metalearning and autoML

Machine Learning
(Aprendizagem Computacional)
Carlos Soares
csoares@fe.up.pt

[some slides shamelessly stolen from J. Vanschoren]

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

plan

FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

- the world where automated ml lives
  - a world of **many models**
  - needs **model management**
  - **metalearning/automl** can help
  - but **opportunities and challenges** are still open

the world where automated ml lives

lots of data

+

lots of detail

+

lots of problems

+

lots of models

=

extreme data mining

(adapted from Soulié-Fogelman)

+ lots of models

U.PORTO
FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

- more specific knowledge
- ... that is, models for smaller subsets
  - e.g. [Fogelman 06]
    - “broadband communications company moved from 5 cross-sell models per year to 1600;
    - A wireless communications company that produces 700 CRM models per year;”
- ... eventually, individual entities
  - e.g. a recommendation model for each customer
  - e.g. soft sensors
  - e.g. UPV’s project with large retail company
    - 50 million models to predict the sales of products

todo

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

- the world where automated ml lives
  - a world of **many models**
  - needs **model management**
  - **metalearning/automl** can help
  - but **opportunities and challenges** are still open

extreme data mining

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

lots of data

+

lots of detail

+

lots of problems

+

lots of models

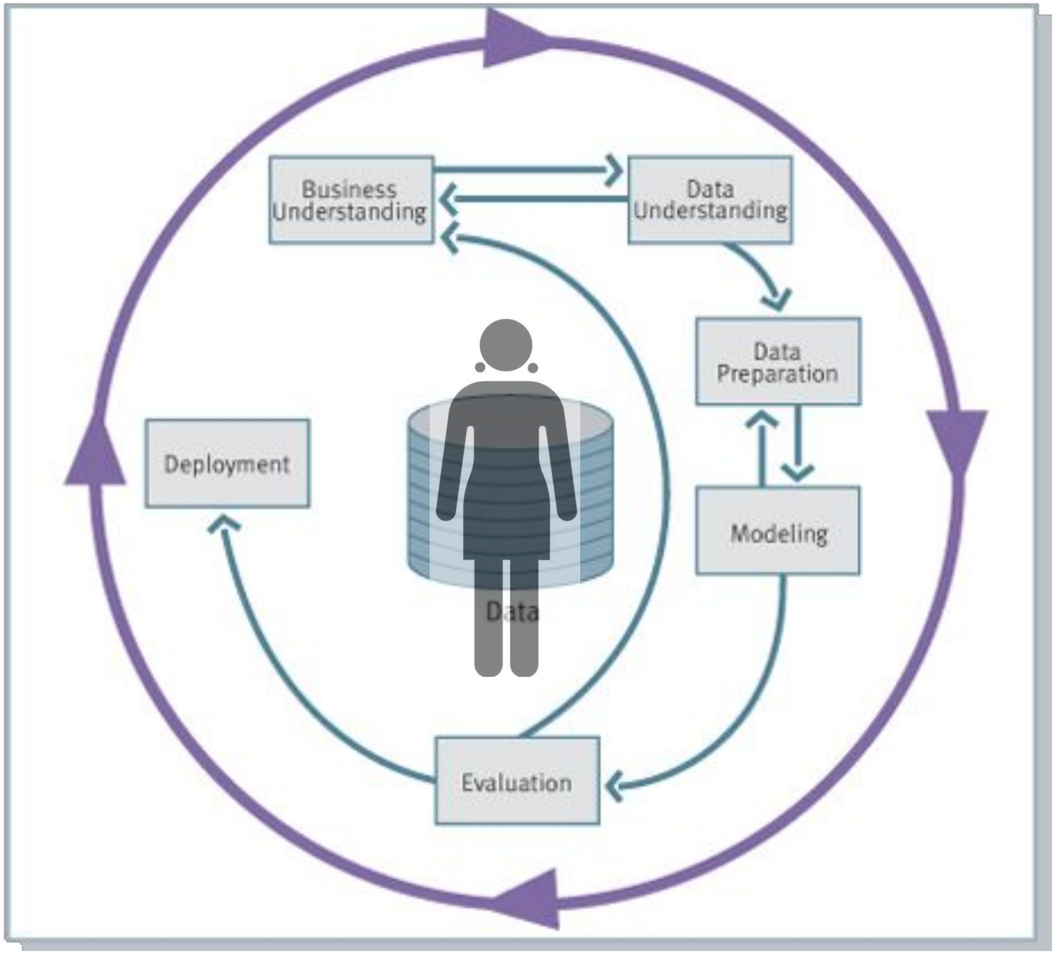
=

extreme data mining

lots of data mining

(adapted from Soulié-Fogelman)

traditional DM methodology

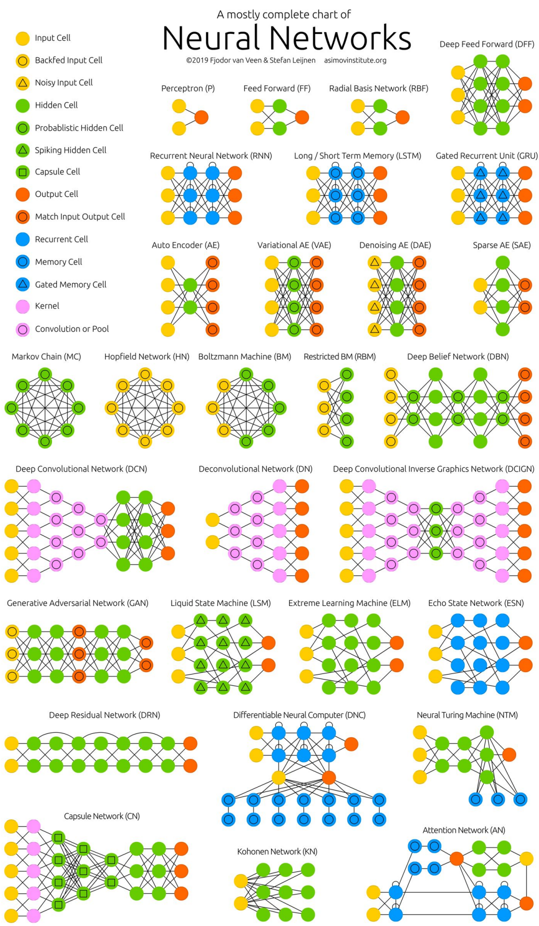


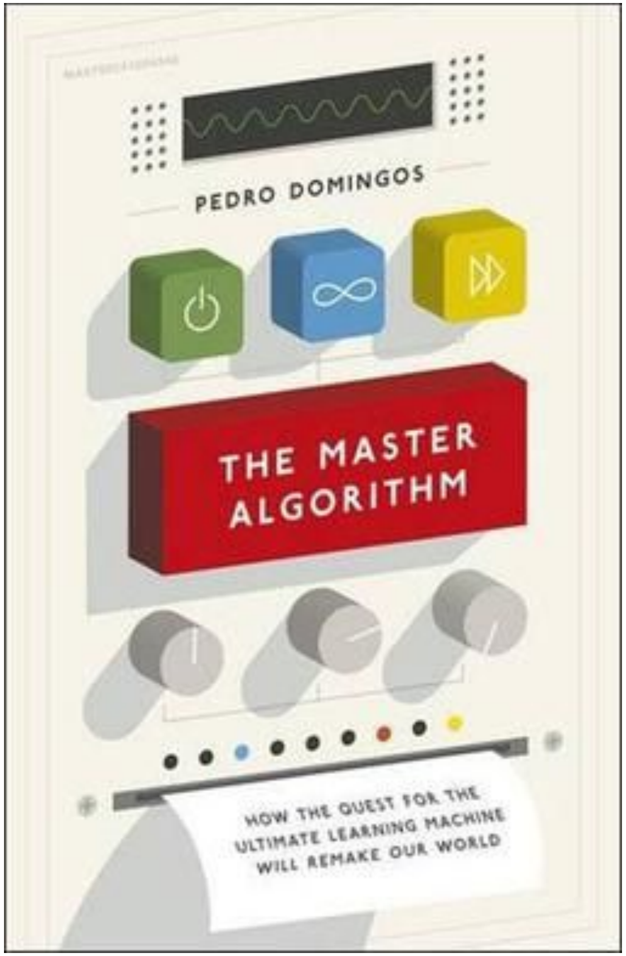
source: Chapman P, Clinton J, Kerber R, et al.
CRISP-DM 1.0: Step-by-Step Data Mining Guide. 2000.

this is not possible!

