could be risks of AutoML?

1. Users apply AutoML without understanding anything.
   - Users might wonder why (Auto-)ML does not perform well after they passed in poor data.
2. Users over-trust AutoML too much.
   - humans might not use human reasoning skills and do not second guess machine decisions
3. We enable non-ML experts to use ML without knowing the risks and consequences of ML.
4. Could result in deployment of ...
   - inaccurate ML models due to lack of understanding of statistical concepts, e.g., sampling bias, overfitting, concept drift, ...
   - biased and unfair models due to lack of understanding ethical practices and use of features such as gender and race for predicting outcomes

See [Bond et al. 2019]

Snippet of Auto-AI Hierarchy



Goals of the Lecture

You will be able to . . .

1. use AutoML tools
2. develop AutoML tools
3. have a good overview over the state-of-the-art in AutoML
4. do research on AutoML yourself
    - perfect opportunity to do a master project or thesis with us afterwards

Course Overview

- Introduction
- Background
  - Design spaces in ML
  - Experimentation and visualization
- Hyperparameter optimization (HPO)
  - Bayesian optimization
  - Other black-box techniques
- Speeding up HPO with multi-fidelity optimization
- Pentecost (Holiday) – no lecture
- Architecture search I + II
- Meta-Learning
- Learning to learn & optimize
- Beyond AutoML: algorithm configuration and control
- Project announcement and closing

Course Format

- Concepts over details
  - we provide references and links to papers s.t. you can read up details!
- Interactive lecture
  - more efficient learning through self-reflection
- Practical exercises
  - implement it, use it and play with it!

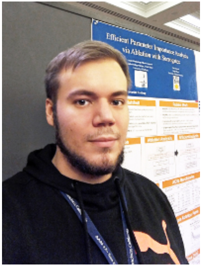
Team - Lectures

Prof. Dr. Frank Hutter

Dr. Marius Lindauer

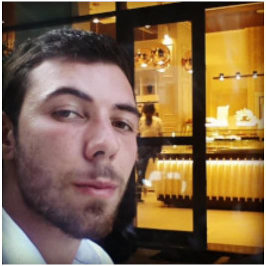
Dr. Noor Awad (guest lecturer)

Team – Exercise



André Biedenkapp

Katharina Eggensperger



Arber Zela

Organization (Lectures)

- 6 ECTS
- Every week at Monday: 14:15 (s.t) - 15:45
  (Building: 106 Room: SR 00 007)
- *Interactive* Lecture
  - We will ask you questions in the lectures
  - Kahoot quiz at the end of each lecture
- Course material on our homepage
  `ml.informatik.uni-freiburg.de/teaching/ss2019/automl/`
- Slides will be online before the lectures
- No video recording!

Organization (Exercises)

- Every Friday at: 14:15 - 15:45
  (Building: 106 Room: SR 00 007)
  - No meeting this week, but first exercise sheet!

Organization (Exercises)

- Every Friday at: 14:15 - 15:45
  (Building: 106 Room: SR 00 007)
  - No meeting this week, but first exercise sheet!
- Every week new exercise sheet and discussion of last exercise
- Most exercises will be practical, i.e., you have to implement something
- Team work mandatory, team size: 2!
- Submit solutions via bitbucket (git)

Organization (Exercises)

- Every Friday at: 14:15 - 15:45
  (Building: 106 Room: SR 00 007)
  - No meeting this week, but first exercise sheet!
- Every week new exercise sheet and discussion of last exercise
- Most exercises will be practical, i.e., you have to implement something
- Team work mandatory, team size: 2!
- Submit solutions via bitbucket (git)
- Cheating:
  - First time cheating: 0 points for exercise
  - Second time cheating: failing the course

Organization (Exercises)

- Every Friday at: 14:15 - 15:45
  (Building: 106 Room: SR 00 007)
  - No meeting this week, but first exercise sheet!
- Every week new exercise sheet and discussion of last exercise
- Most exercises will be practical, i.e., you have to implement something
- Team work mandatory, team size: 2!
- Submit solutions via bitbucket (git)
- Cheating:
  - First time cheating: 0 points for exercise
  - Second time cheating: failing the course
- You have to obtain at least 50% points in the exercises
  in order to get a grade for the course
  - Please enroll for the SL in HisInOne

