AutoML: Automated construction of classification models

Hugo Jair Escalante

http://ccc.inaoep.mx/~hugojair

Coordinación de Ciencias Computacionales

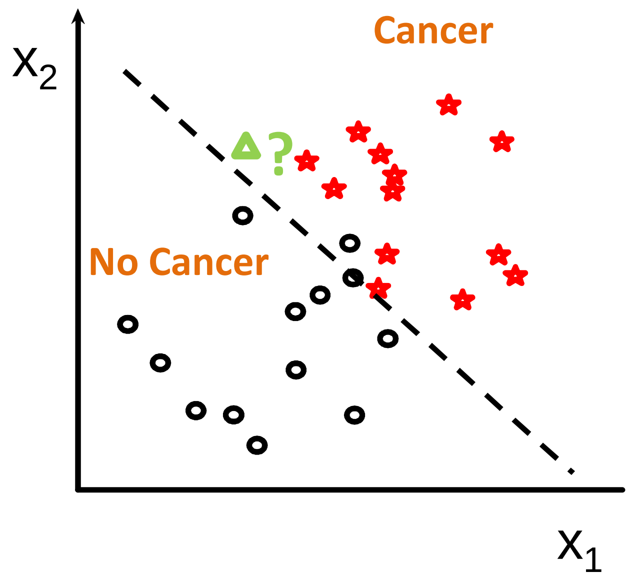
AUTOMATED CONSTRUCTION OF CLASSIFICATION MODELS

Outline

- Pattern classification
- Model selection in a broad sense: FMS
- Related work (overview)
- PSO for full model selection
- Experimental results and applications
- Conclusions and future work directions

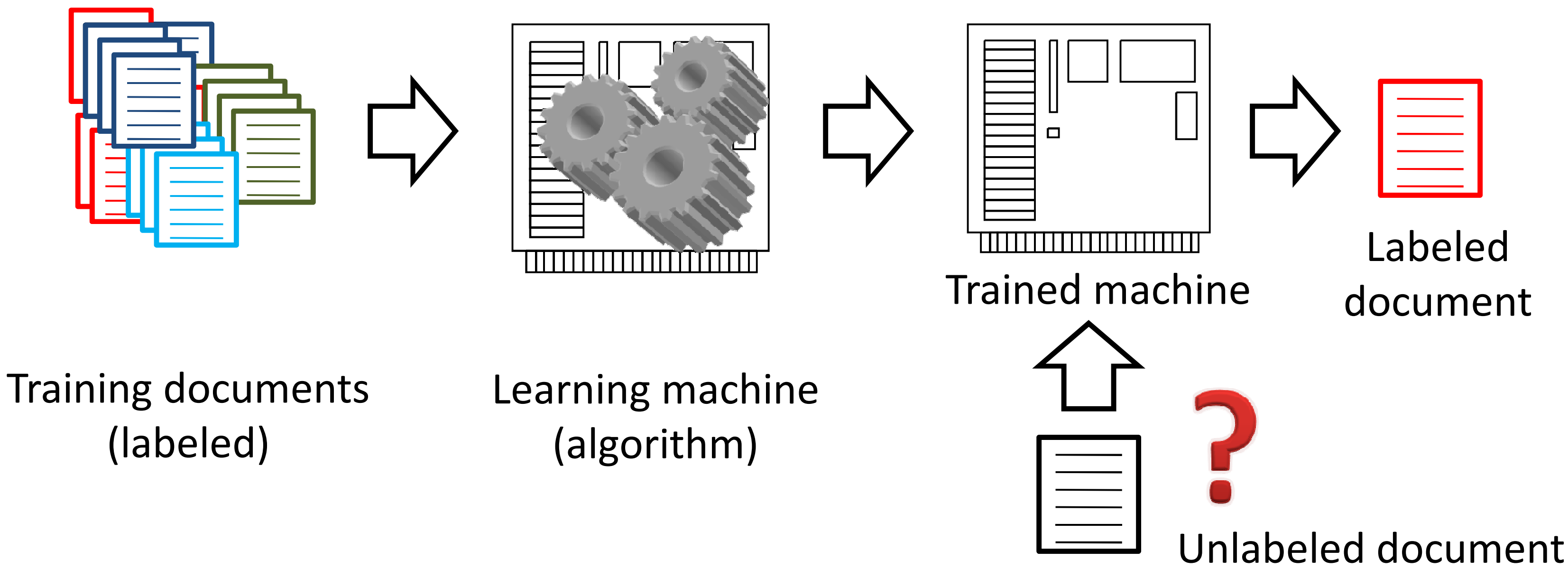
Pattern classification

- To learn a model able to make predictions regarding a variable of interest, using a set of other variables. E.g., classifying:
  - Emails as spam vs. safe-email
  - Topographies as tumor vs. non-tumor, or malign vs. benign
  - Face recognition
  - Hand-written character recognition

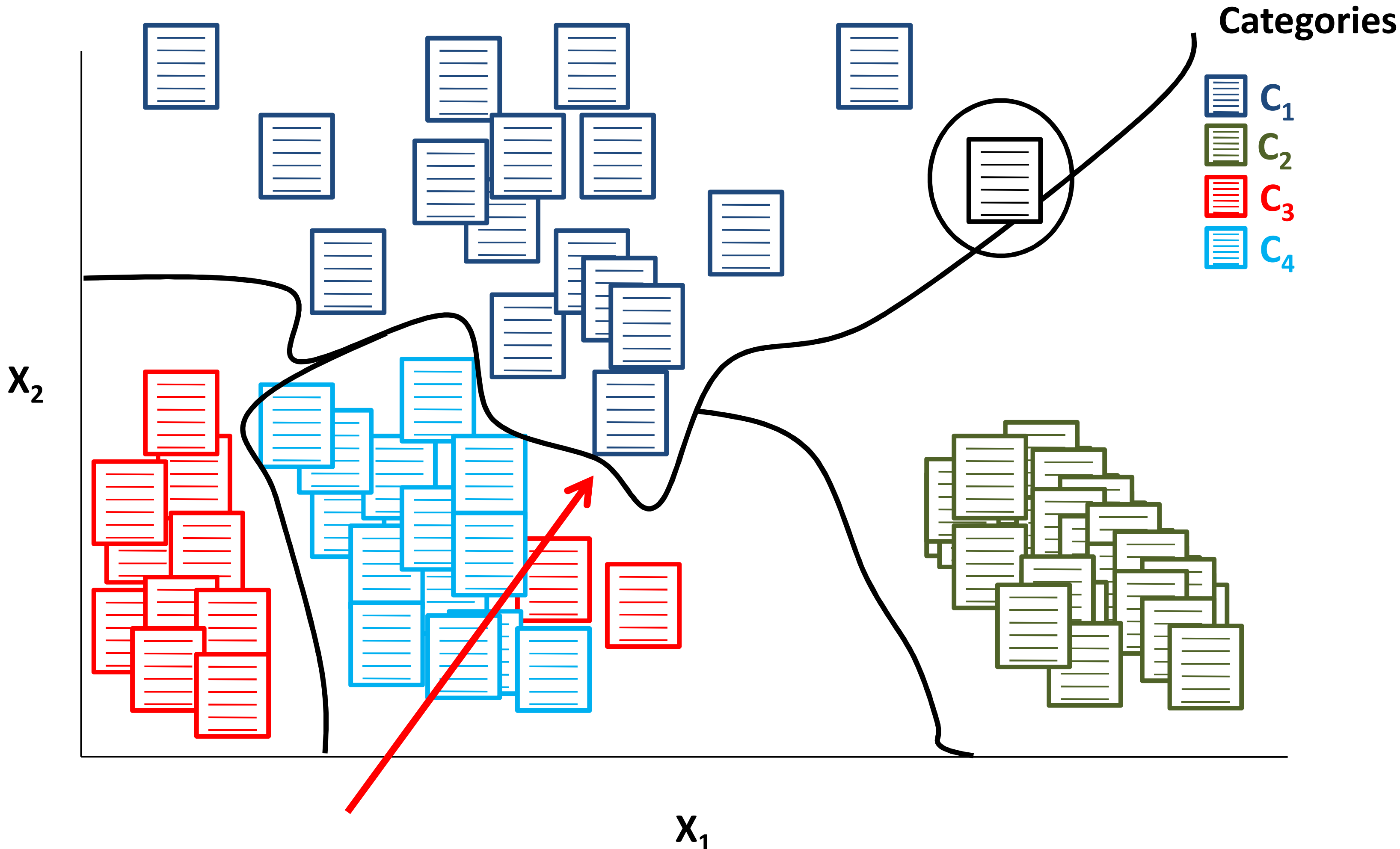


Machine learning: classification

- To learn a model able to make predictions regarding a variable of interest, using a set of other variables. Example: *text categorization*



Machine learning: classification



How to learn these functions?

What is a classifier?

- A function:

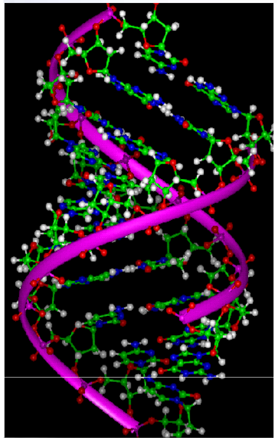
$$f : \Re^d \to C \qquad C = \{C_1, \ldots, C_K\}$$

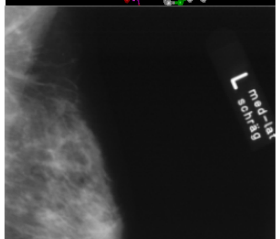
$$f : (\Re^d, C) \to \{0,1\}$$

- Given:

$$D = \{(\mathbf{x}_i, y_i)\}_{1,\ldots,N}$$

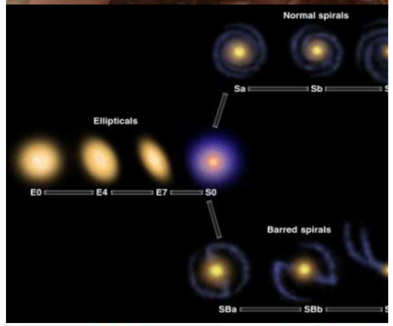
$$\mathbf{x}_i \in \Re^d ; y_i \in C$$

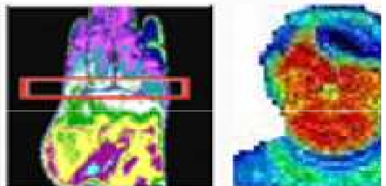


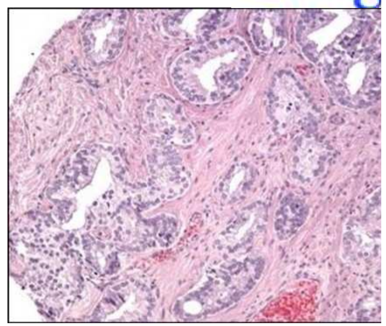


Applications

- Natural language processing,
- Computer vision,
- Robotics,
- Information technology,
- Medicine,
- Science,
- Entertainment,
- ….