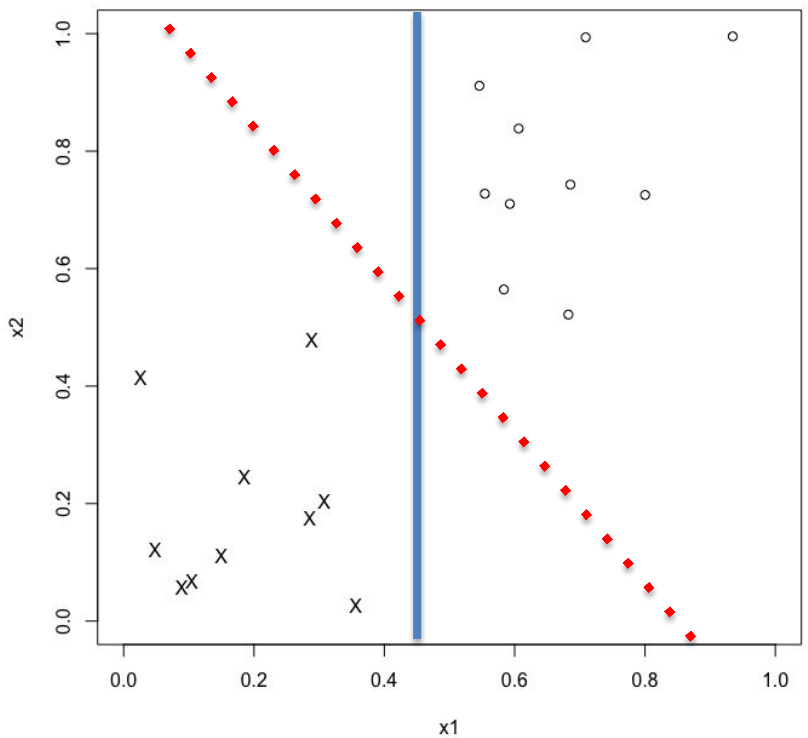
wait, is this really a problem?

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO





bias: can't live with it



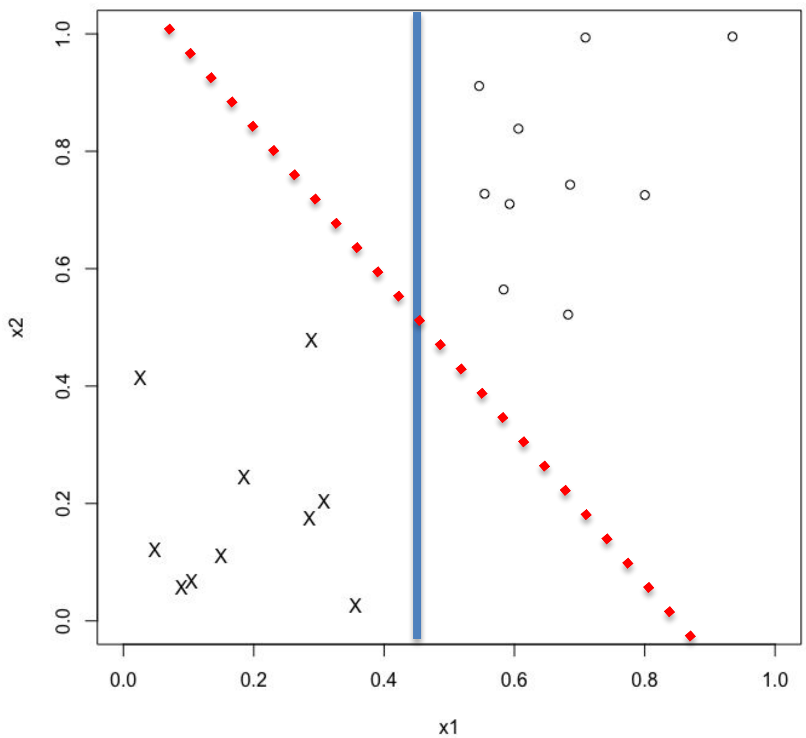
- given
  - dataset
  - learning algorithm
- not every model is possible
  - e.g. DT and LR
  - … but not DT and LR

… and can't live without it

- bias-free learning is futile (Mitchell 97, Ch. 2)
  - an algorithm that assumes nothing concerning the function it is trying to learn has no rational basis to classify unknown cases

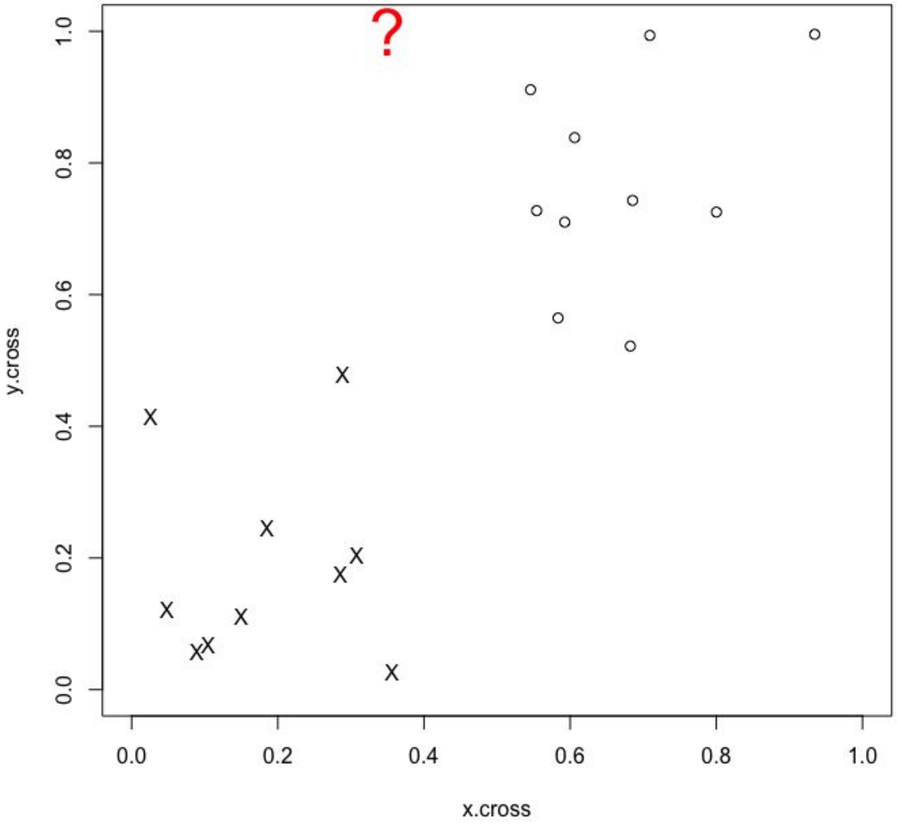
- bias = criteria to prefer one model relative to another
- … so, how to select the best model if all models are considered equally suitable?

the bias-free algorithm



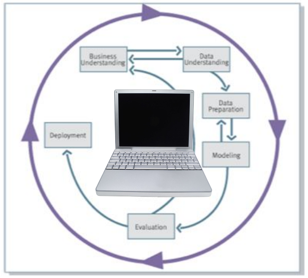
- … can learn any model
  - e.g. the DT and the LR
- … but doesn't have any preference for one over the other
  - … or for one over the other

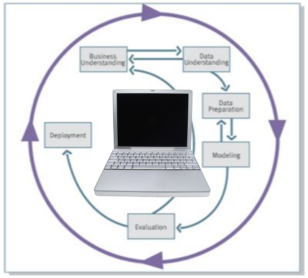
… right?

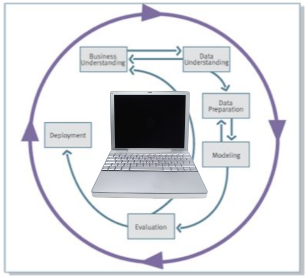


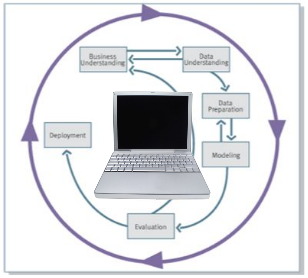
so, if this is not possible?...