Organization (Exercises)

- Every Friday at: 14:15 - 15:45
  (Building: 106 Room: SR 00 007)
  - No meeting this week, but first exercise sheet!
- Every week new exercise sheet and discussion of last exercise
- Most exercises will be practical, i.e., you have to implement something
- Team work mandatory, team size: 2!
- Submit solutions via bitbucket (git)
- Cheating:
  - First time cheating: 0 points for exercise
  - Second time cheating: failing the course
- You have to obtain at least 50% points in the exercises
  in order to get a grade for the course
  - Please enroll for the SL in HisInOne
- The number of points per sheet will slightly increase over time
- If you need help or have questions:
  1. ILIAS forum (preferred)
  2. `automl-lecture@informatik.uni-freiburg.de`
     (only for personal matters)

Requirements

- Knowledge and hands-on exp. in Machine Learning (mandatory)
  - Classification, regression, clustering, decision tree, training-test split, cross validation, pre-processing ...
  - to catch up (if nec.): `https://www.coursera.org/learn/machine-learning`

Requirements

- Knowledge and hands-on exp. in Machine Learning (mandatory)
  - Classification, regression, clustering, decision tree, training-test split, cross validation, pre-processing ...
  - to catch up (if nec.): `https://www.coursera.org/learn/machine-learning`
- Knowledge and hands-on exp. in Deep Learning (mandatory)
  - feed-forward network, recurrent network, convolutions, learning rates, regularization, ...
  - to catch up (if nec.): `https://course.fast.ai/`

Requirements

- Knowledge and hands-on exp. in Machine Learning (mandatory)
  - Classification, regression, clustering, decision tree, training-test split, cross validation, pre-processing ...
  - to catch up (if nec.): `https://www.coursera.org/learn/machine-learning`
- Knowledge and hands-on exp. in Deep Learning (mandatory)
  - feed-forward network, recurrent network, convolutions, learning rates, regularization, ...
  - to catch up (if nec.): `https://course.fast.ai/`
- Experience in Python and git (strongly recommended)
  - nearly all exercises will require that you implement something in Python and submit the solution to a git repo

Final Oral Exam

- Implement a larger project (worth $1 - 2$ weeks fulltime)
  - No teamwork!
- Exam
  - Present the project in the first 15 minutes (including some questions from us)
  - Answer questions about further course material in the second 15 minutes
- tentative date: end of September

Additional Resources

- To get a deep understanding of AutoML, you should also read some papers
- We will provide links to papers at the end of each lecture
- New AutoML book: `https://www.automl.org/book/`
  - Draft available online
- NeurIPS tutorial on NAS and meta-learning: `https://videoken.com/embed/5A4xbv5nd8c`

Opportunities and Risks

AutoML is an advanced lecture and we update it each time.

Opportunities and Risks

AutoML is an advanced lecture and we update it each time.

Opportunities:

- All presented topics are close to state-of-the-art; there is active research on these topics
- The course will provide a solid background for doing a master project/thesis in our group

Risks:

- You will find some typos and issues in the slides; please tell us if you find something

Opportunities and Risks

AutoML is an advanced lecture and we update it each time.

Opportunities:

- All presented topics are close to state-of-the-art; there is active research on these topics
- The course will provide a solid background for doing a master project/thesis in our group

Risks:

- You will find some typos and issues in the slides; please tell us if you find something

→ Give us some feedback and we will improve the course!

Opportunities and Risks

AutoML is an advanced lecture and we update it each time.

Opportunities:

- All presented topics are close to state-of-the-art;
  there is active research on these topics
- The course will provide a solid background
  for doing a master project/thesis in our group

Risks:

- You will find some typos and issues in the slides;
  please tell us if you find something

→ Give us some feedback and we will improve the course!

Note: AutoML was already partially covered in our old lecture ML4AAD.
If you successfully attended ML4AAD, please don't attend AutoML.

Introduce yourself!

- Why have you chosen this course?
- Background knowledge? (ML, DL, ...)
- Experience with such problems?
- Are you still looking for a team member?