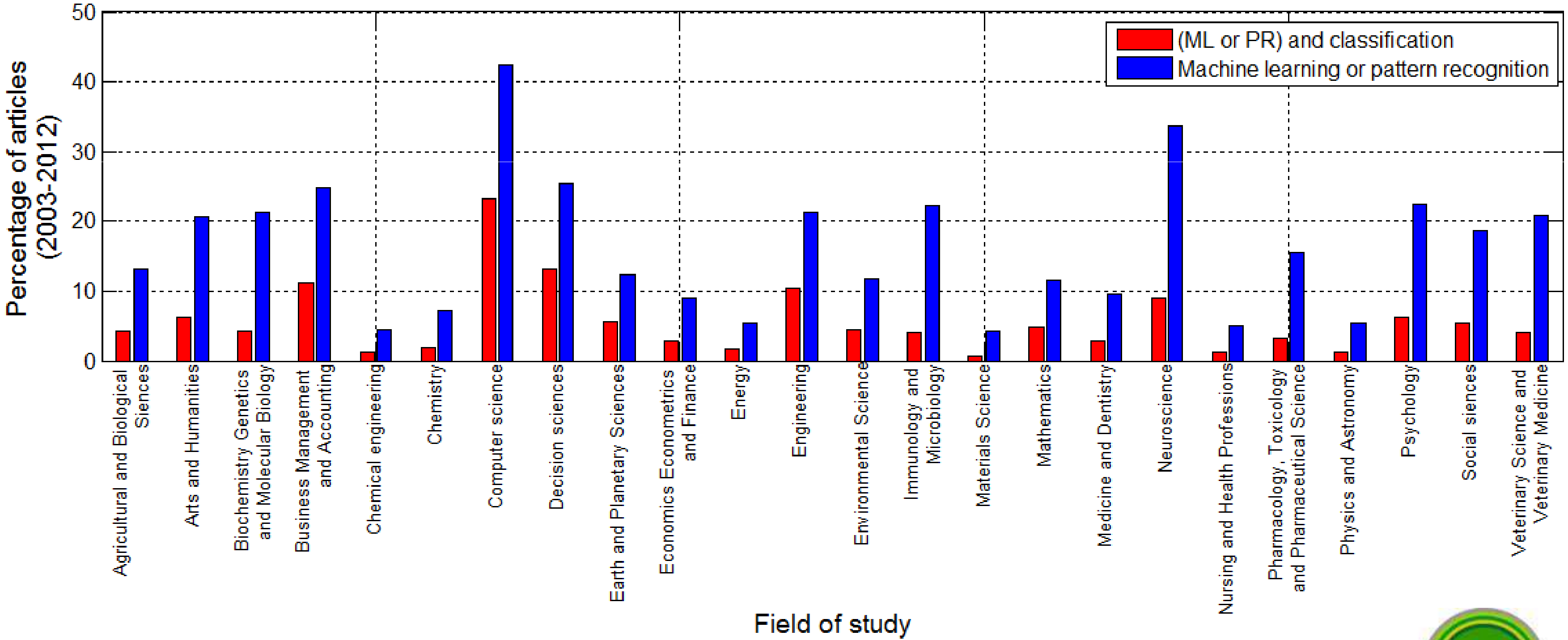






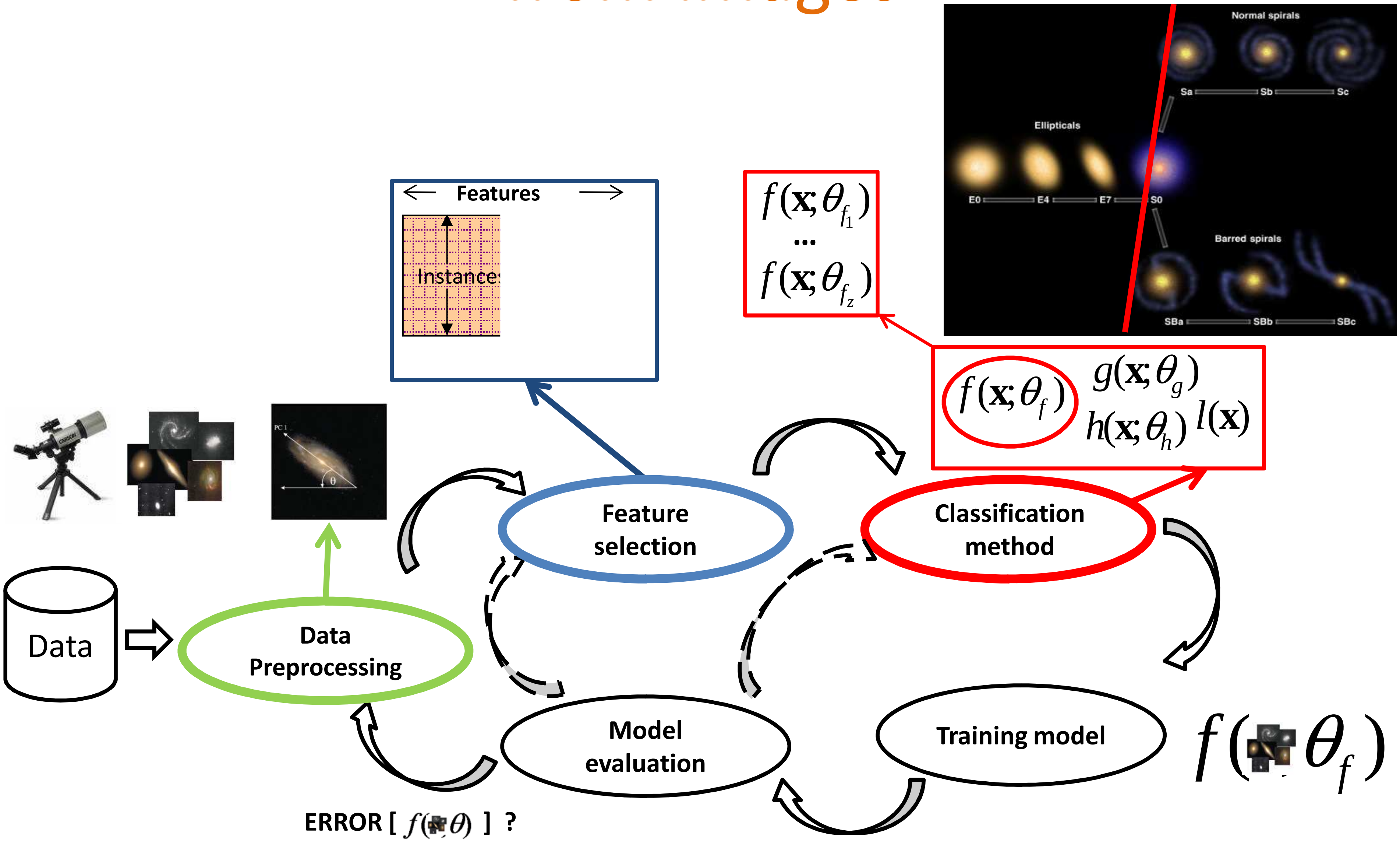
Pattern classification

- Ubiquitous problem within computer science, (popular in other sciences and non-science applications)

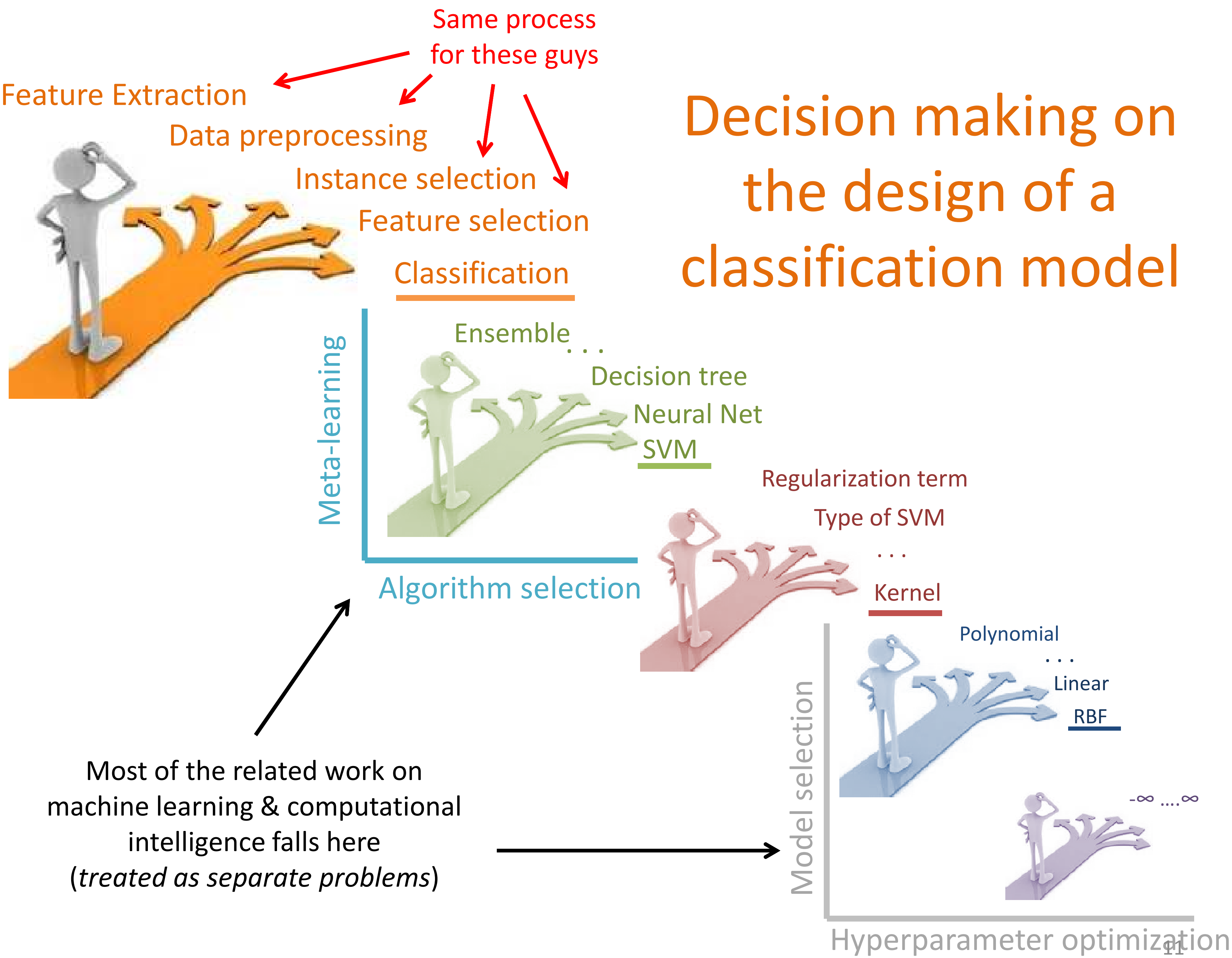


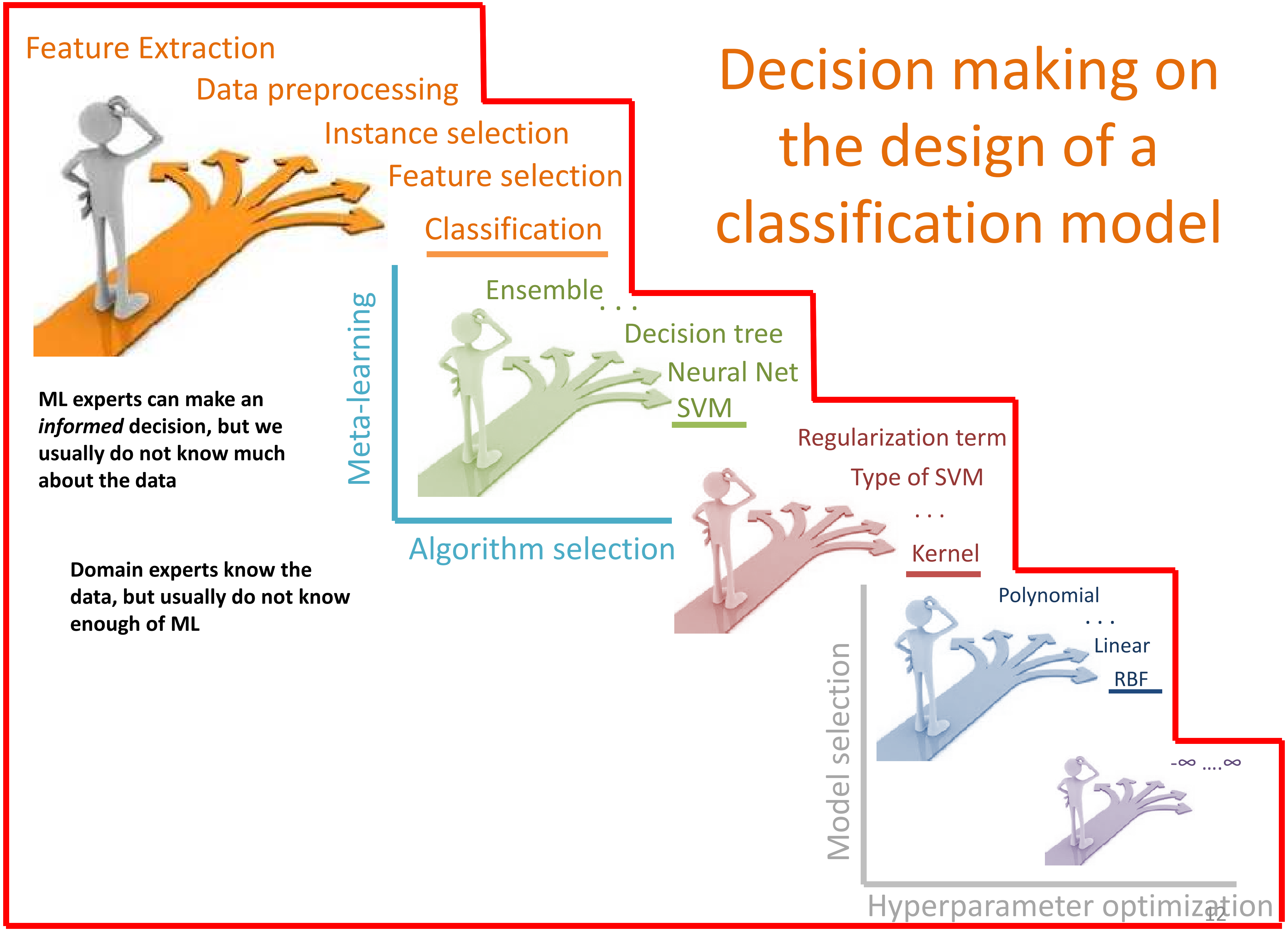
Per-field percentage of articles published in journals (2003-2012) including the terms: **1) machine learning or pattern recognition**; **2) machine learning or pattern recognition and classification**

Example: galaxy classification from images



Decision making on the design of a classification model




Decision making on the design of a classification model
Feature Extraction
Data preprocessing
Instance selection
Feature selection
Classification
ML experts can make an *informed* decision, but we usually do not know much about the data
Domain experts know the data, but usually do not know enough of ML
Meta-learning
Ensemble
. . .
Decision tree
Neural Net
SVM
Algorithm selection
Regularization term
Type of SVM
. . .
Kernel
Polynomial
. . .
Linear
RBF
-∞ ....∞
Model selection
Hyperparameter optimization

Some issues with the cycle of design

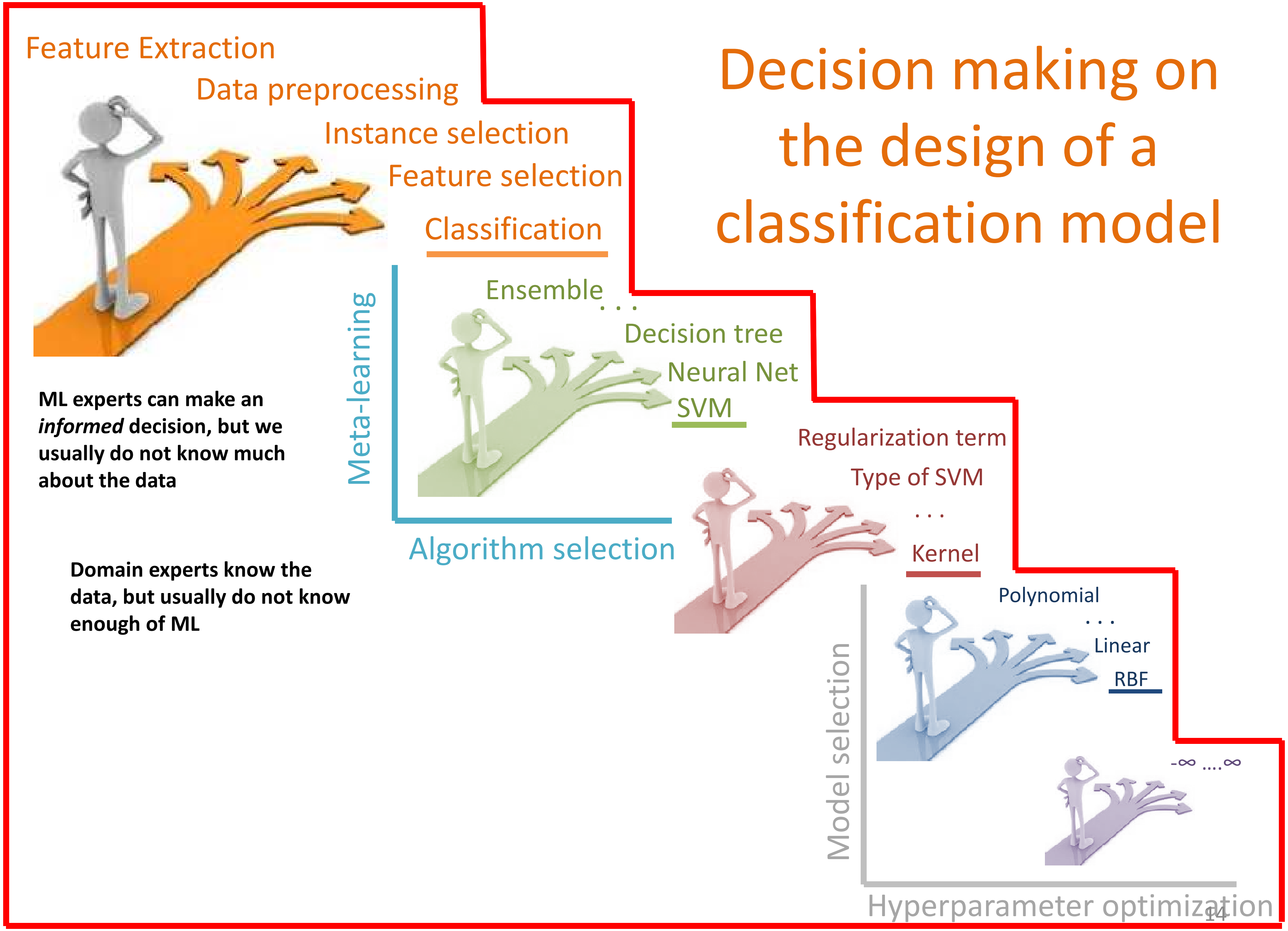
- The above issues are usually fixed manually:
  - Domain expert's knowledge
  - Machine learning specialists' knowledge
  - Trial and error