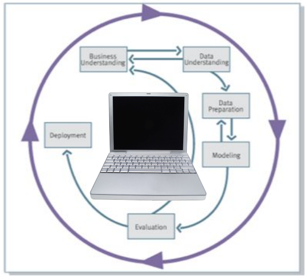
[the dream]

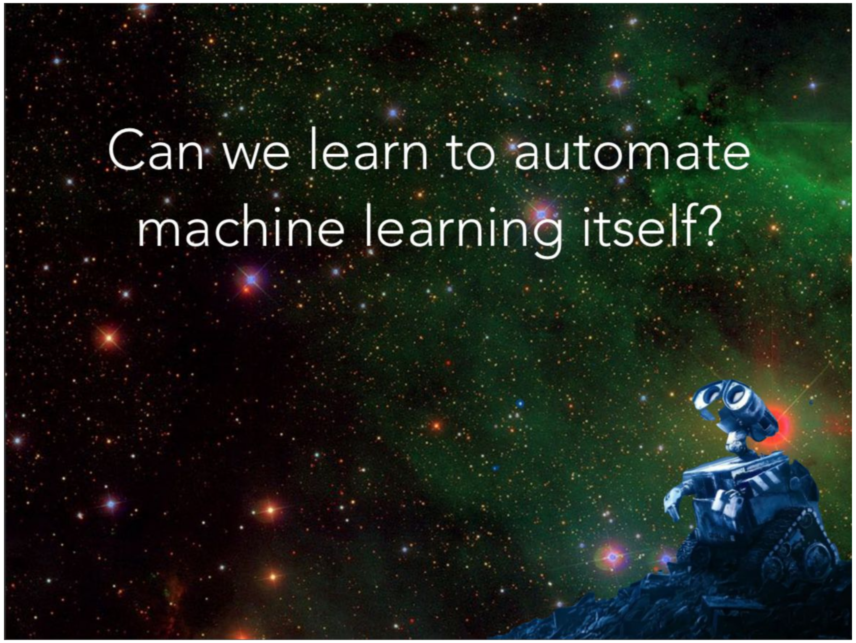








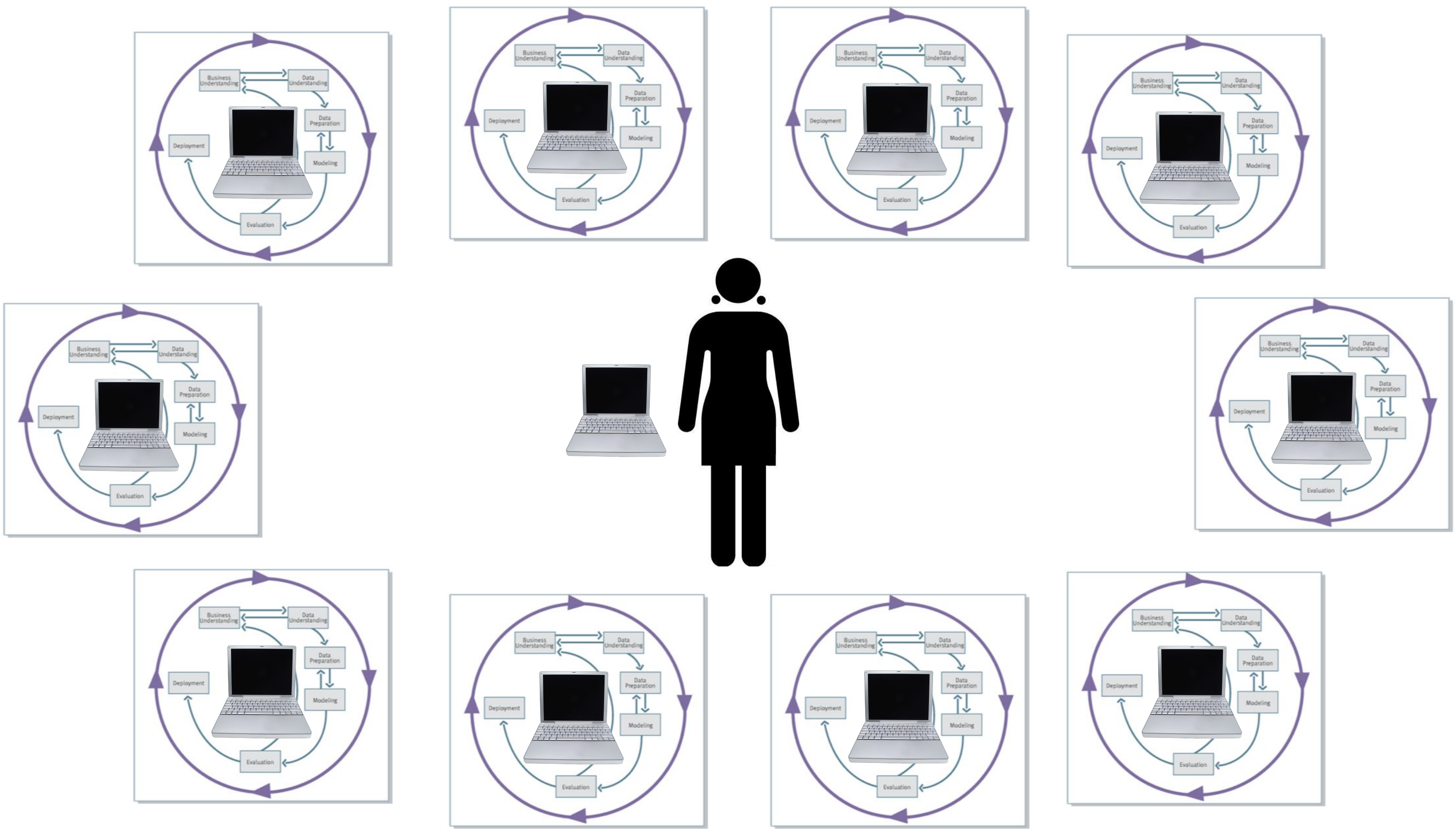




shameless plagiarism of someone who prepares more beautiful slides than I do

… but maybe this is…

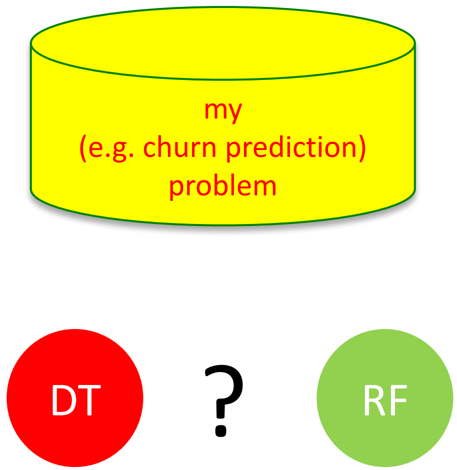
U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO



todo

- the world where automated ml lives
  - a world of **many models**
  - needs **model management**
  - **metalearning/automl** can help
  - but **opportunities and challenges** are still open

a simple autoML problem: algorithm selection



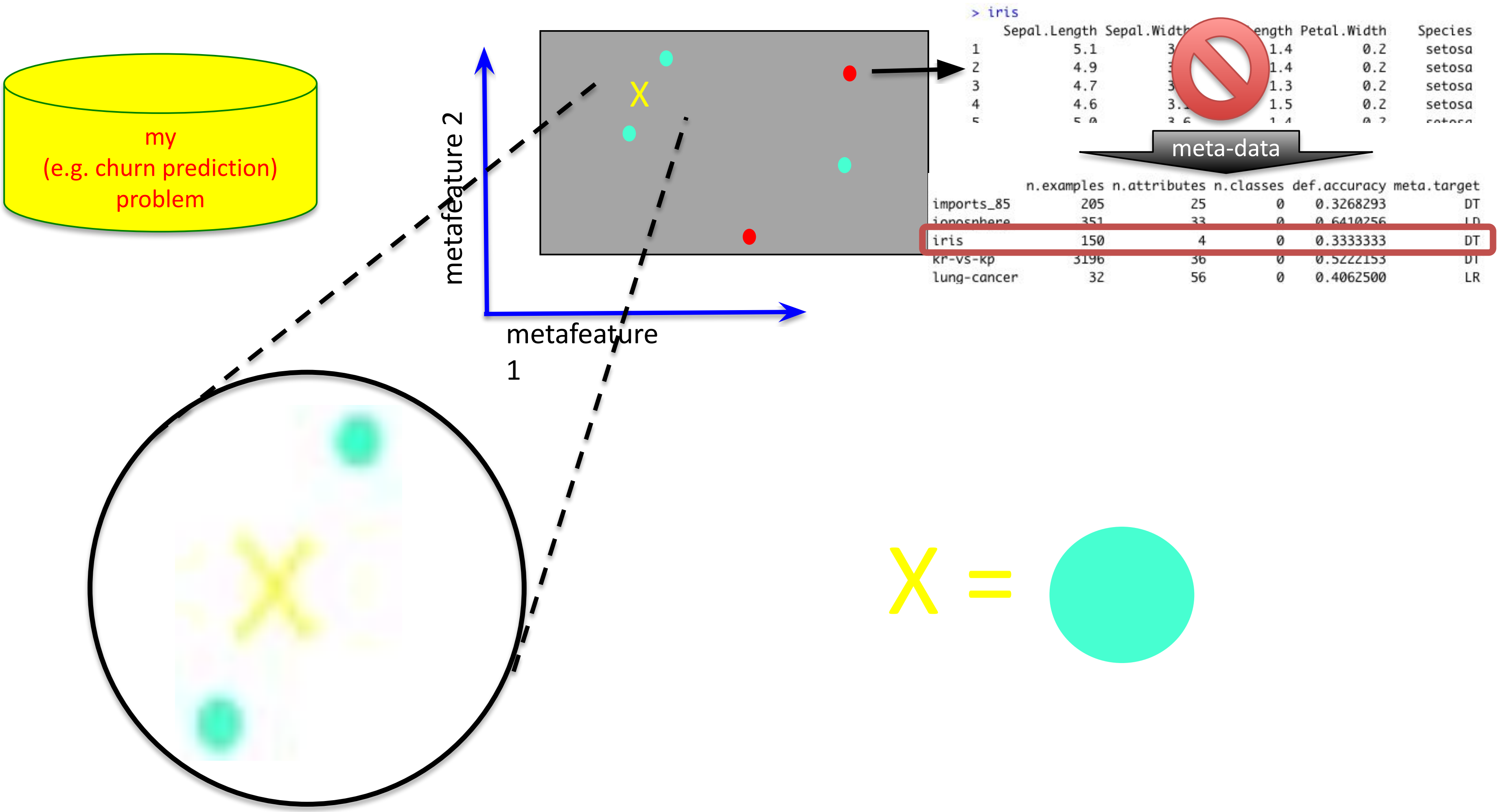
how can I use previous experience to help me choose the best algorithm?

```
> iris
  Sepal.Length Sepal.Width Petal.Length Petal.Width    Species
1          5.1         3.5          1.4         0.2     setosa
2          4.9         3.0          1.4         0.2     setosa
3          4.7         3.2          1.3         0.2     setosa
4          4.6         3.1          1.5         0.2     setosa
5          5.0         3.6          1.4         0.2     setosa
```


```
> house_votes_84
    V1   V2 V3   V4   V5   V6   V7   V8   V9 V10  V11  V12  V13  V14  V15  V16        V17
1    n    y  n    y    y    y    n    n    n   y <NA>    y    y    y    n    y republican
2    n    y  n    y    y    y    n    n    n   n    n    y    y    y    n <NA> republican
3 <NA>    y  y <NA>    y    y    n    n    n   n    y    n    y    y    n    n   democrat
4    n    y  y    n <NA>    y    n    n    n   n    y    n    y    n    n    y   democrat
5    y    y  y    n    y    y    n    n    n   n    y <NA>    y    y    y    y   democrat
```