- The design/development of a pattern classification system relies on the knowledge and biases of humans, which may be risky, expensive and time consuming

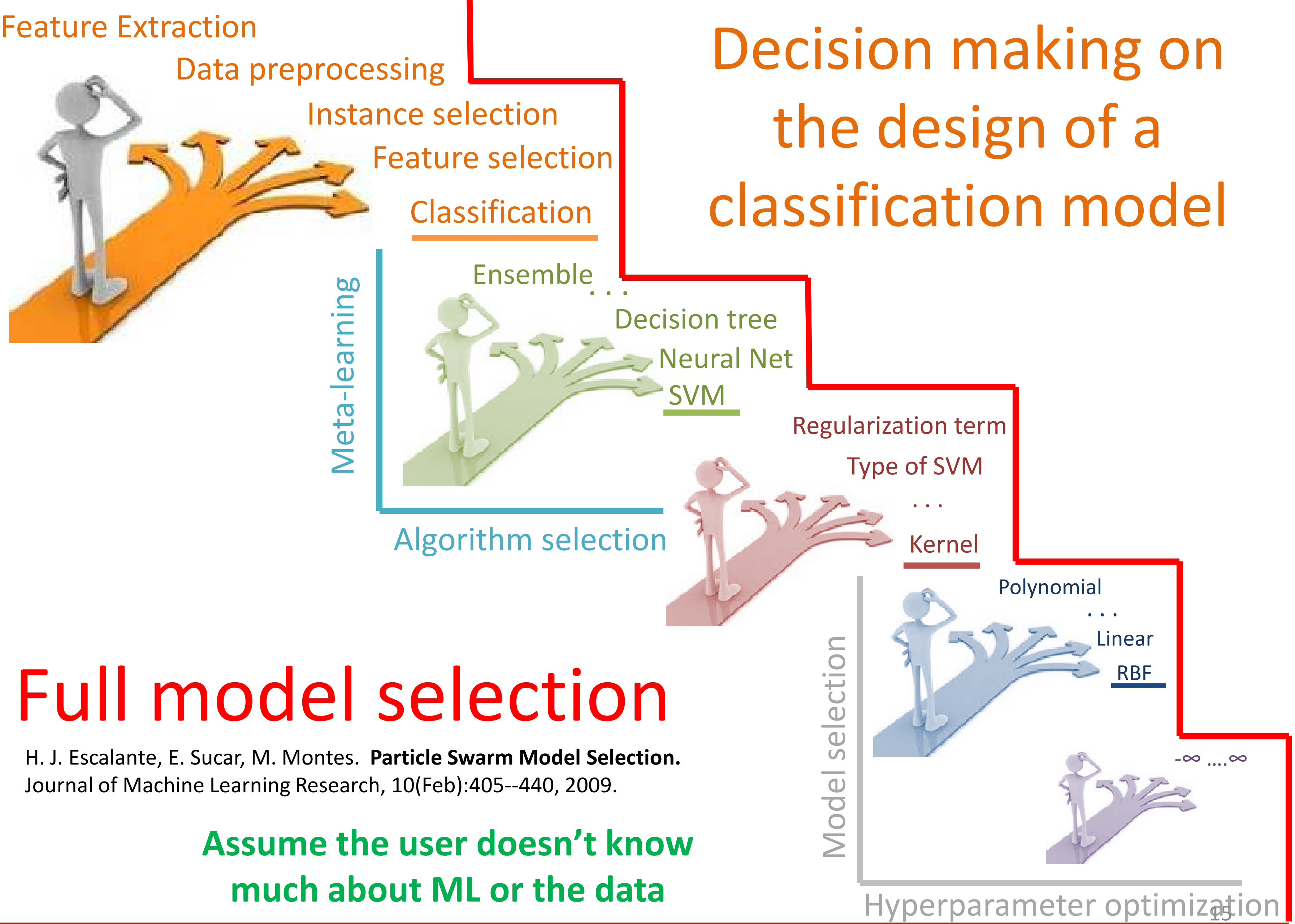
- Automated solutions are available but only for particular processes (e.g., either feature selection, or classifier selection but not both)

It is possible to automate the whole process?

Decision making on the design of a classification model



Decision making on the design of a classification model

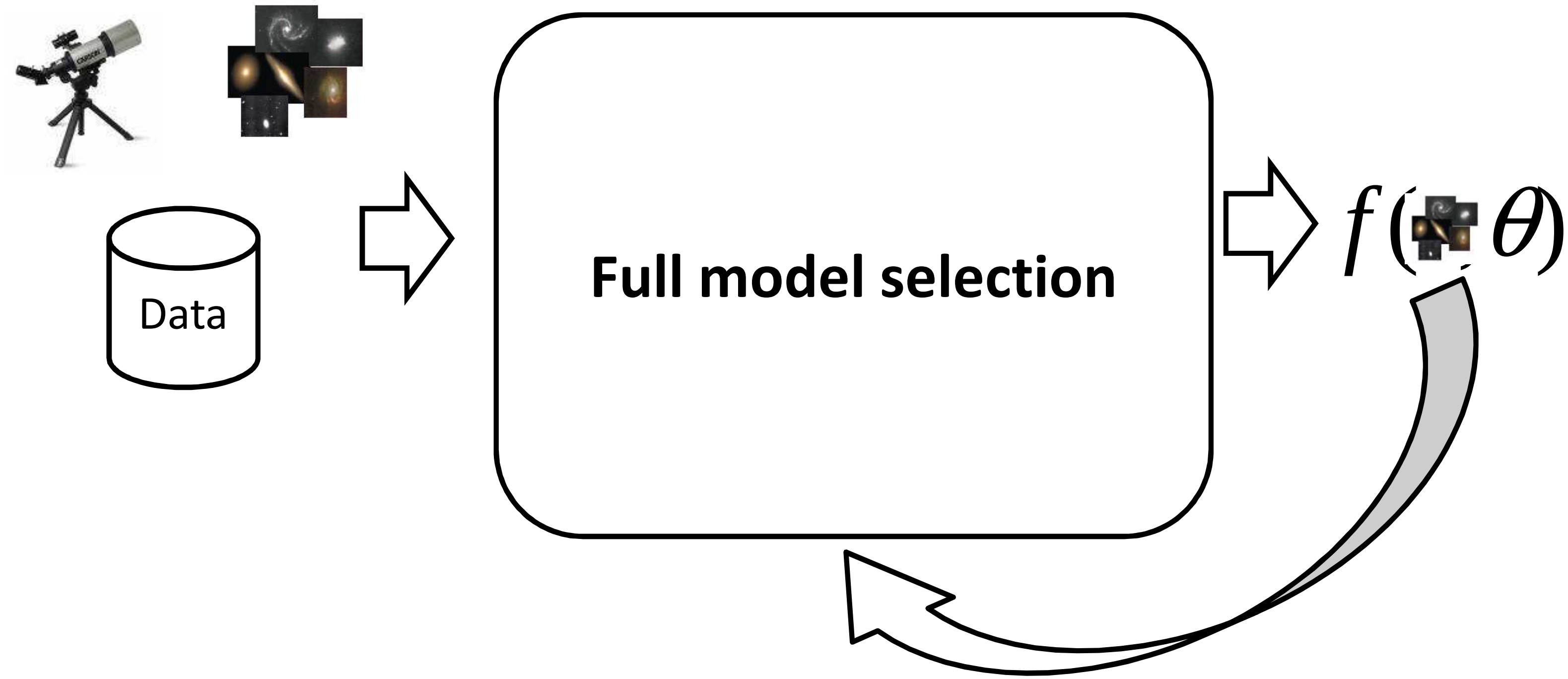


Full model selection

H. J. Escalante, E. Sucar, M. Montes. **Particle Swarm Model Selection.** Journal of Machine Learning Research, 10(Feb):405--440, 2009.

Assume the user doesn't know much about ML or the data

Full model selection



H. J. Escalante. **Towards a Particle Swarm Model Selection algorithm.** *Multi-level inference workshop and model selection game, NIPS, Whistler, VA, BC, Canada, 2006.*

H. J. Escalante, E. Sucar, M. Montes. **Particle Swarm Model Selection,** *In Journal of Machine Learning Research, 10(Feb):405--440, 2009.*

Full model selection

- Given a set of methods for data preprocessing, feature selection and classification select the combination of methods (together with their hyperparameters) that minimizes an estimate of classification performance

H. J. Escalante, E. Sucar, M. Montes. **Particle Swarm Model Selection,** *In Journal of Machine Learning Research, 10(Feb):405--440, 2009.*

Full model selection

- **Full model:** A model composed of methods for data preprocessing, feature selection and classification

- **Example:**

```
chain
 {
  1:standardize center=1
  2:svcrfe svc kernel linear coef0=0 degree=1 gamma=0 shrinkage=0.001 f_max=Inf
  3:pc_extract f_max=2000
  4:svm kernel linear C=Inf ridge=1e-013 balanced_ridge=0 nu=0 alpha_cutoff=-1 b0=0 nob=0
 }
```

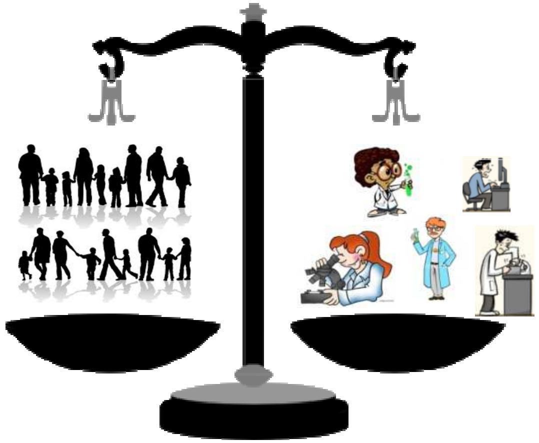
Full model selection

- Pros
  - The job of the data analyst is considerably reduced
  - Neither knowledge on the application domain nor on machine learning is required
  - Different methods for preprocessing, feature selection and classification are considered
  - It can be used in any classification problem

- Cons
  - It is real function + combinatoric optimization problem
  - Computationally expensive
  - Risk of overfitting

… and full models for all

- **Short-term goal:** provide data analysts with tool that allows them to build effective classification systems without much effort
- **Long-term goal:** An APP that allows anyone to build a classification model from their data (photographs, smart phone data, tweets, etc.)



OVERVIEW OF RELATED WORKS

Model selection via heuristic optimization

- A single model is considered and their hyperparameters are optimized via heuristic optimization:
  - Swarm optimization,
  - Genetic algorithms,
  - Pattern search,
  - Genetic programming
  - ...

B. Zhang and H. Muhlenbein . **Evolving optimal neural networks using genetic algorithms with Occam's razor.** *Complex Systems*, Vol. 7 (1993), pp. 199-220

T. Howley, M. Madden. **The Genetic Kernel Support Vector Machine: Description and Evaluation.** Artificial Intelligence Review, Vol 24(3-4): 379—395, 2005.

M. Momma, K. Bennett. **A Pattern SearchMethod for ModelSelection of SupportVector Regression.** Proceedings of the SIAM conference on data mining, pp. 261—274, 2002.

GEMS

- **GEMS** (**G**ene **E**xpression **M**odel **S**election) is a system for automated development and valuation of high-quality cancer diagnostic models and biomarker discovery from microarray gene expression data

GEMS

Gene Expression Model Selector

A. Statnikov, I. Tsamardinos, Y. Dosbayev, C.F. Aliferis. **GEMS: A System for Automated Cancer Diagnosis and Biomarker Discovery from Microarray Gene Expression Data**. International Journal of Medical Informatics, 2005 Aug;74(7-8):491-503.

N. Fananapazir, A. Statnikov, C.F. Aliferis. **The Fast-AIMS Clinical Mass Spectrometry Analysis System**. Advaces in Bioinformatics, 2009, Article ID 598241.

GEMS

- The user specifies the models, and methods to be considered
- GEMS explores all of the combinations of methods, using grid search to optimize hyperparameters

Gene Expression Model Selector

<u>Nested cross-validation:</u>

1. Repeat N times:
    - *Training set* ← *N-1* subsets;
    - *Testing set* ← remaining subset;

    1.1. Repeat for $i = 1, \ldots, m$:

    a. Repeat *N-1* times (<u>for samples only in the *training set*</u>):
    - ○ *Training_validation* set ← *N-2* subsets;
    - ○ *Testing_validation set* ← remaining subset;
    - ○ Train the classifier *A* on the *training_validation set* using parameter α_i;
    - ○ Test it on the *testing_validation set*.

    b. Record $P(i)$, the average performance of *A* over *N-1 testing_validation sets*.

    1.2. Determine α_j, where $j = \text{argmax } P(i)$ for $i = 1, \ldots, m$;

    1.3. Train the classifier *A* on the *training set* using parameter α_j.

    - Test the classifier obtained in step 1.3 on the *testing set*.
2. Return ρ, the average performance of *A* over *N testing sets*.

GEMS

- The user specifies the models, and methods to be considered
- GEMS explores all of the combinations of methods, using grid search to optimize hyperparameters

GEMS
Gene Expression Model Selector

Nested cross-validation:

1. Repeat N times:
   - *Training set* ← *N-1* subsets;
   - *Testing set* ← remaining subset;

   1.1. Repeat for $i = 1, \ldots, m$:
      a. Repeat *N-1* times (for samples only in the *training set*):
         - *Training_validation* set ← *N-2* subsets;
         - *Testing_validation set* ← remaining subset;
         - Train the classifier *A* on the *training_validation set* using parameter α_i;
         - Test it on the *testing_validation set*.

      b. Record $P(i)$, the average performance of *A* over *N-1 testing_validation sets*.

   1.2. Determine α_j, where $j = \text{argmax } P(i)$ for $i = 1, \ldots, m$;

   1.3. Train the classifier *A* on the *training set* using parameter α_j.

   - Test the classifier obtained in step 1.3 on the *testing set*.
2. Return ρ, the average performance of *A* over *N testing sets*.

Model type selection for regression

- Genetic algorithms are used for the selection of model type (learning method, feature selection, preprocessing) and parameter optimization for regression problems

http://www.sumo.intec.ugent.be/

D. Gorissen, T. Dhaene, F. de Turck. **Evolutionary Model Type Selection for Global Surrogate Modeling.** *In Journal of Machine Learning Research,* 10(Jul):2039-2078, 2009

Meta-learning: learning to learn

- Evaluates and compares the application of learning algorithms on (many) previous tasks/domains to suggest learning algorithms (combinations, rankings) for new tasks

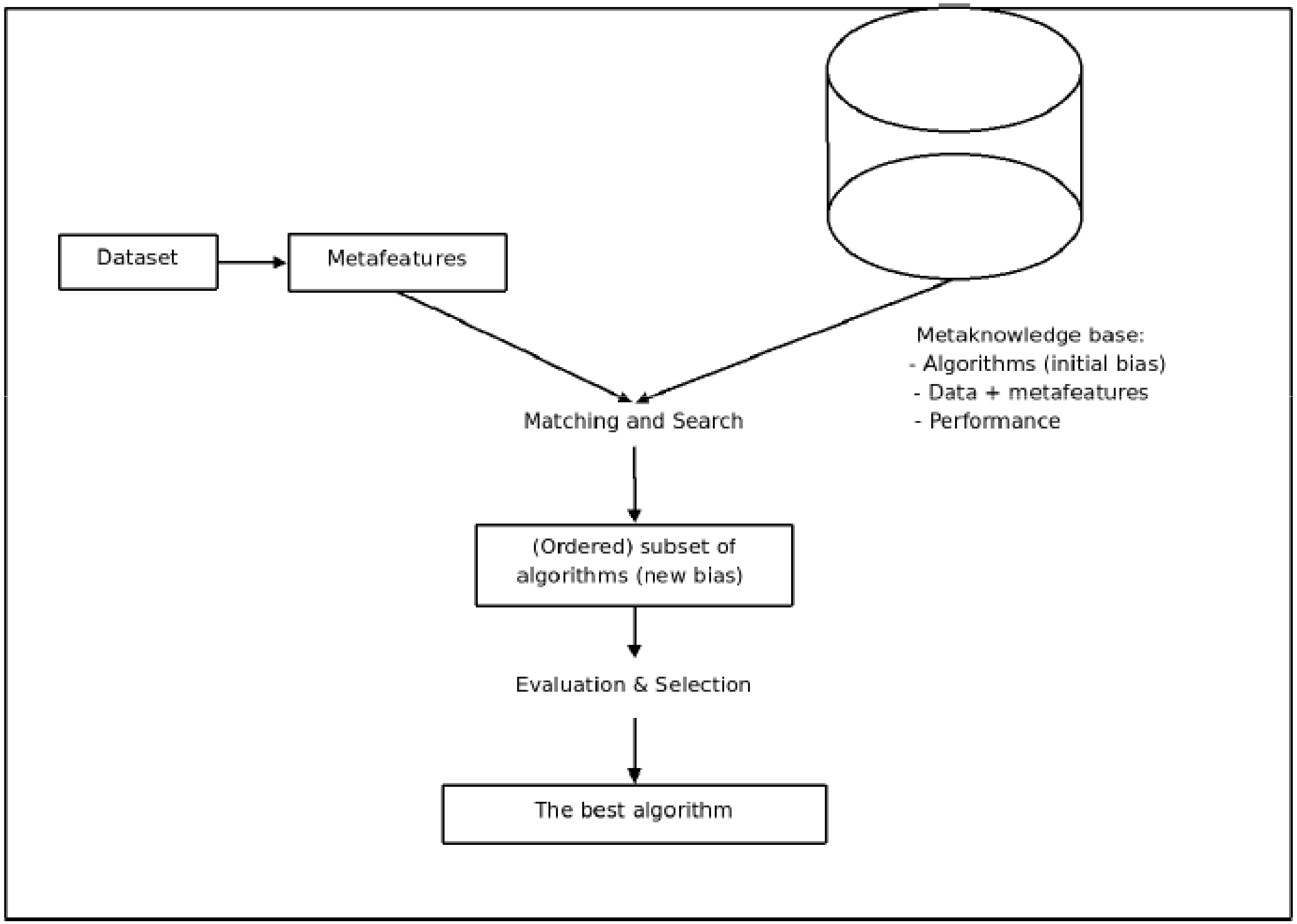
- Focuses on the relation between tasks/domains and learning algorithms

- Accumulating experience on the performance of multiple applications of learning methods

Brazdil P., Giraud-Carrier C., Soares C., Vilalta R. **Metalearning: Applications to Data Mining.** Springer Verlag. ISBN: 978-3-540-73262-4, 2008.

Brazdil P., Vilalta R, Giraud-Carrier C., Soares C.. **Metalearning.** Encyclopedia of Machine learning. Springer, 2010.

Meta-learning: learning to learn



Brazdil P., Giraud-Carrier C., Soares C., Vilalta R. **Metalearning: Applications to Data Mining.** Springer Verlag. ISBN: 978-3-540-73262-4, 2008.

Brazdil P., Vilalta R, Giraud-Carrier C., Soares C.. **Metalearning.** Encyclopedia of Machine learning. Springer, 2010.

Google prediction API

Google Developers

- "Machine learning as a service in the cloud"
- Upload your data, train a model and perform queries
- **Nothing is for free!**

Make Big Data Analysis Easy — Google

With machine learning as a web service
- Don't need to provision large number of machines
- Don't require substantial investment upfront
- Don't require deep machine learning expertise
- Easy to integrate with existing apps and deploy models

https://developers.google.com/prediction/

Google prediction API

Google Developers

Three steps to use the Prediction API

Google

1. Upload	Upload your training data to Google Storage	Use the Google Storage API or gsutil to upload your {data} to Google Storage
2. Train	Build a model from your data	POST prediction/v1/training/{data} make a training request GET prediction/v1/training/{data} get training status
3. Predict	Make new predictions	POST prediction/v1/training/{data}/predict make a prediction request

Your data become property of Google!

IBM's SPSS modeler

Modeling: Automated

The automated modeling nodes estimate and compare a number of different modeling methods, allowing you to try out a variety of approaches in a single modeling run. You can select the modeling algorithms to use, and the specific options for each. The node explores every possible combination of options, ranks each candidate model based on the measure you specify, and saves the best for use in scoring or further analysis.

$$$$$

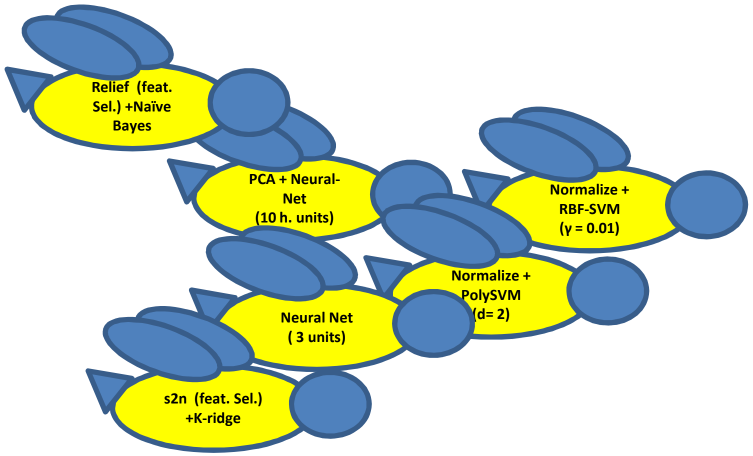
Automated machine learning

- Interest from diverse fronts (recently):
  - Research agencies (e.g., IARPA, DARPA)
  - Industry (e.g., Google, IBM, ORACLE, Microsoft)
  - Researchers (machine learning, computational intelligence)

PARTICLE SWARM MODEL SELECTION

PSMS: Our approach to full model selection

- **Particle swarm model selection**: Use ***particle swarm optimization*** for exploring the search space of full models in a particular ML-toolbox



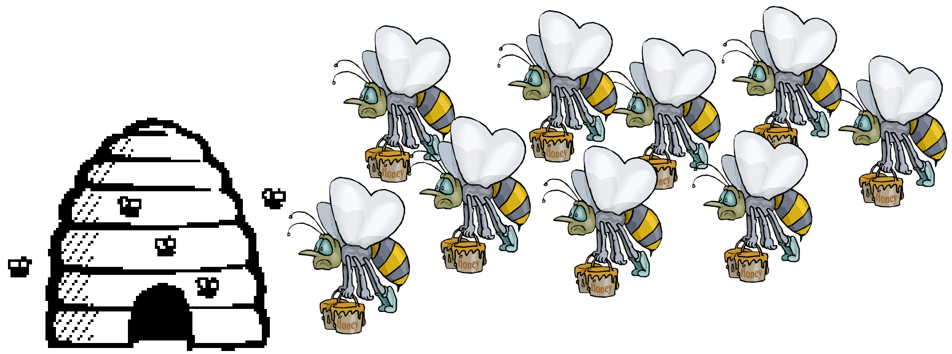
H. J. Escalante. **Towards a Particle Swarm Model Selection algorithm.** *Multi-level inference workshop and model selection game, NIPS, Whistler, VA, BC, Canada, 2006.*

H. J. Escalante, E. Sucar, M. Montes. **Particle Swarm Model Selection,** *In Journal of Machine Learning Research, 10(Feb):405--440, 2009.*

Particle swarm optimization

- Population-based search heuristic

- Inspired on the behavior of biological communities that exhibit local and social behaviors



A. P. Engelbrecht. **Fundamentals of Computational Swarm Intelligence.** Wiley,2006

Particle swarm optimization

- Each individual (particle) *i* has:
  - A position in the search space (**X_i^t**), which represents a solution to the problem at hand,
  - A velocity vector (**V_i^t**), which determines how a particle explores the search space
- After random initialization, particles update their positions according to:

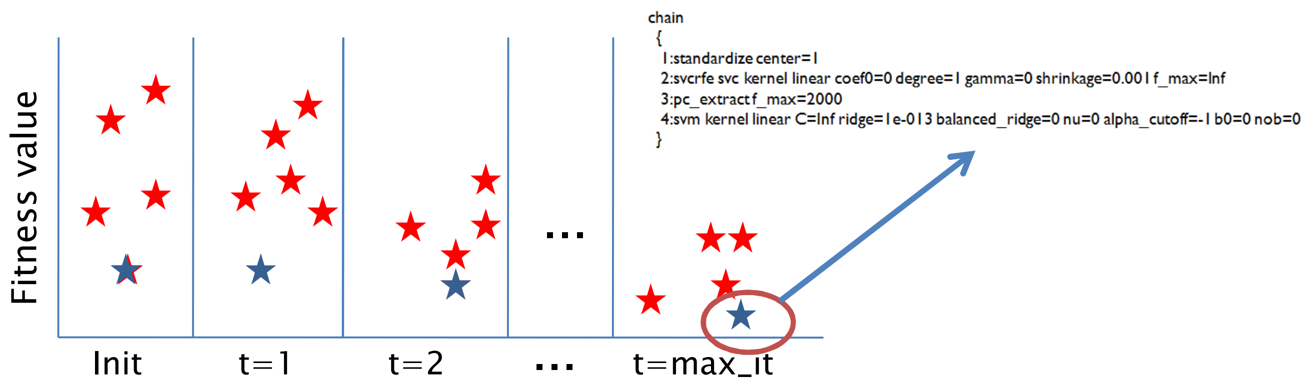
$$\mathbf{x}^{t+1}_i = \mathbf{v}^{t+1}_i + \mathbf{x}^t_i$$

$$\mathbf{v}^{t+1}_i = \phi_0 \times \mathbf{v}^t_i + \phi_1 \boxed{\times (\mathbf{p}_i - \mathbf{x}^t_i)} + \phi_2 \boxed{\times (\mathbf{p}_g - \mathbf{x}^t_i)}$$

A. P. Engelbrecht. **Fundamentals of Computational Swarm Intelligence.** Wiley,2006

Particle swarm optimization

1. Randomly initialize a population of particles (i.e., the swarm)
2. Repeat the following iterative process until stop criterion is meet:
   a) Evaluate the fitness of each particle
   b) Find personal best ($\mathbf{p}_i$) and global best ($\mathbf{p}_g$)
   c) Update particles
   d) Update best solution found (if needed)
3. Return the best particle (solution)



PSMS : PSO for full model selection

- Set of methods (not restricted to this set)