autoML approaches (1/2): metalearning for algorithm selection

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO



Brazdil P, Soares C, Costa J. Ranking Learning Algorithms: Using {IBL} and Meta-Learning on Accuracy and Time Results. Machine Learning. 2003;50(3):251-277.

experimental results: an example

U.PORTO FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

- when to prune decisions trees
  - 3 classes: prune, don’t prune, doesn’t matter
  - selected **64** datasets from UCI
- metafeatures
  - entropy of classes (target attribute)
  - mean entropy of symbolic attributes
- positive but not excellent
  - very simple example
  - better examples in different contexts

algorithm	accuracy (%)
default	41
dt	41
ld	41
rf	47
svm	41
nn	45

very hard problem

summary

- metalearning for algorithm selection
  - induce model from *metadata* to predict the best algorithm on a new dataset

(meta)data

i	$x_{i,1}$	$x_{i,2}$	$x_{i,3}$	*decision*
1	0.7	327.2	0	A
2	-0.6	1234.2	1	B
3	...	...	...	...

-0.8	37.2	1	?
0.2	14.32	1	?
...	...	...	...

predict

autoML is old

U.PORTO FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

THE ALGORITHM SELECTION PROBLEM

John R. Rice
Computer Science Department
Purdue University
West Lafayette, Indiana 47907

July 1975

CSD-TR 152

(This is a revised version of CSD-TR 116, 117 and 130)

(To appear in Advances in Computers, Vol. 15, Academic Press, 1976)

autoML is hard

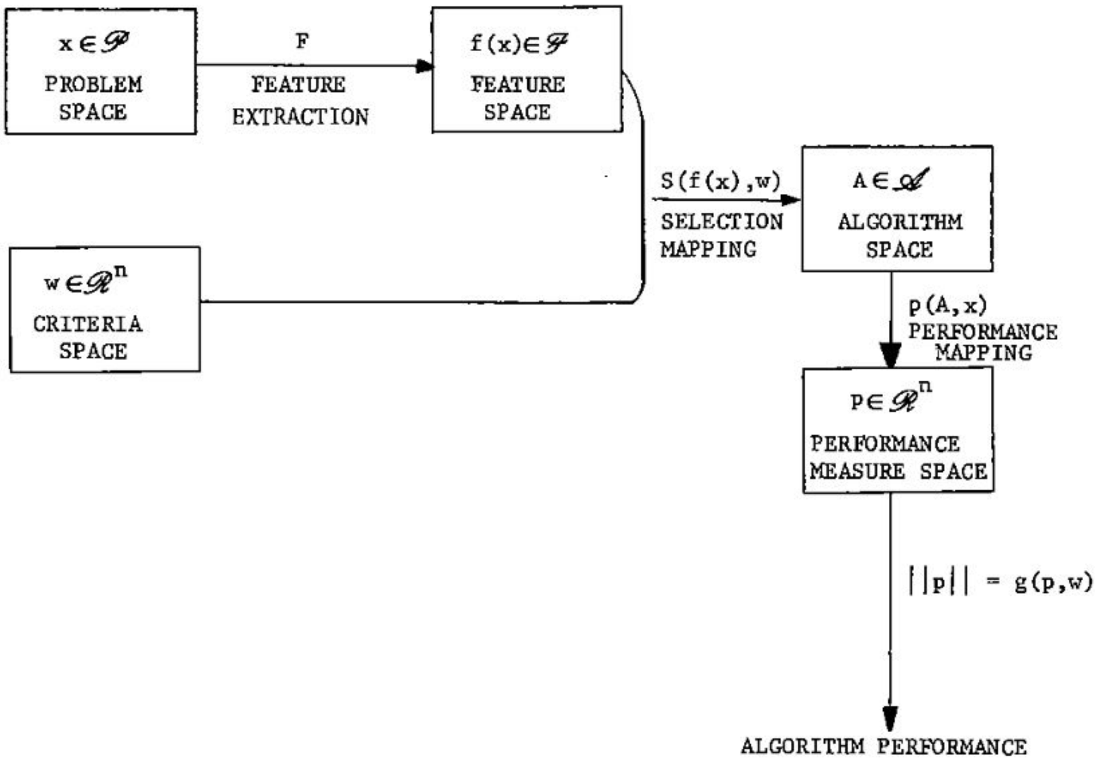
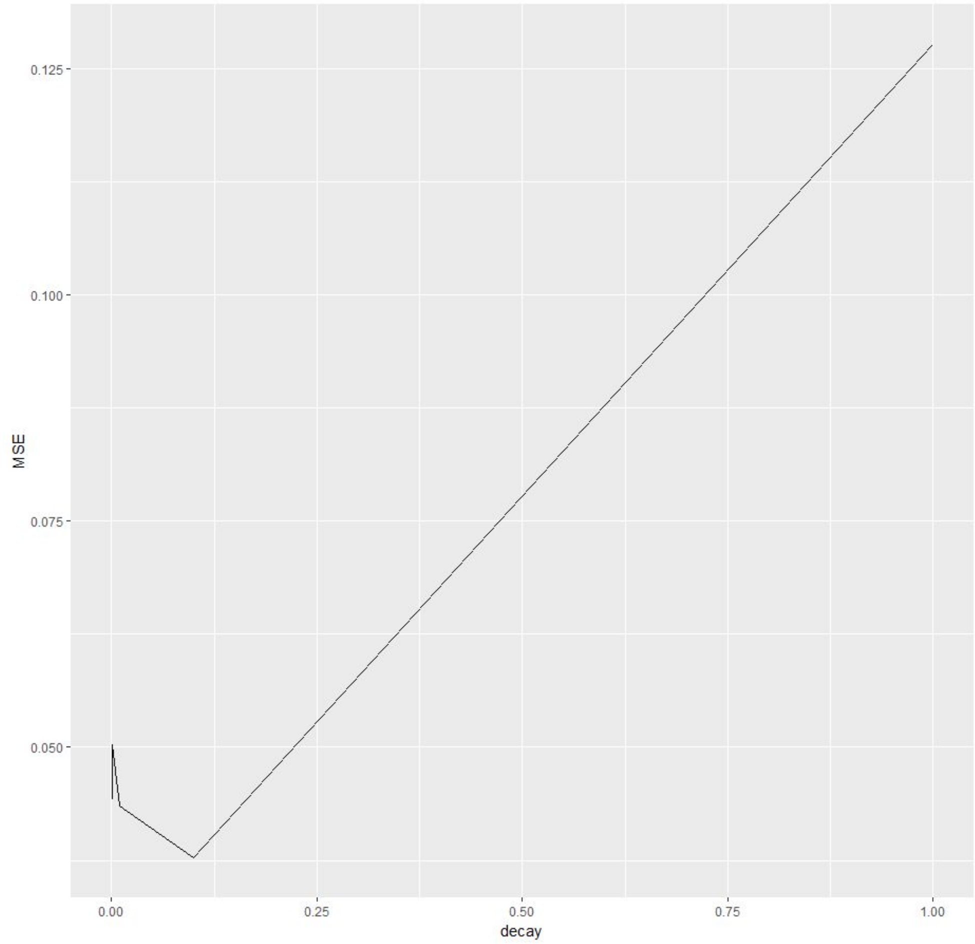


Figure 4. Schematic diagram of the model with selection based on problem features and variable performance criteria.