	Object name	Type	# pars.	Description
Classification	*zarbi*	C	0	Linear classifier
	naive	C	0	Naïve Bayes
	logitboost	C	3	Boosting with trees
	neural	C	4	Neural network
	svc	C	4	SVM classifier
	kridge	C	4	Kernel ridge regression
	rf	C	3	Random forest
	lssvm	C	5	Kernel ridge regression
Feature selection	*Ftest*	F	4	F-test criterion
	Ttest	F	4	T-test criterion
	aucfs	F	4	AUC criterion
	odds-ratio	F	4	Odds ratio criterion
	relief	F	3	Relief ranking criterion
	Pearson	F	4	Pearson correlation coefficient
	ZFilter	F	2	Statistical filter
	s2n	F	2	Signal-to-noise ratio
	$pc - extract$	F	1	Principal components analysis
	svcrfe	F	1	SVC- recursive feature elimination
Preprocessing	*normalize*	P	1	Data normalization
	standardize	P	1	Data standardization
	$shift - scale$	P	1	Data scaling

http://clopinet.com/CLOP

PSMS : PSO for full model selection

- Codification of solutions as real valued vectors

Hyperparameters for the selected methods

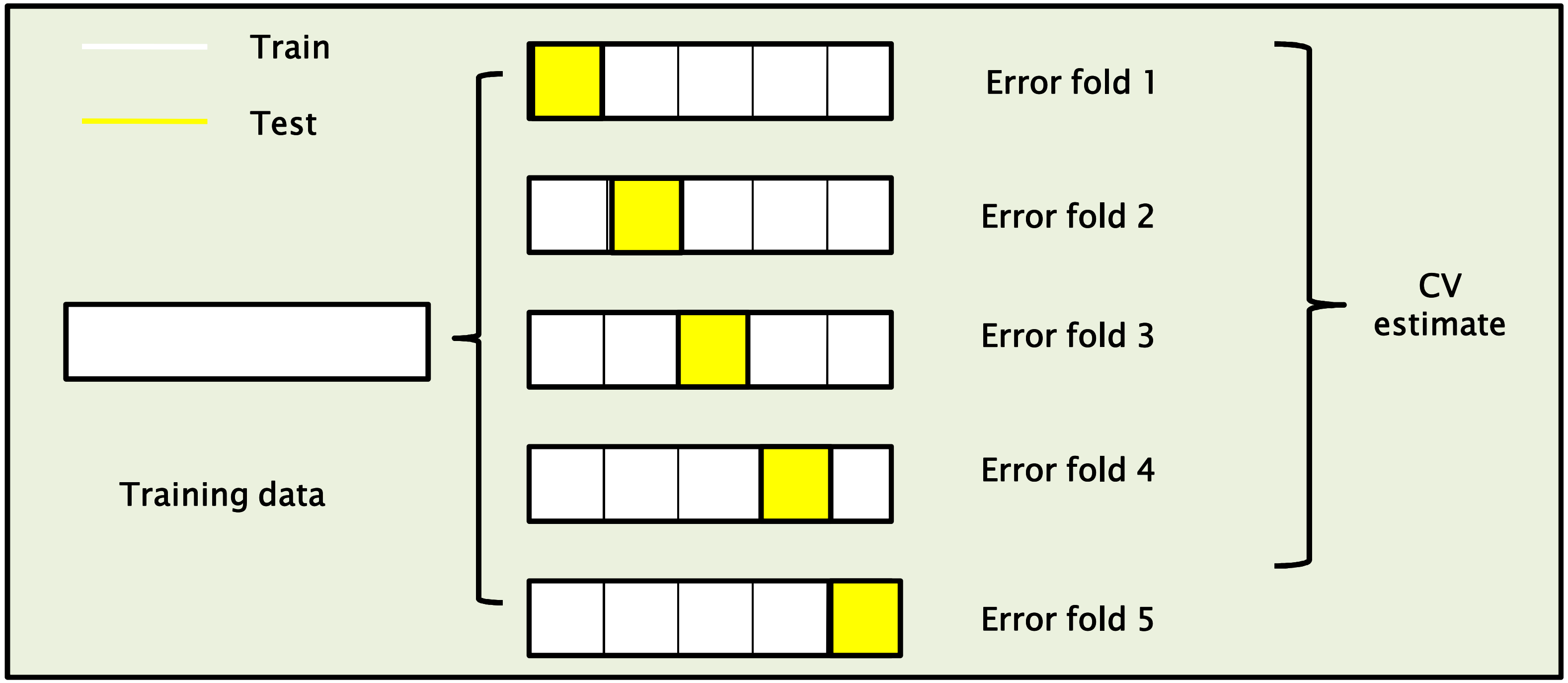
Preprocessing before feature selection?

$$\mathbf{x}_i = \left\langle \boxed{x_{\mathrm{i,pre}}}, \boxed{y_{\mathrm{i,1...Npre}}}, \boxed{x_{\mathrm{i,fs}}}, \boxed{y_{\mathrm{i,1...Nfs}}}, \boxed{x_{\mathrm{i,sel}}}, \boxed{x_{\mathrm{i,class}}}, \boxed{y_{\mathrm{i,1...Nclass}}} \right\rangle$$

Choice of methods for preprocessing, feature selection and classification

PSMS : PSO for full model selection

- Fitness function:
  - K-fold cross-validation balanced error rate
  - K-fold cross-validation area under the ROC curve



SOME EXPERIMENTAL RESULTS

PSMS in the ALvsPK challenge

- Five data sets for binary classification
- Goal: to obtain the best classification model for each data set
- Two tracks:
  - Prior knowledge
  - Agnostic learning

Dataset	Domain	Number of examples (training/validation /test)	Positive class (% num. ex.)	Number of features	
				Raw data (for PK)	Preprocessed (for AL)
ADA	Marketing	4147 / 415 / 41471	28.4	14	48
GINA	HWR	3153 / 315 / 31532	49.2	784	970
HIVA	Drug discovery	3845 / 384 / 38449	3.5	Molecules	1617
NOVA	Text classification	1754 / 175 / 17537	28.5	Text	16969
SYLVA	Ecology	13086 / 1309 / 130857	6.2	108	216

http://www.agnostic.inf.ethz.ch/

PSMS in the ALvsPK challenge

- Best configuration of PSMS:

Entry	Description	Ada	Gina	Hiva	Nova	Sylva	Overall	Rank
Interim-all-prior	Best PK	17.0	2.33	27.1	4.71	0.59	10.35	1st
psmsx_jmlr_run_I	PSMS	16.86	**2.41**	**28.01**	5.27	**0.62**	**10.63**	2nd
Logitboost-trees	Best AL	**16.6**	3.53	30.1	**4.69**	0.78	11.15	8th

Comparison of the performance of models selected with PSMS with that obtained by other techniques in the ALvsPK challenge

Data	SF	Model	Time (m)	Test-BER
Ada	1	chain({logitboost(units=469,shrinkage=0.4,depth=1),bias}	368.12	16.86
Gina	2	chain({sns(1),relief(fmax=487),gkridge,bias}	482.23	2.41
Hiva	3	chain({norm(1),rffs(fmax=1001),lssvm(gamma=0.096),bias}	124.54	28.01
Nova	1	chain({rffs(fmax=338),norm(1),std(1),sns(1),gkridge,bias}	82.12	5.27
Sylva	10	chain({sns(1),odds-ratio(fmax=60),gkridge,bias}	787.58	0.62

Models selected with PSMS for the different data sets

http://www.agnostic.inf.ethz.ch/results.php

PSMS in the ALvsPK challenge

- Official ranking:

Rank	Method	Balanced Error			Area Under Curve			Date	Name
		Train	Valid	Test	Train	Valid	Test		
1	interim all prior	0.0305	0.0934	0.1035	0.9893	0.9113	0.9332	2007-01-31 21:54:23	reference (gcc)
2	psmsx_jmlr_run_I	0.0482	0.0669	0.1063	0.9508	0.9355	0.8941	2008-10-18 01:29:25	H. Jair Escalante
3	psmsx_jmlr_run	0.0481	0.0691	0.1065	0.9513	0.9348	0.8938	2008-10-18 01:20:04	H. Jair Escalante
4	the bad	0.033	0.1002	0.1085	0.9876	0.9269	0.9332	2006-11-11 19:08:00	reference
5	the ugly	0.034	0.1016	0.1086	0.9873	0.9222	0.9328	2006-11-12 15:35:54	reference
6	vn3	0.0634	0.0744	0.1095	0.9464	0.9267	0.8949	2007-07-27 04:56:02	Vladimir Nikulin
7	cross-indexing-prior-1	0.0434	0.1284	0.1099	0.9835	0.9183	0.9308	2007-01-25 16:50:26	Juha Reunanen
8	cross-indexing-prior-1a	0.0497	0.0682	0.11	0.9782	0.9712	0.9312	2007-02-14 20:25:07	Juha Reunanen
9	Doubleboost	0.0411	0.0429	0.1114	0.9586	0.9605	0.8896	2007-02-20 14:30:27	Roman Lutz
10	LogitBoost with trees	0.0585	0.1056	0.1115	0.974	0.9323	0.9303	2006-10-10 18:05:48	Roman Lutz
11	LogitBoost with trees	0.0415	0.0432	0.1117	0.9583	0.9598	0.8892	2007-02-20 13:47:47	Roman Lutz
12	cross-indexing-prior-3	0.0415	0.049	0.1119	0.9846	0.9811	0.932	2007-06-10 01:14:52	Juha Reunanen
13	LogitBoost with trees	0.0585	0.1076	0.1121	0.974	0.9298	0.9301	2006-10-09 13:42:41	Roman Lutz
14	the good	0.0384	0.1088	0.1125	0.9827	0.9182	0.9299	2006-11-12 15:59:31	reference
15	SVM+GbO+trees	0.0382	0.0958	0.1139	0.9601	0.9086	0.882	2006-11-15 00:31:02	Vladimir Nikulin
16	cross-indexing-prior-2	0.0415	0.049	0.1141	0.9846	0.9811	0.9315	2007-02-25 21:42:16	Juha Reunanen
17	serate quadratic lssvm	0.0361	0.1219	0.1142	0.9841	0.9158	0.9281	2006-10-20 10:44:13	reference
18	out1-fs-nored-val (Intel final 1)	0.0698	0.0616	0.1142	0.9313	0.9406	0.8859	2007-07-31 02:14:16	IDEAL, Intel
19	vn1	0.0723	0.0721	0.1145	0.9517	0.9469	0.8987	2007-07-27 04:50:52	Vladimir Nikulin
20	cubic lssvm	0.0365	0.1057	0.1145	0.9842	0.9221	0.9301	2006-10-11 20:15:41	reference

http://www.agnostic.inf.ethz.ch/results.php

Some results in benchmark data

- Comparison of PSMS and pattern search

ID	Data set	PATSMS test-BER	PSMS test-BER	PATSMS CV-BER	PSMS CV-BER
1	Breast-cancer	36.98 ± 0.08	**33.59 ± 0.12**	**32.64 ± 0.06**	32.96 ± 0.01
2	Diabetes	26.07 ± 0.03	**25.37 ± 0.02**	**25.39 ± 0.02**	26.48 ± 0.05
3	Flare-solar	32.87 ± 0.02	**32.65 ± 0.01**	**32.69 ± 0.01**	33.13 ± 0.01
4	German	28.65 ± 0.02	**28.28 ± 0.02**	**31.00 ± 0.00**	31.02 ± 0.00
5	Heart	19.50 ± 0.19	**17.35 ± 0.06**	**16.96 ± 0.07**	19.93 ± 0.03
6	Image	3.58 ± 0.01	**2.50 ± 0.01**	**11.54 ± 0.10**	15.88 ± 0.04
7	Splice	13.94 ± 0.99	**9.46 ± 0.25**	**18.01 ± 0.05**	19.15 ± 0.07
8	Thyroid	10.84 ± 0.39	**5.98 ± 0.06**	**11.15 ± 0.20**	15.49 ± 0.12
9	Titanic	29.94 ± 0.00	**29.60 ± 0.00**	**27.19 ± 0.13**	27.32 ± 0.13

H. J. Escalante, E. Sucar, M. Montes. **Particle Swarm Model Selection,** *In Journal of Machine Learning Research, 10(Feb):405--440, 2009.*

Some results in benchmark data

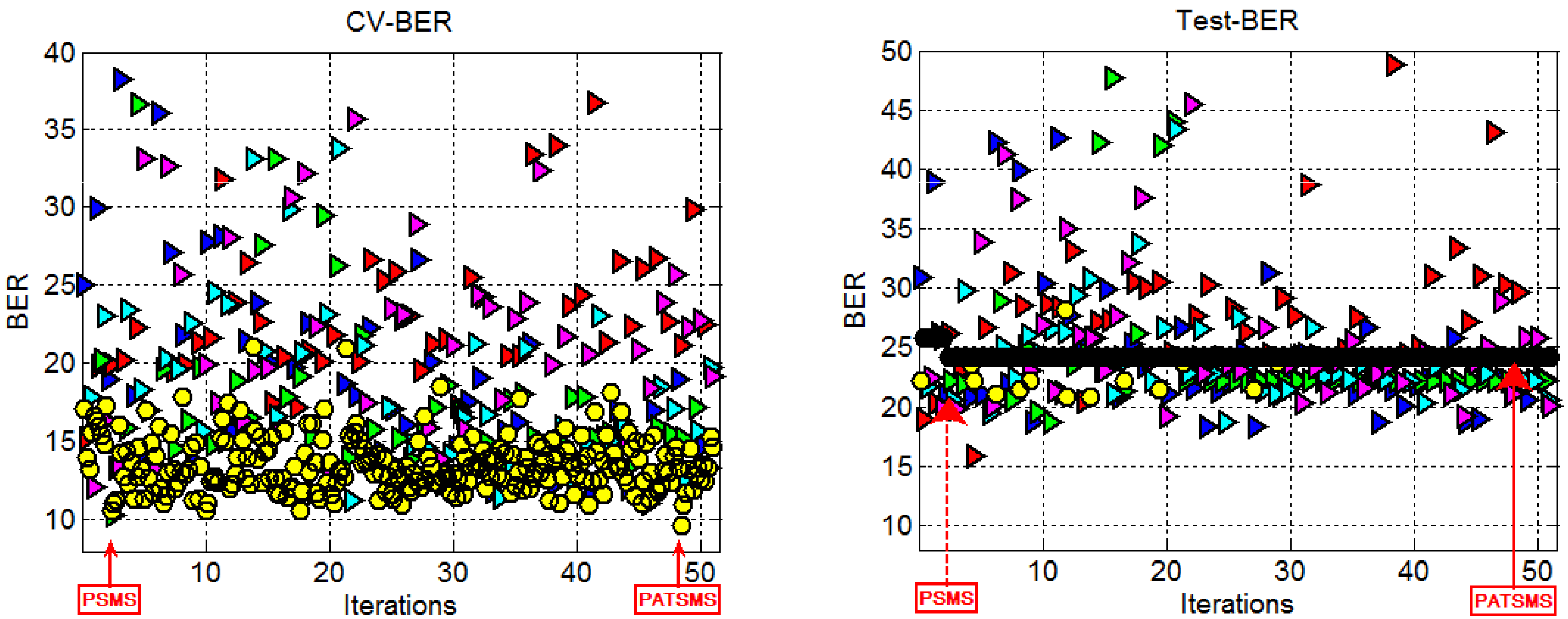
- Comparison of PSMS and pattern search

ID	Data set	PATSMS test-BER	PSMS test-BER	PATSMS CV-BER	PSMS CV-BER
1	Breast-cancer	36.98 ± 0.08	**33.59 ± 0.12**	**32.64 ± 0.06**	32.96 ± 0.01
2	Diabetes	26.07 ± 0.03	**25.37 ± 0.02**	**25.39 ± 0.02**	26.48 ± 0.05
3	Flare-solar	32.87 ± 0.02	**32.65 ± 0.01**	**32.69 ± 0.01**	33.13 ± 0.01
4	German	28.65 ± 0.02	**28.28 ± 0.02**	**31.00 ± 0.00**	31.02 ± 0.00
5	Heart	19.50 ± 0.19	**17.35 ± 0.06**	**16.96 ± 0.07**	19.93 ± 0.03
6	Image	3.58 ± 0.01	**2.50 ± 0.01**	**11.54 ± 0.10**	15.88 ± 0.04
7	Splice	13.94 ± 0.99	**9.46 ± 0.25**	**18.01 ± 0.05**	19.15 ± 0.07
8	Thyroid	10.84 ± 0.39	**5.98 ± 0.06**	**11.15 ± 0.20**	15.49 ± 0.12
9	Titanic	29.94 ± 0.00	**29.60 ± 0.00**	**27.19 ± 0.13**	27.32 ± 0.13

H. J. Escalante, E. Sucar, M. Montes. **Particle Swarm Model Selection,** *In Journal of Machine Learning Research, 10(Feb):405--440, 2009.*

Some results in benchmark data

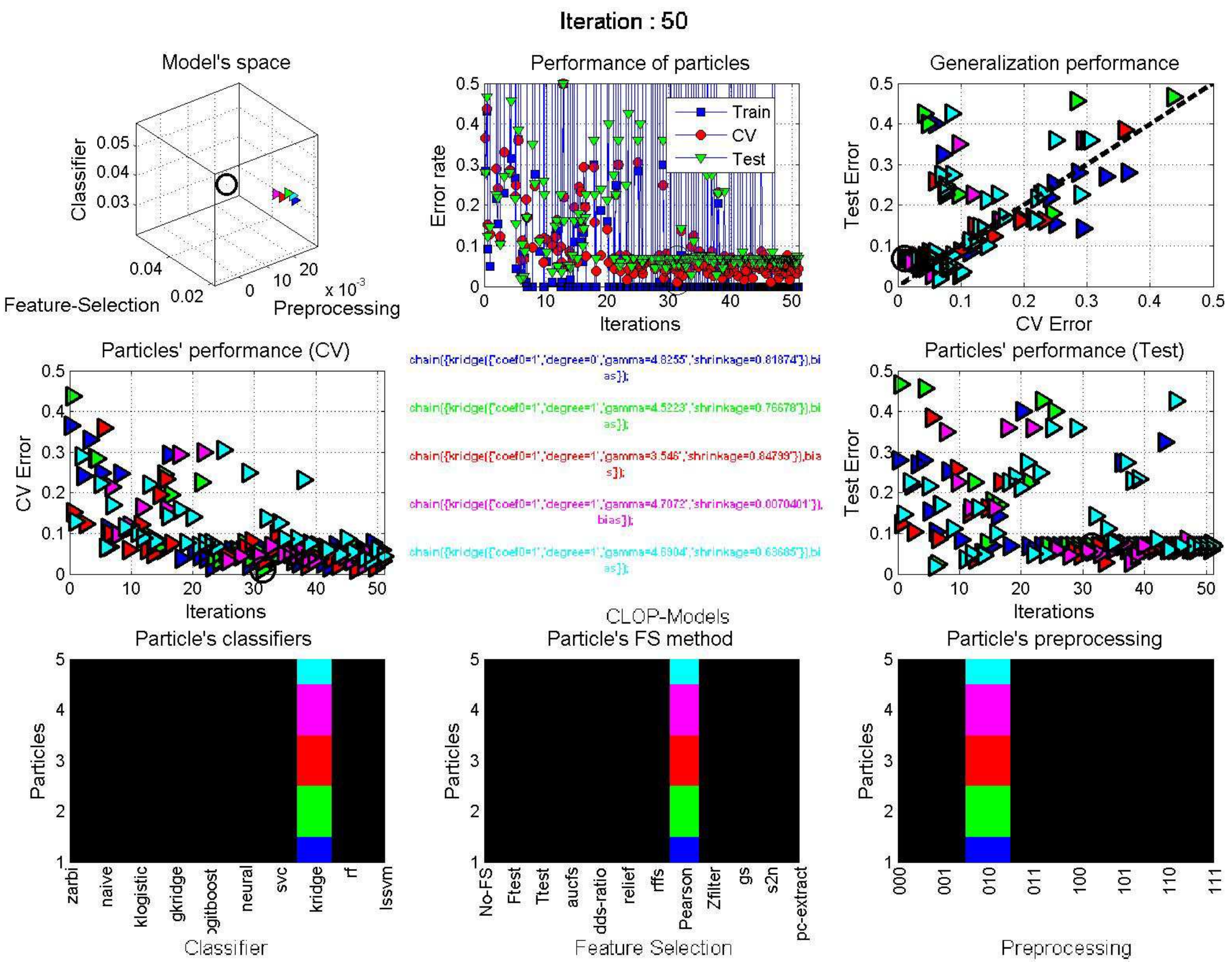
- Comparison of PSMS and pattern search



H. J. Escalante, E. Sucar, M. Montes. **Particle Swarm Model Selection,** *In Journal of Machine Learning Research, 10(Feb):405--440, 2009.*

PSMS: Interactive demo

http://clopinet.com/CLOP



Isabelle Guyon, Amir Saffari, Hugo Jair Escalante, Gokan Bakir, and Gavin Cawley, **CLOP: a Matlab Learning Object Package.** *NIPS 2007 Demonstrations, Vancouver, British Columbia, Canada 2007.*

Other applications of PSMS/EPSMS

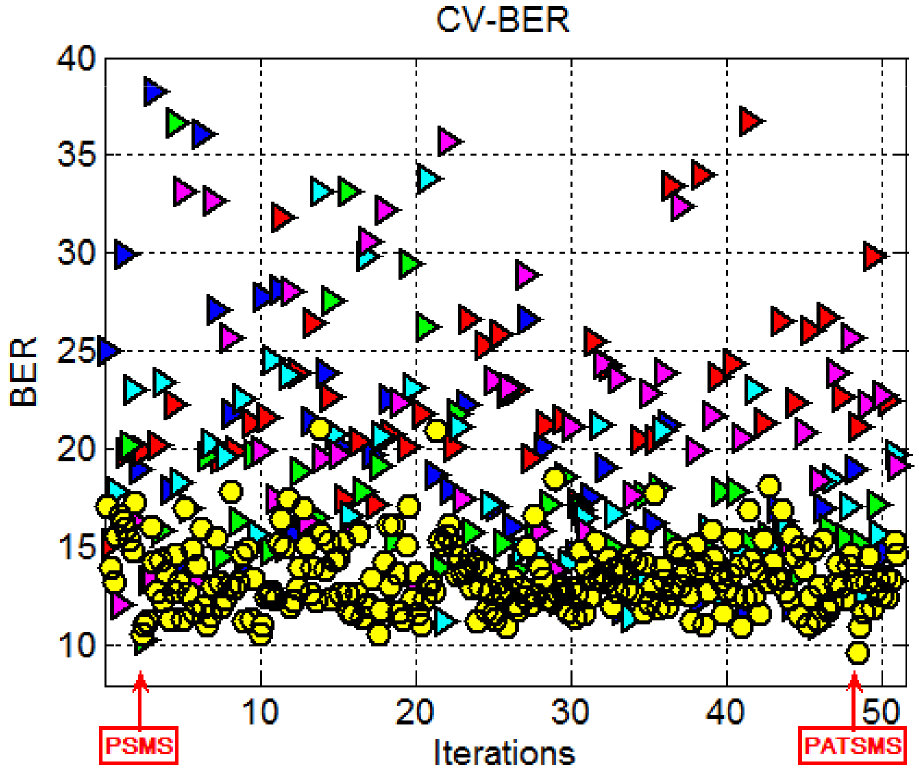
- Successful:
  - Acute leukemia classification
  - Authorship verification (Spanish/English)
  - Authorship attribution
  - Region labeling
  - ML Challenges

- Not successful:
  - Review recommendation (14 features)
  - Region labeling (~90 classes)
  - Sentiment analysis on speech signals (high p – small n)
  - Plagiarism detection (a few samples)
  - ML Challenges

ENSEMBLE PSMS

Ensemble PSMS

- Many models are evaluated during the search process of PSMS; although a single model is selected



Ensemble PSMS

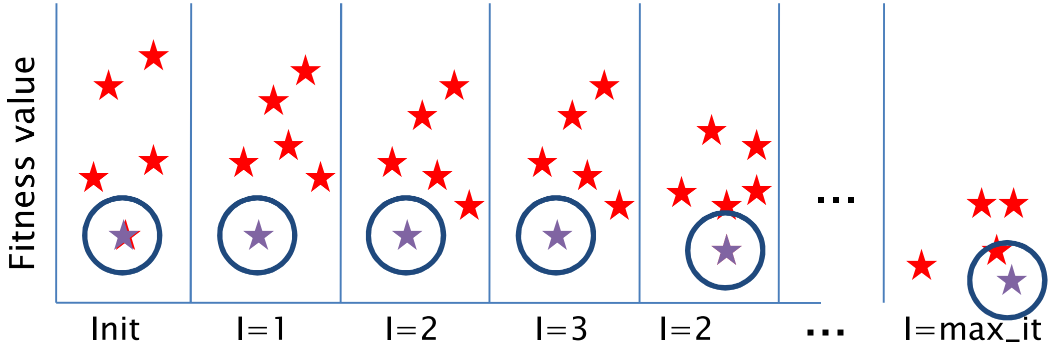
- **Idea:** taking advantage of the large number of models that are evaluated during the search for building ensemble classifiers
- **Problem:** How to select the partial solutions from PSMS so that they are accurate and diverse to each other
- **Motivation:** The success of ensemble classifiers depends mainly in two key aspects of individual models: ***Accuracy and diversity***

Ensemble PSMS

- How to select potential models for building ensembles?
  - **BS:** store the global best model in each iteration
  - **BI:** the best model in each iteration
  - **SE:** combine the outputs of the final swarm

- How to fuse the outputs of the selected models?
  - Simple (un-weighted) voting

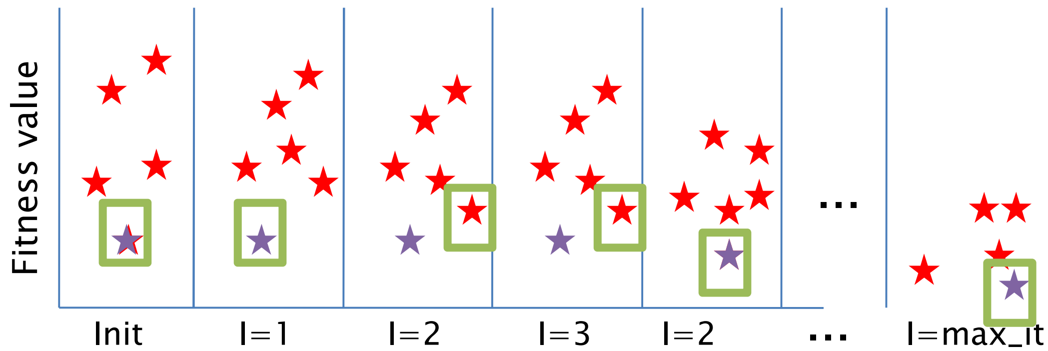
$$g(E) = \frac{1}{L}\sum_{l=1}^{L} f_l(\mathbf{p})$$



Ensemble PSMS

- How to select potential models for building ensembles?
  - **BS:** store the global best model in each iteration
  - **BI:** the best model in each iteration
  - **SE:** combine the outputs of the final swarm
- How to fuse the outputs of the selected models?
  - Simple (un-weighted) voting

$$g(E) = \frac{1}{L}\sum_{l=1}^{L} f_l(\mathbf{p})$$

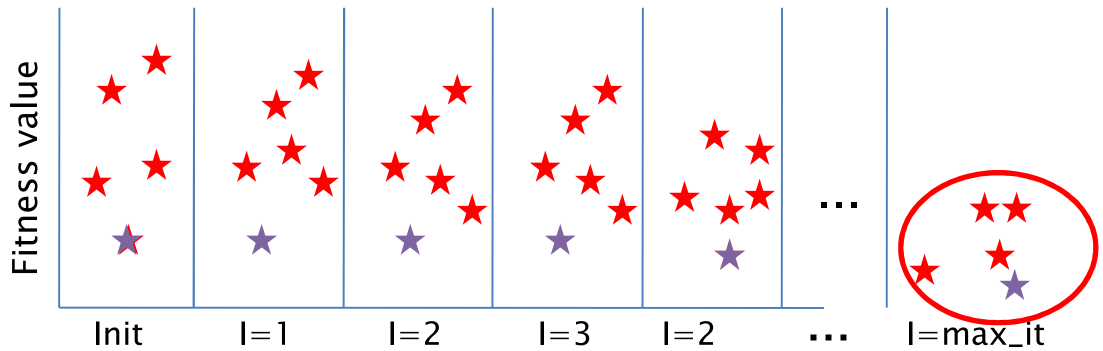


Ensemble PSMS

- How to select potential models for building ensembles?
  - **BS:** store the global best model in each iteration
  - **BI:** the best model in each iteration
  - **SE:** combine the outputs of the final swarm

- How to fuse the outputs of the selected models?
  - Simple (un-weighted) voting

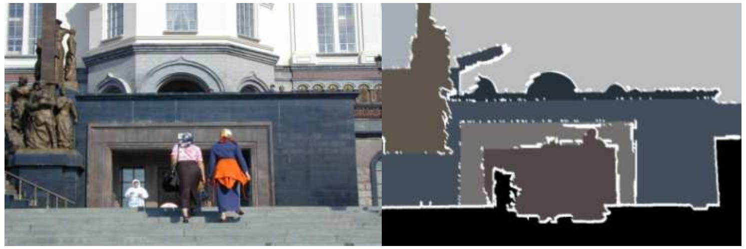
$$g(E) = \frac{1}{L}\sum_{l=1}^{L} f_l(\mathbf{p})$$



Experimental results

- Data:
  - 9 Benchmark machine learning data sets (binary classification)
  - 1 Object recognition data set (multiclass, 10 classes)

ID	Data set	Training	Testing	Features
1	Breast-cancer	200	77	9
2	Diabetes	468	300	8
3	Flare solar	666	400	9
4	German	700	300	20
5	Heart	170	100	13
6	Image	1300	1010	20
7	Splice	1000	2175	60
8	Thyroid	140	75	5
9	Titanic	150	2051	3
OR	*SCEF*	*2378*	*3300*	*50*



H. J. Escalante, M. Montes, E. Sucar. **Ensemble Particle Swarm Model Selection.** Proceedings of the International Joint Conference on Neural Networks (IJCNN2010 – WCCI2010), pp. 1814--1821, IEEE,, 2010 [**Best Student Paper Award**].