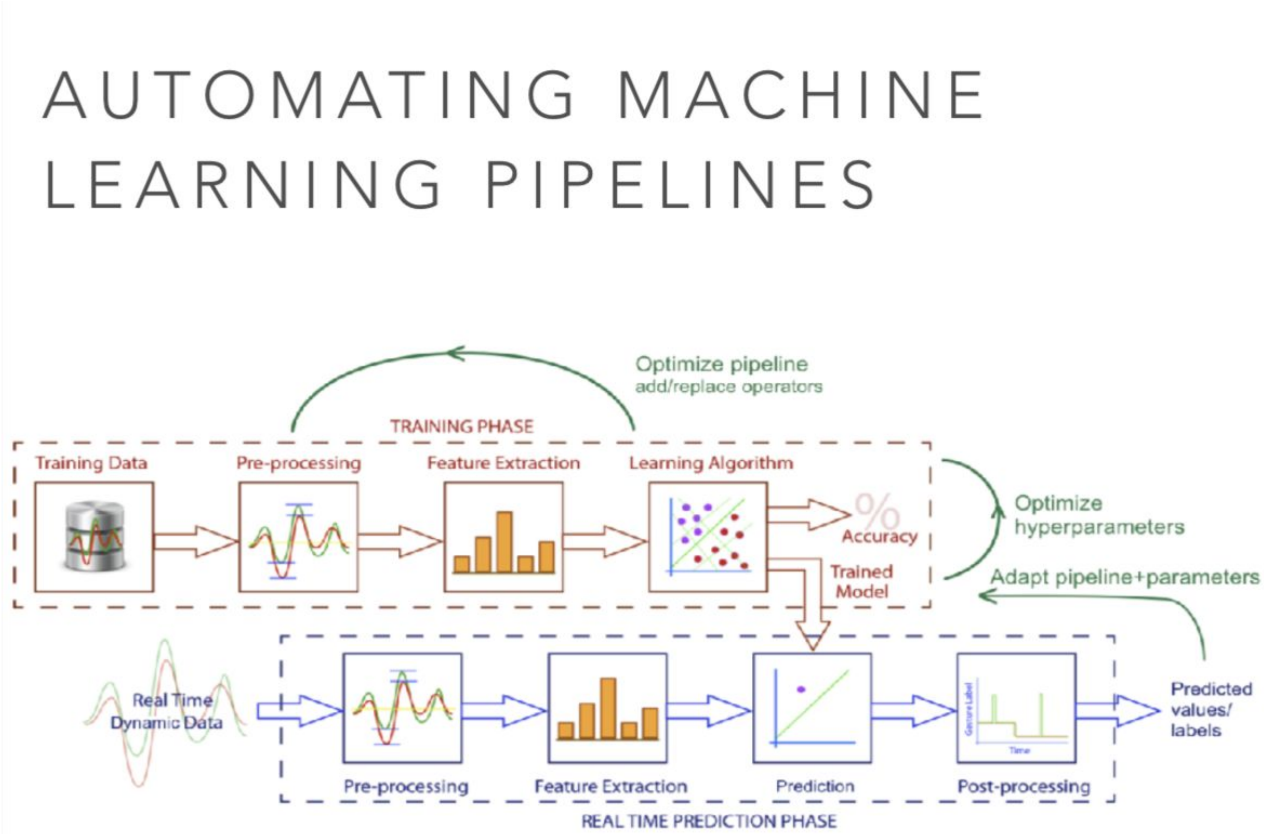
Rice, J. R. (1976). The Algorithm Selection Problem. In *Advances in Computers* (Vol. 15, pp. 65–118). https://doi.org/10.1016/S0065-2458(08)60520-3

autoML is very hard



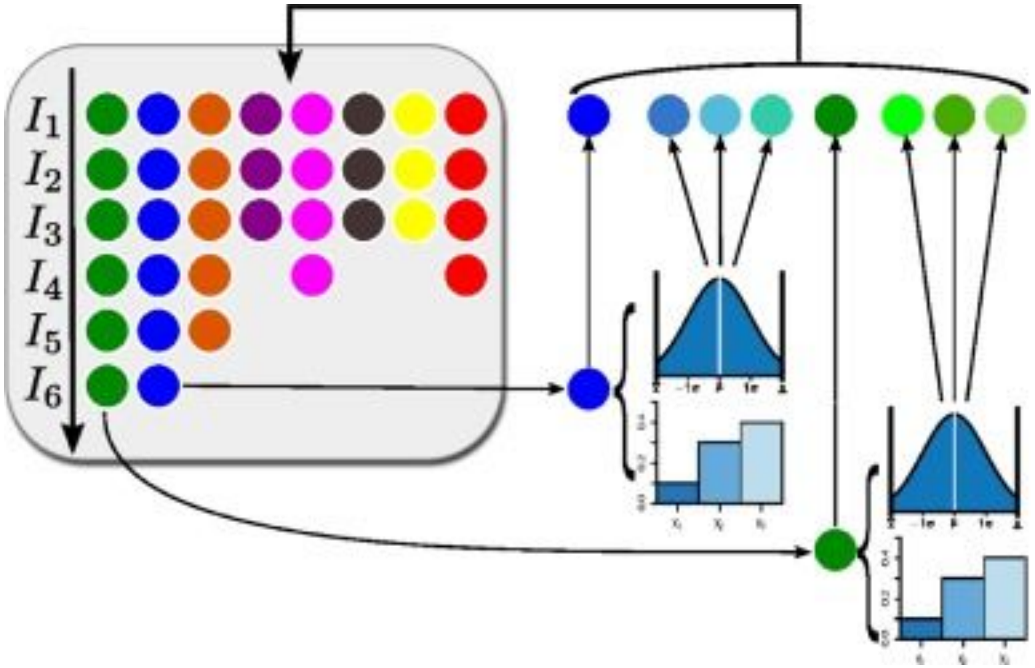
thanks Catarina Félix!

autoML is extremely hard!



can’t resist it: his slides look so much better than mine!