Experimental results

- Evaluation:
  - Average of area under the ROC curve (performance)
  - Coincident failure diversity (ensemble diversity)

$$CFD = \begin{cases} \dfrac{1}{1-p_0} \displaystyle\sum_{r=1}^{L} \dfrac{L-r}{L-1} p_r & \text{If } p_0 < 1 \\ 0 & \text{If } p_0 = 1 \end{cases}$$

W. Wang. **Some fundamental issues in ensemble methods.** Proc. Of IJCNN07, pp. 2244–2251, IEEE, July 2007.

Experimental results: performance

- **Benchmark data sets:** better performance is obtained by ensemble methods

ID	PSMS	EPSMS-BS	EPSMS-SE	EPSMS-BI
1	72.03±2.24	73.40±0.78	74.05±0.91	**74.35±0.49**
2	82.11±1.29	82.60±1.52	74.07±13.7	**83.42±0.46**
3	68.81±4.31	69.38±4.53	70.13±7.48	**72.16±1.42**
4	73.92±1.23	73.84±1.53	74.70±0.72	**74.77±0.69**
5	85.55±5.48	87.40±2.01	87.07±0.75	**88.36±0.88**
6	97.21±3.15	98.85±1.45	95.27±3.04	**99.58±0.33**
7	97.26±0.55	98.02±0.64	96.99±1.21	**98.84±0.26**
8	96.00±4.75	98.18±0.94	97.29±1.54	**99.22±0.45**
9	73.24±1.16	73.50±0.95	**75.37±1.05**	74.40±0.91
Avg.	82.90±2.68	83.91±1.59	82.77±3.38	**85.01±0.65**

Average accuracy over 10-trials of PSMS and EPSMS in benchmark data

H. J. Escalante, M. Montes, E. Sucar. **Ensemble Particle Swarm Model Selection.** Proceedings of the International Joint Conference on Neural Networks (IJCNN2010 – WCCI2010), pp. 1814--1821, IEEE,, 2010 [**Best Student Paper Award**].

Experimental results: Diversity of ensemble

- Diversity results

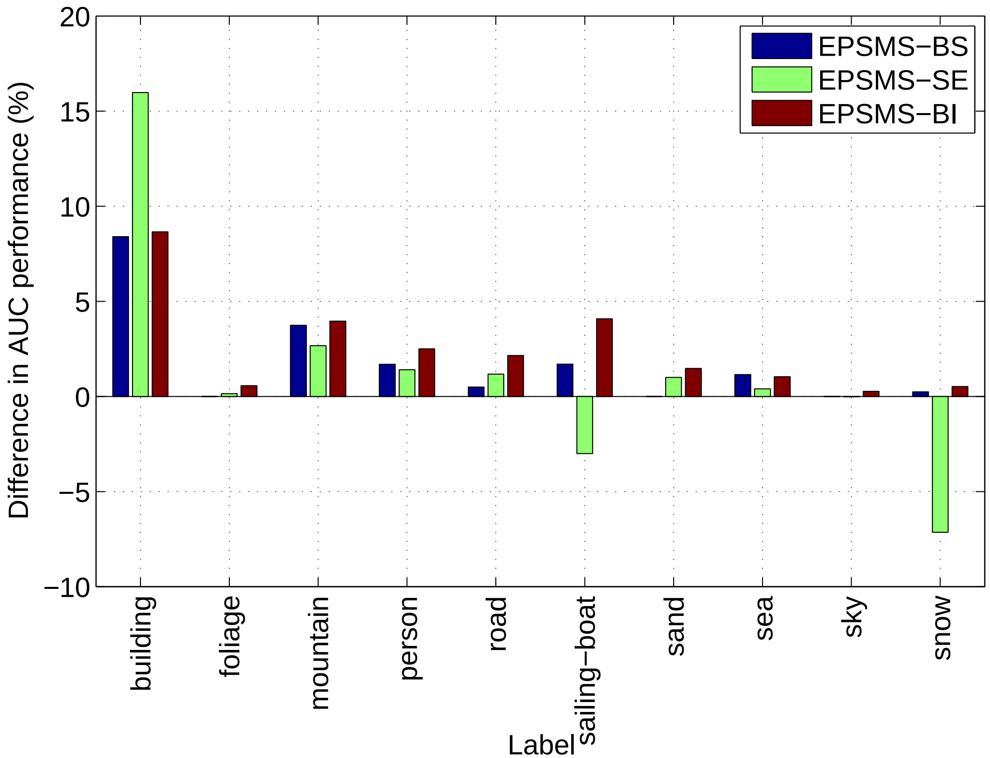
ID	EPSMS-BS	EPSMS-SE	EPSMS-BI
1	0.2055±0.1498	**0.5422±0.0550**	0.5017±0.1149
2	0.3547±0.1711	**0.6241±0.0619**	0.5081±0.0728
3	0.1295±0.1704	**0.4208±0.1357**	0.4012±0.1071
4	0.3019±0.1732	**0.5159±0.0596**	0.4296±0.0490
5	0.2733±0.1714	**0.5993±0.0925**	0.5647±0.0655
6	0.7801±0.0818	0.7555±0.0524	**0.8427±0.0408**
7	0.5427±0.3230	0.7807±0.0585	**0.8050±0.0294**
8	0.6933±0.1558	0.8173±0.0626	**0.8514±0.0403**
9	0.7473±0.0089	0.7473±0.0089	0.7473±0.0089
Avg.	0.4476±0.1562	**0.6448±0.0603**	0.6280±0.0588

EPSMS-SE models are more diverse

H. J. Escalante, M. Montes, E. Sucar. **Ensemble Particle Swarm Model Selection.** Proceedings of the International Joint Conference on Neural Networks (IJCNN2010 – WCCI2010), pp. 1814--1821, IEEE,, 2010 [**Best Student Paper Award**].

Experimental results: region labeling

ID	PSMS	EPSMS-BS	EPSMS-SE	EPSMS-BI
AUC	91.53±6.8	93.27±5.6	92.79±7.4	**94.05±5.3**
MCC	69.58%	76.59%	79.13%	**81.49%**

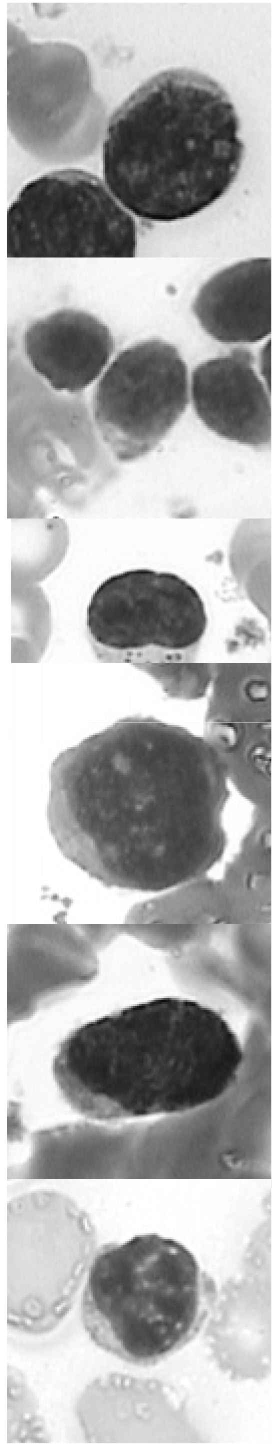


Per-concept improvement of EPSMS variants over straight PSMS

EPSMS for acute leukemia classification

- **Acute leukemia:** a malignant disease that affects a considerable portion of the world population

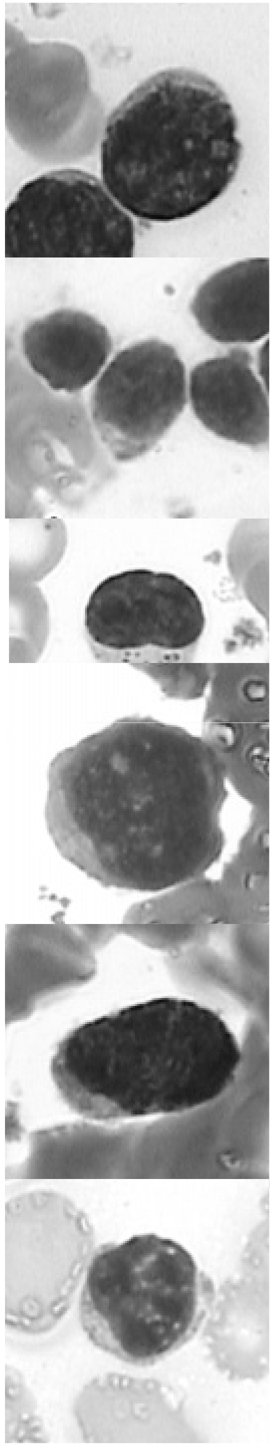
- There are different types and subtypes of acute leukemia, requiring different treatments.



EPSMS for acute leukemia classification

- Different types/subtypes:
  - Acute Lymphocytic Leukemia (ALL): L1, L2, L3
  - Acute Myelogenous Leukemia (AML): M0, M1, M2, M3, M4, M5, M6, M7

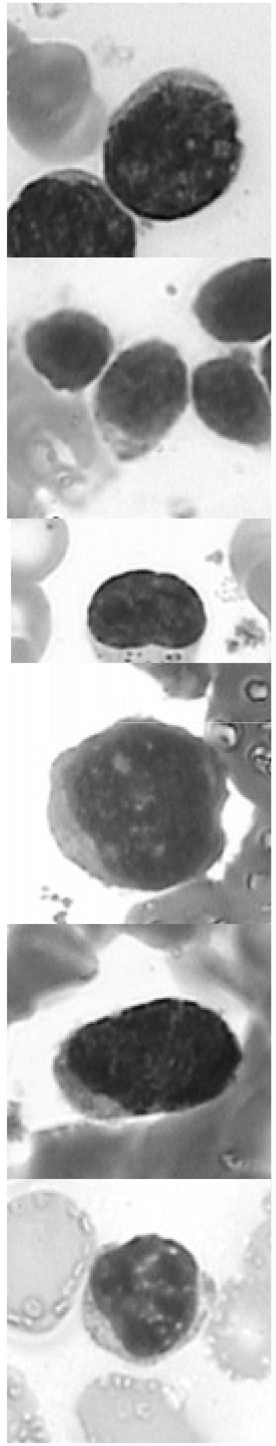
- Considered tasks:
  - Binary
    - ALL vs AML
    - L1 vs L2
    - M2 vs M3, M5, M3 vs. M2, M5, M3 vs M2, M5,
  - Multiclass
    - M1 vs M2 vs M3
    - L1 vs L2 vs M1 vs M2 vs M3



EPSMS for acute leukemia classification

- Despite the fact that there are advanced and precise methods to identify leukemia types, they are very expensive and unavailable in most of hospitals of third world countries

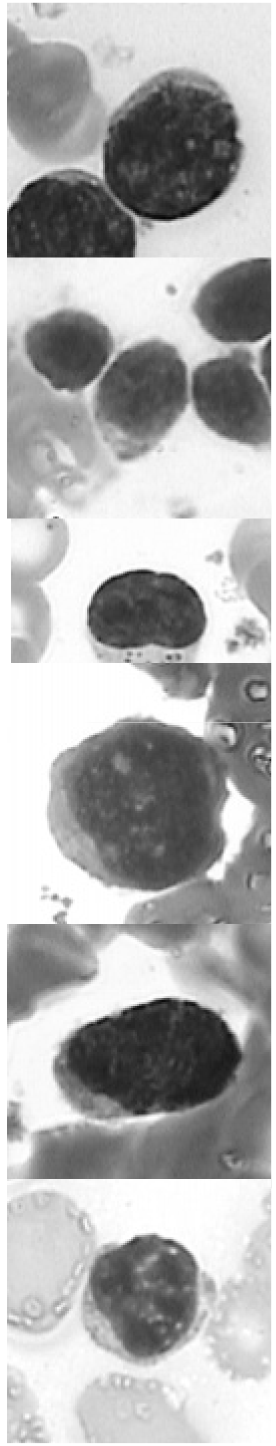
- A Cheaper alternative: *morphological acute leukemia classification from bone marrow cell images*



EPSMS for acute leukemia classification

- Morphological classification:
  - Image registration
  - Image segmentation
  - Feature extraction (Morphological, statistical, texture)
  - Classifier construction

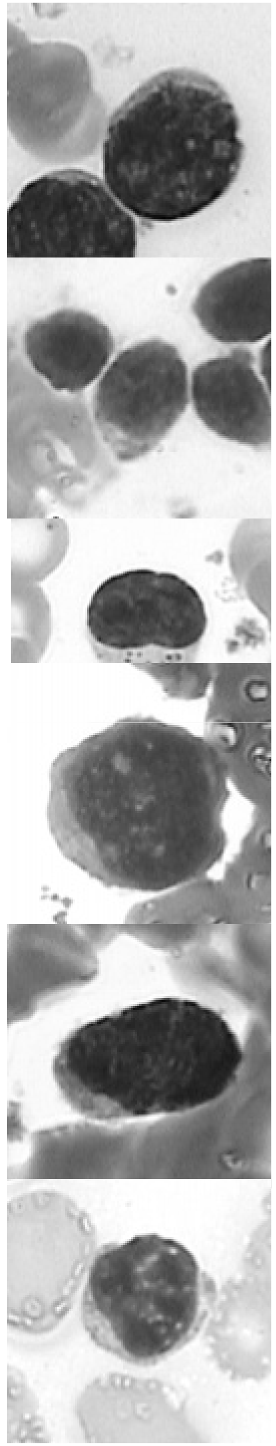
J. A. Gonzalez et al. **Leukemia Identification from Bone Marrow Cells Images using a Machine Vision and Data Mining Strategy**, Intelligent Data Analysis, Vol 15(3):443—462, 2011.