trendy autoML approaches: search and metalearning

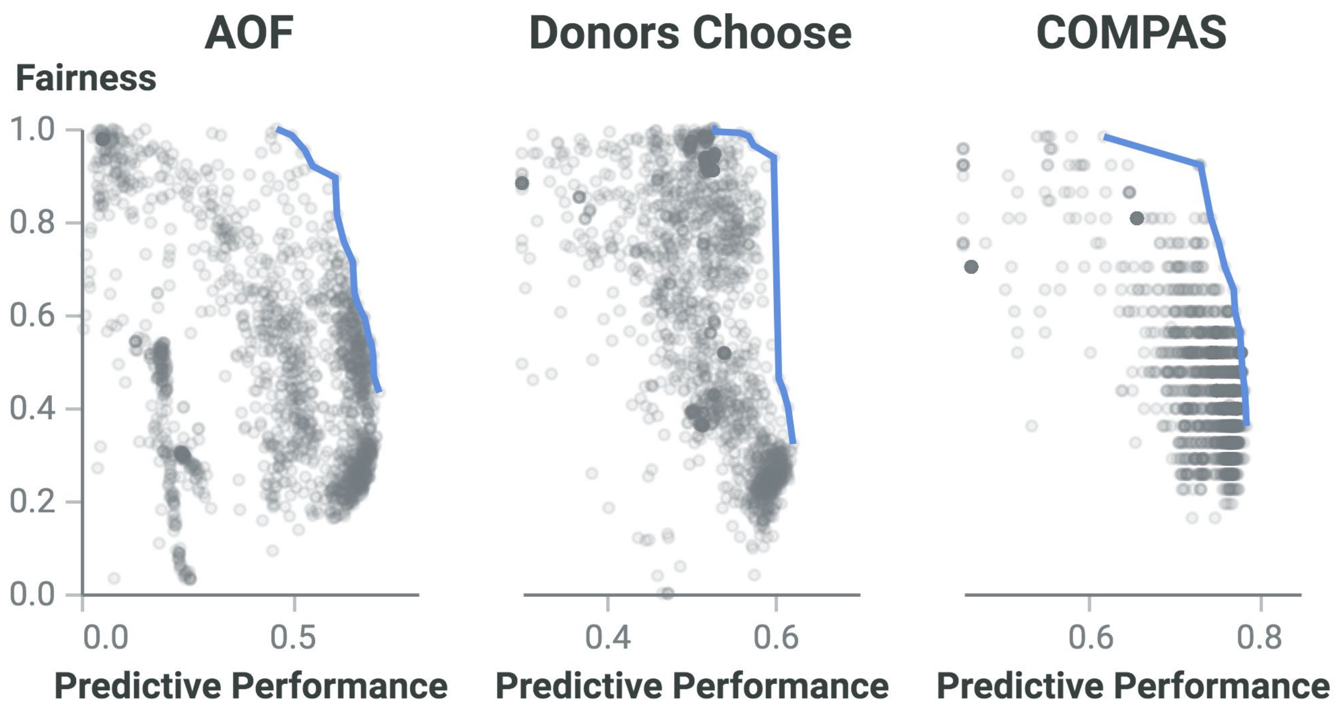


M. López-Ibáñez, J. Dubois-Lacoste, L. P. Cáceres, M. Birattari, T. Stützle, The irace package: Iterated racing for automatic algorithm configuration, Operations Research Perspectives, Volume 3, 2016, Pages 43-58, ISSN 2214-7160, https://doi.org/10.1016/j.orp.2016.09.002.

and it's not only about predictive performance!

U.PORTO
FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

- responsible AI
  - is *accuracy vs fairness* a real problem?
    - i.e. if I want to promote fair models, I have to sacrifice predictive performance



A. Cruz, P. Saleiro, C. Belém, C. Soares, and P. Bizarro (2021), "Promoting Fairness through Hyperparameter Optimization", IEEE International Conference on Data Mining

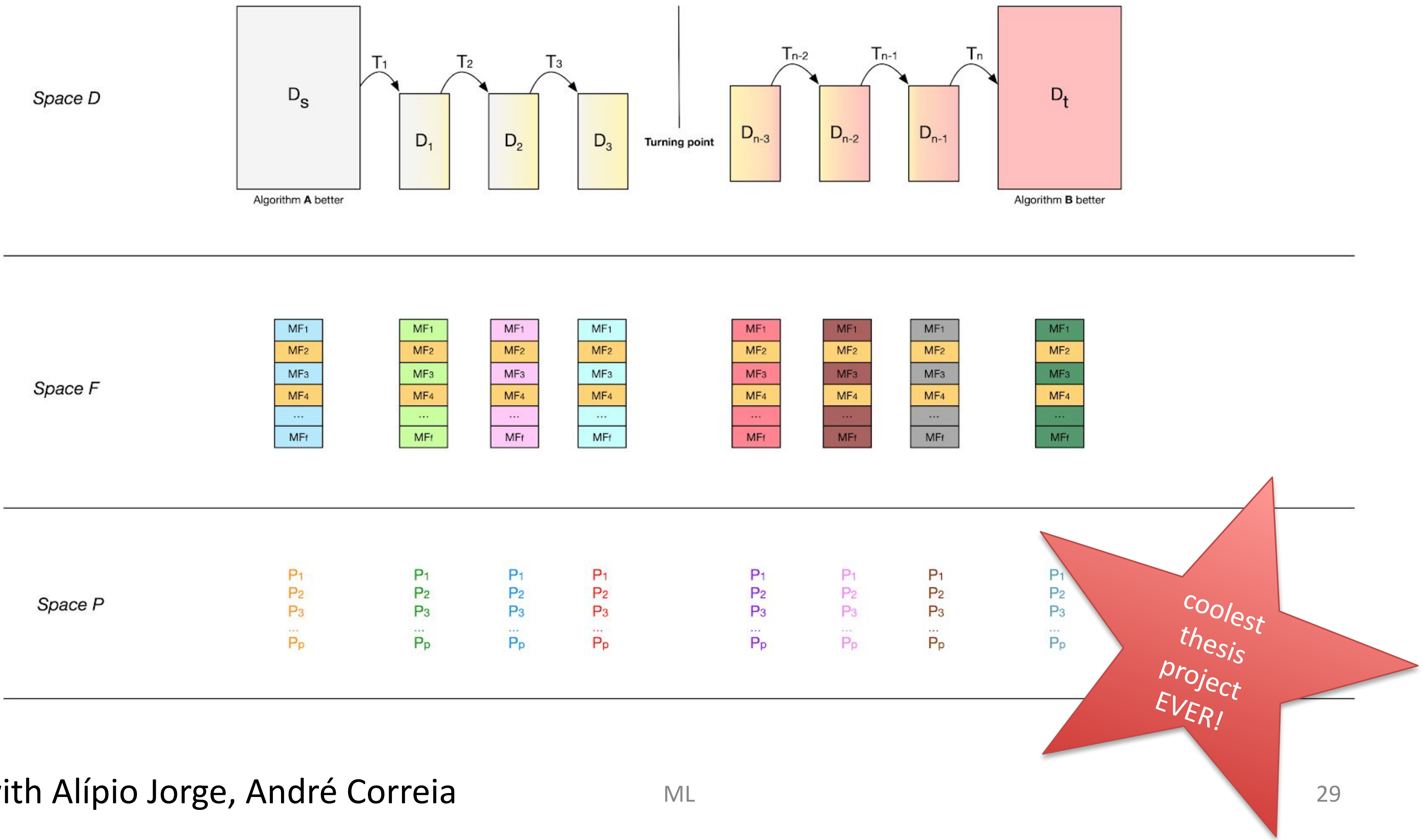
todo

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

- the world where automated ml lives
  - a world of **many models**
  - needs **model management**
  - **metalearning/automl** can help
  - but **opportunities and challenges** are still open

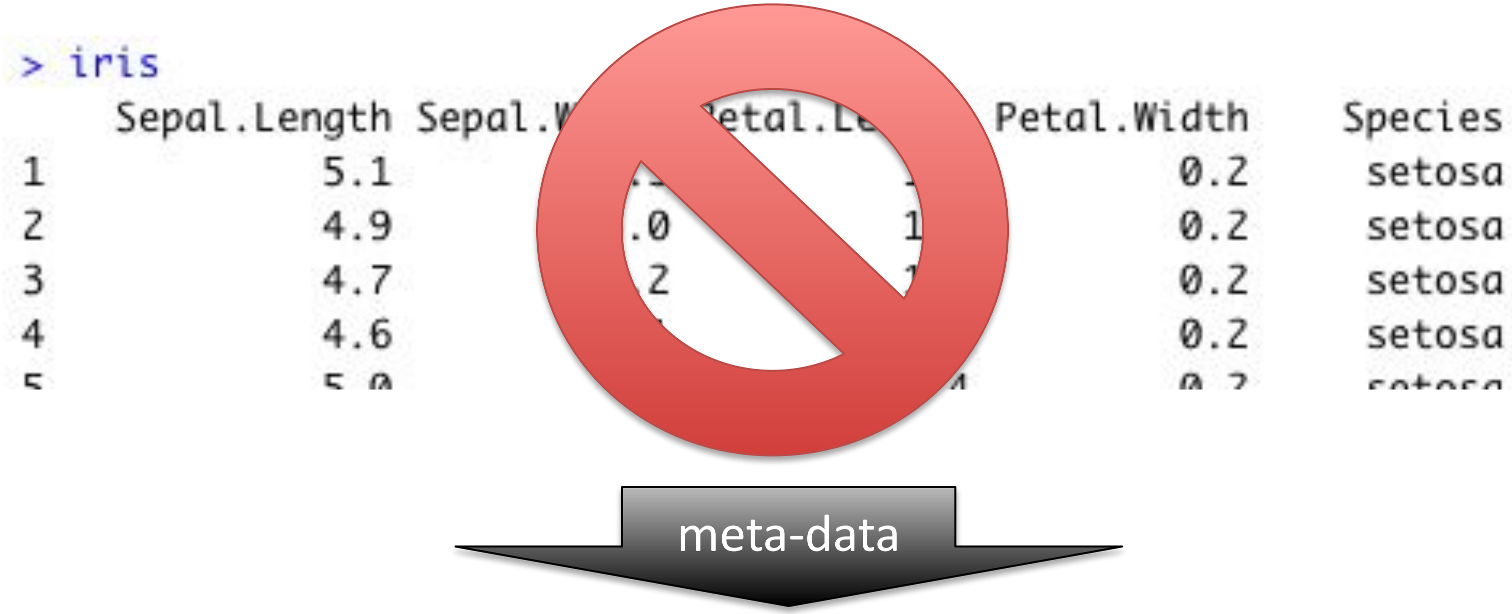
coolest metalearning ever (1/2): dataset morphing to understand ML algorithm behavior

U.PORTO FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO



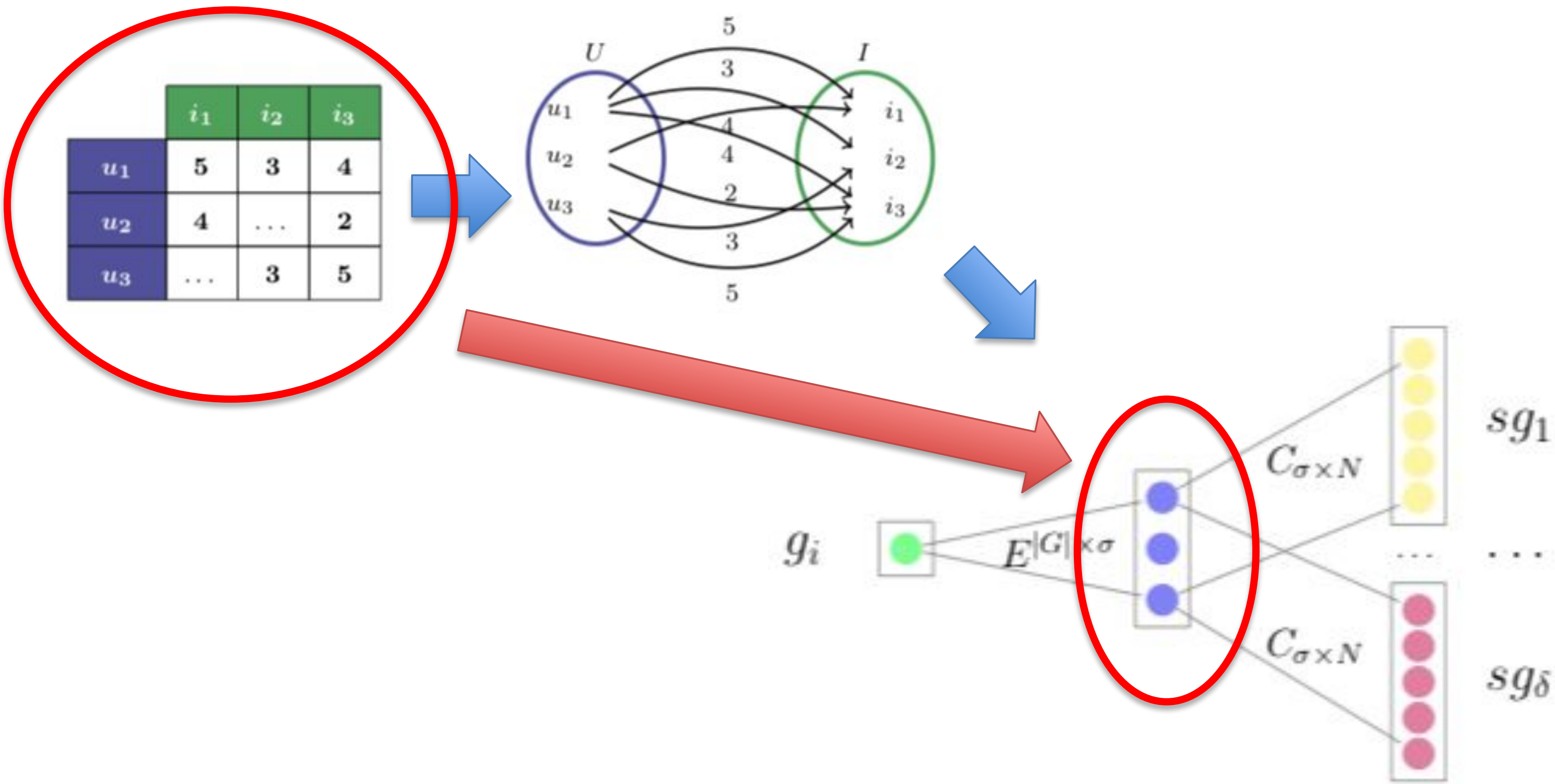
with Alípio Jorge, André Correia

coolest metalearning ever (2/2): dataset characterisation



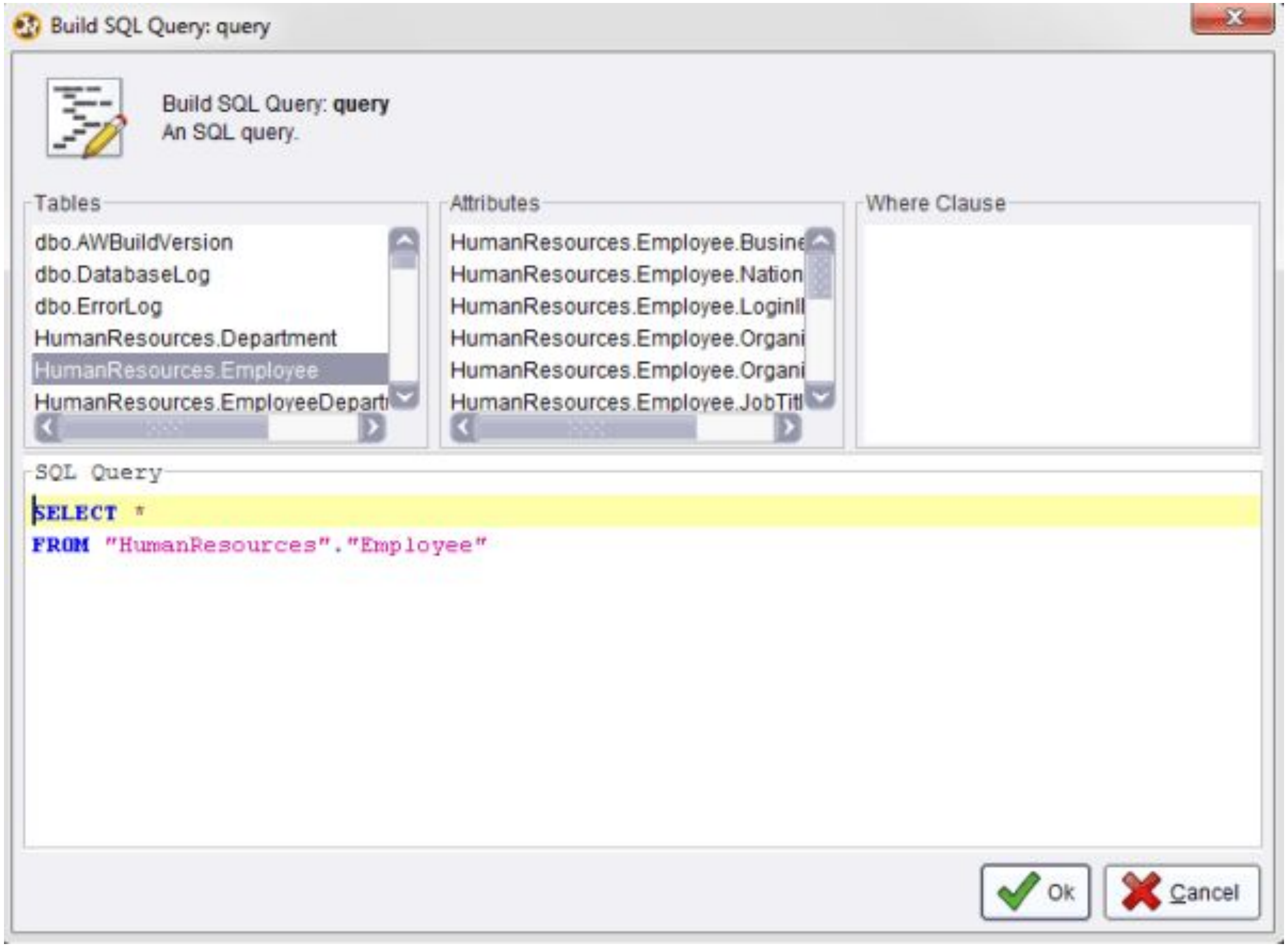
	n.examples	n.attributes	n.classes	def.accuracy	meta.target
imports_85	205	25	0	0.3268293	DT
ionosphere	351	33	0	0.6410256	LD
iris	150	4	0	0.3333333	DT
kr-vs-kp	3196	36	0	0.5222153	DT
lung-cancer	32	56	0	0.4062500	LR

coolest metalearning ever (2/2): dataset embeddings



Cunha, T., Soares, C., & Carvalho, A.C. (2018). cf2vec: Collaborative Filtering algorithm selection using graph distributed representations. https://arxiv.org/abs/1809.06120v1

but what is fundamentally wrong?



https://datamodelprototype.wordpress.com/2013/10/25/
how-to-import-and-export-files-in-rapidminer/

how to make it fundamentally right?

U.PORTO FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

```
CREATE TABLE LoanRequest (
     ID INT PRIMARY KEY,
    CustomerID INT REFERENCES Customer(ID) NOT
NULL,
    Value INT,
    Default ENUM('y','n') TARGET,
);

SELECT ID, PREDICT(Default) FROM LoanRequest;
```

the way to a (real) empirical science of machine learning

U.PORTO FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

- metalearning to understand the behavior of algorithms
  - eg when to do resampling in unbalanced datasets

Coverage	LWNorm ($\times 10^{-2}$)	Ranking	Conditions
No preproc.			
21	3.3992	j>c>d>e>bf>i>a>h>g	*statistical_kurtosis* >=17.9168
21	2.8934	j>b>d>e>f>c>i>a>g>h	*statistical_cov* <= 0.0234
21	2.7911	j>b>d>c>f>e>a>i>g>h	*statistical_eigenvalues* <= 0.2581
21	2.7911	j>b>d>c>f>e>a>i>g>h	*statistical_var* <= 0.2581
42	2.3839	j>c>d>ef>b>i>h>a>g	*general_nr_inst* >= 376.0
21	2.8448	j>c>e>d>f>b>h>i>g>a	*complexity_n1* <= 0.0675
41	2.8327	j>c>e>d>b>f>h>i>g>a	*complexity_l2* <= 0.0421
24	2.6046	j>c>d>e>f>b>i>h>a>g	*complexity_t3* <= 0.0031
41	2.2675	j>c>e>d>b>h>f>gi>a	*complexity_n4* <= 0.0611
21	2.4585	j>c>d>e>f>b>h>i>g>a	*typology_border* <= 0.0858
41	2.2161	j>c>d>b>e>f>h>i>g>a	*typology_safe* >= 0.5334
41	2.9260	j>c>e>d>b>f>h>i>g>a	*landmarking_linear_discr* >= 0.9225
21	2.7286	j>e>c>d>b>f>h>i>g>a	*landmarking_nn* >= 0.9750
41	2.7001	j>c>e>d>b>f>h>i>a>g	*landmarking_nn* >= 0.9052
Do preproc.			
21	3.0865	h>c>d>f>ae>i>g>b>j	*statistical_kurtosis* <= -1.3063
21	2.2963	h>a>ci>b>d>f>j>e>g	*statistical_sparsity* >= 0.4085
21	2.7049	h>c>bg>i>d>f>a>e>j	*typology_border* >= 0.6555
21	2.4990	h>a>d>b>c>e>g>fi>j	*complexity_t3* >= 0.0668
22	2.2101	h>a>g>c>f>e>d>i>b>j	*complexity_t2* >= 0.1250
41	2.1513	c>h>f>b>a>d>e>i>j>g	*complexity_f3* >= 0.9831
41	2.1513	c>h>f>b>a>d>e>i>j>g	*complexity_f4* >= 0.9831
21	2.3544	h>c>b>a>f>d>ei>g>j	*landmarking_elite_nn* <= 0.5788
21	2.9550	h>c>a>b>f>dg>i>e>j	*landmarking_best_node* <= 0.6557

Costa, Santos, Soares & Abreu. A Meta-Analysis on the Recommendation of Imbalance Strategies *(submitted)*

wrap-up

- model management
  - exciting field
    - e.g. autoML (http://www.automl.org/)
  - new challenges
- do not forget the basic issues
  - … not all of them, at least
- learn from other areas
  - e.g., algorithm portfolios

Smith-Miles. Cross-disciplinary perspectives on meta-learning for algorithm selection.
ACM Comput. Surv. 2008;41(1):1-25

Kotthoff, Algorithm Selection for Combinatorial Search Problems: A Survey
AI Magazine, 2014

acknowledgements

- Supervisors: P. Brazdil (UP), J.P. Costa (UP)
- Colleagues: METAL project, A. Carvalho (USP), R. Prudêncio (UFPE), C. Giraud-Carrier (BYU), R. Vilalta (UT), P. Flach (UB), H. Ferreira (UP)
- Students: P. Abreu (UP), C. Félix (UP), C. Gomes (UP), F. Pinto (UP), M. Nozari (UP), T. Cunha (UP), T. Gomes (UFPE), J. Kanda (USP), T. Lucas (UFPE), P. Miranda (UFPE), E. Partodikromo (UL), F. Pinto (UP), C. Rebelo (UP), A. Rossi (USP), B. Souza (USP)

“THE” Book

Metalearning – Applications to Data Mining

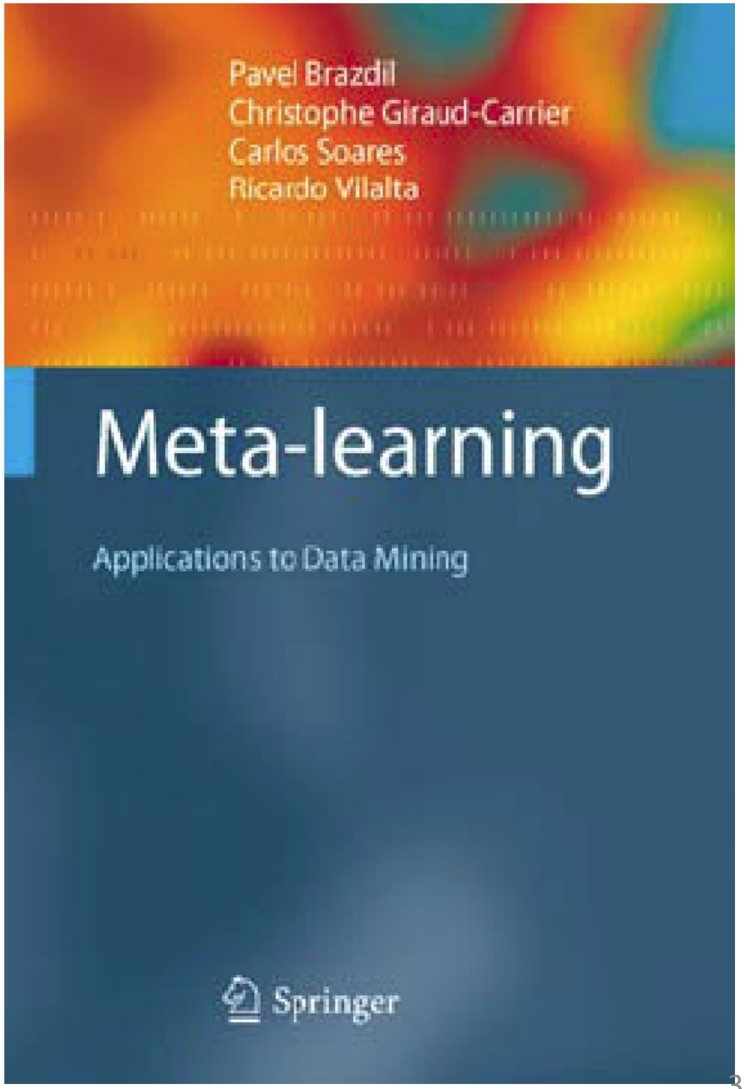
Pavel Brazdil

Christophe Giraud-Carrier

Carlos Soares

Ricardo Vilalta

http://www.springer.com/computer/artificial/book/978-3-540-73262-4



another book (not so interesting… ;-))

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

AutoML.org Freiburg

Home Blog AutoML AAD Analysis Book Jobs Events Team & Partners

AUTOML: METHODS, SYSTEMS, CHALLENGES (NEW BOOK)

Editors: Frank Hutter, Lars Kotthoff, Joaquin Vanschoren

We're in the process of finishing this edited book, and it will be ready for sale by NIPS 2018. Next to publishing, we will keep the book open access. Below, we share preliminary versions of the chapters; at this point in time, **these are all drafts, before copy editing.**

Part 1: AutoML Methods

This part comprises highly up-to-date overview chapters on the common foundations behind all AutoML systems.

Chapter 1: Hyperparameter Optimization. By Matthias Feurer and Frank Hutter

Chapter 2: Meta Learning. By Joaquin Vanschoren

Chapter 3: Neural Architecture Search. By Thomas Elsken, Jan-Hendrik Metzen and Frank Hutter

Part 2: AutoML Systems

This part comprises in-depth descriptions of a broad range of available AutoML systems that can be used for effective machine learning out of the box.

Chapter 4: Auto-WEKA. By Lars Kotthoff and Chris Thornton and Holger H. Hoos and Frank Hutter and Kevin Leyton-Brown

https://www.automl.org/book/