EPSMS for acute leukemia classification

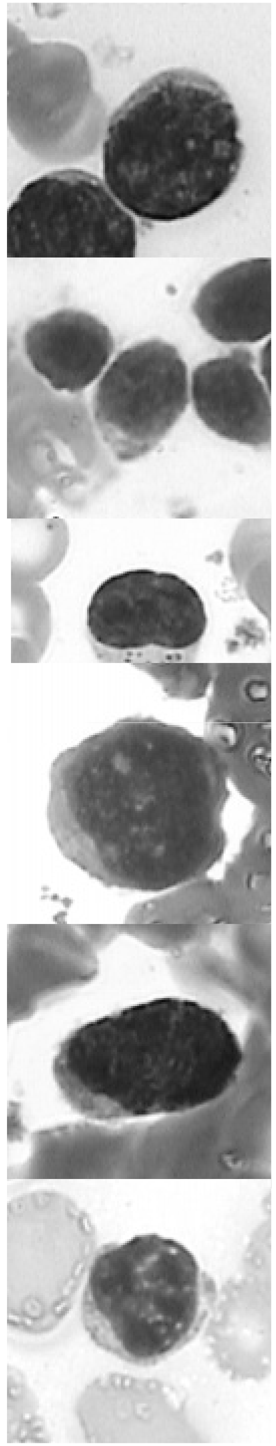
- In previous work either:
  - the same classification model has been considered for different problems, or
  - Models have been manually selected by trial and error
- **Proposal:** Using EPSMS for automatically selecting specific classifiers for the different tasks

H. J. Escalante, J. A. Gonzalez, M. Montes-y-Gomez, P. Gómez, C. Reta, C. A. Reyes, A. Rosales. **Acute Leukemia Classiffcation with Ensemble Particle Swarm Model Selection,** Artificial Intelligence in Medicine, Available online 15 April 2012.



EPSMS for acute leukemia classification

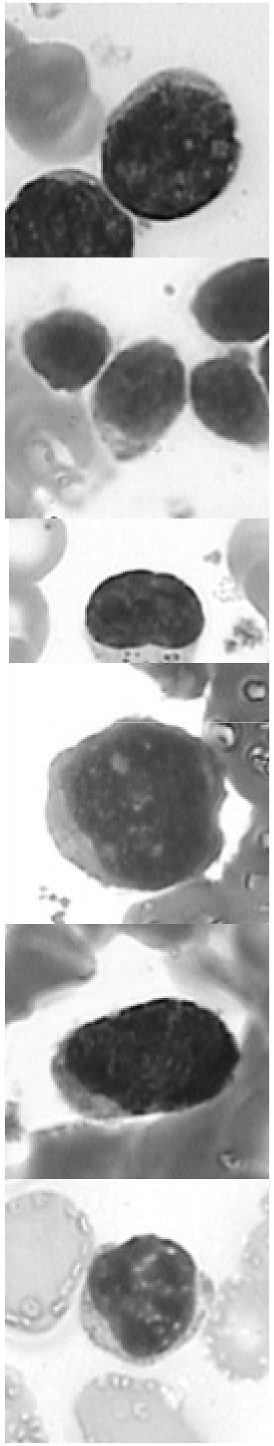
- Experiments were performed with real data collected by a Mexican health institution (IMSS), using 10-fold CV



H. J. Escalante, J. A. Gonzalez, M. Montes-y-Gomez, P. Gómez, C. Reta, C. A. Reyes, A. Rosales. **Acute Leukemia Classiffcation with Ensemble Particle Swarm Model Selection,** Artificial Intelligence in Medicine, Available online 15 April 2012.

EPSMS for acute leukemia classification

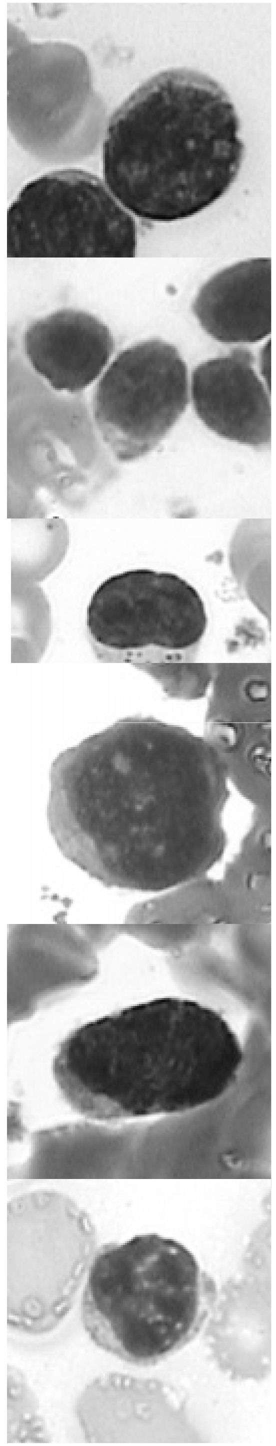
ID	Features	Region	Reference	Baseline	PSMS	E1	E2	E3
ALL vs AML								
1	A	C	89.26	93.63	93.87	**95.14**	**95.14**	92.43
2		N-C	90.65	94.37	95.16	96.06	**96.16**	92.58
3	B	C	79.06	84.38	83.41	**83.43**	83.31	77.82
4		N-C	**81.27**	80.59	79.74	80.32	80.32	76.08
5	C	C	89.92	93.04	93.15	**95.04**	94.87	91.95
6		N-C	92.17	93.17	95.66	96.48	**96.66**	93.77
L1 vs L2								
7	A	C	81.4	91.46	87.68	88.95	88.72	**92.24**
8		N-C	90.69	91.13	88.54	92	92	**94.1**
9	B	C	76.08	79.73	79.07	79.76	79.91	**84.11**
10		N-C	83.67	80.07	74.79	86.8	87.06	**87.41**
11	C	C	82.25	87.85	87.45	89.7	89.71	**92.38**
12		N-C	88.61	88.93	88.06	90.23	90.37	**93.09**
M2 vs M3, M5								
13	A	C	80.45	94.43	91.93	**95.78**	95.13	92.72
14		N-C	95.9	95.58	**96.71**	95.7	95.51	93.01
15	B	C	71.06	73.34	72.99	80.23	**80.33**	72.1
16		N-C	78.93	81.47	79.06	**83.48**	83.44	76.69
17	C	C	84.12	93.2	**95.93**	95.67	95.7	92.85
18		N-C	**94.68**	93.54	93.65	94.24	94.36	89.45
M3 vs M2, M5								
19	A	C	78.82	89.71	88.25	92.01	**92.57**	88.72
20		N-C	87.97	86.88	91.51	**96.62**	96.55	92.28
21	B	C	78.67	74.59	72.93	74.9	75	**79.86**
22		N-C	73.91	72.34	**77.26**	76.4	76.78	76.89
23	C	C	71.01	81.91	92.74	94.02	**94.15**	92.01
24		N-C	89.85	84.25	93.06	**93.24**	92.98	90.44
M5 vs M2, M3								
25	A	C	86.64	92.08	87.32	89.53	89.93	**95.28**
26		N-C	95.52	92.87	92.94	94.54	94.69	**97.68**
27	B	C	73.14	69.87	61.77	66.4	67.73	**75.78**
28		N-C	73.32	69.3	79.24	78.08	78.12	**85.91**
29	C	C	84.98	90.86	92.29	95.85	**96.01**	95.77
30		N-C	**93.54**	92.11	88.97	92.14	92.25	93.07



EPSMS for acute leukemia classification

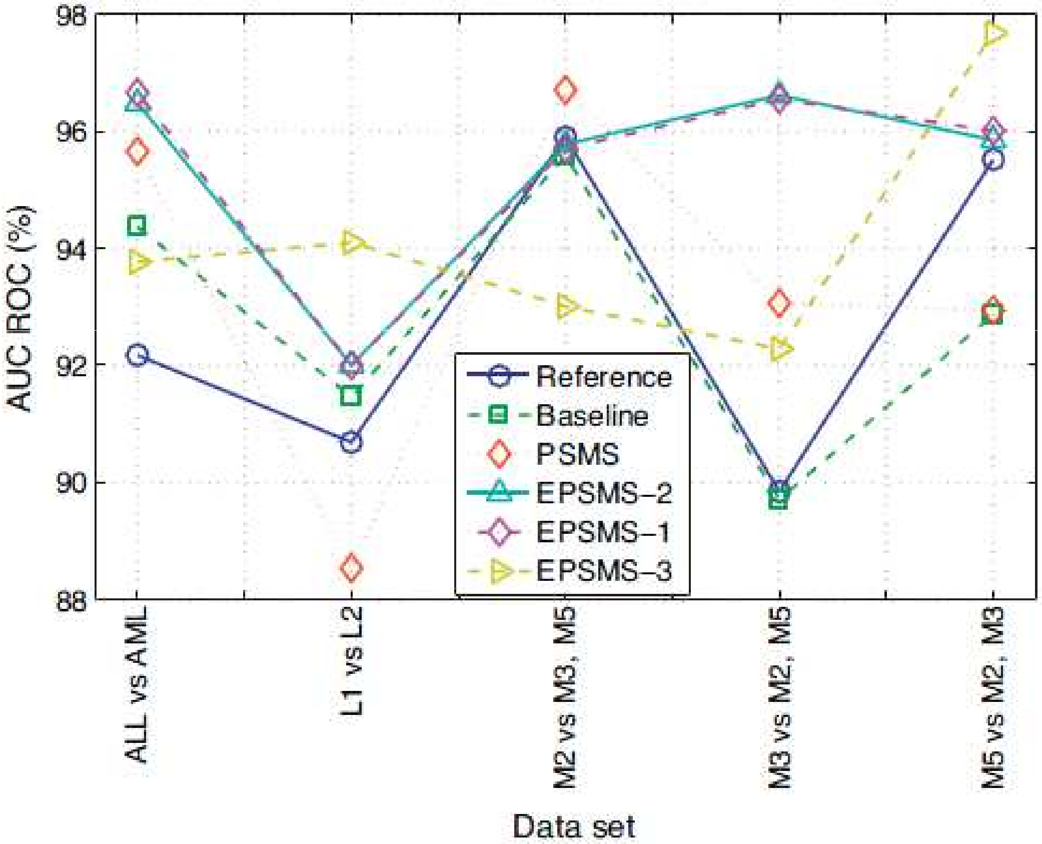
- In general ensembles generated with EPSMS outperformed previous work

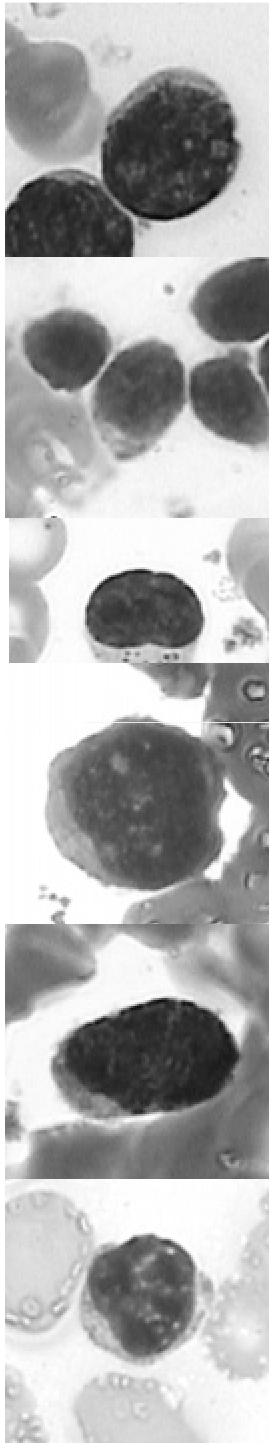
- Models selected with EPSMS were more effective than those selected with straight PSMS



H. J. Escalante, J. A. Gonzalez, M. Montes-y-Gomez, P. Gómez, C. Reta, C. A. Reyes, A. Rosales. **Acute Leukemia Classiffcation with Ensemble Particle Swarm Model Selection,** Artificial Intelligence in Medicine, Available online 15 April 2012.

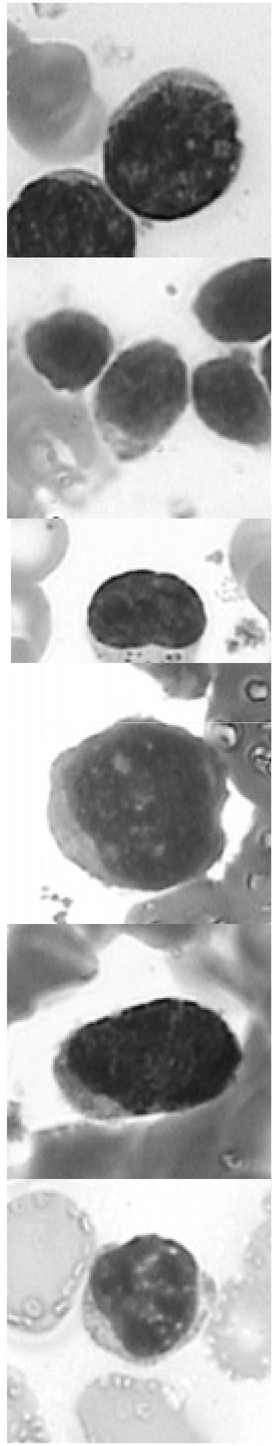
EPSMS for ALC - binary tasks





EPSMS for ALC – Multiclass

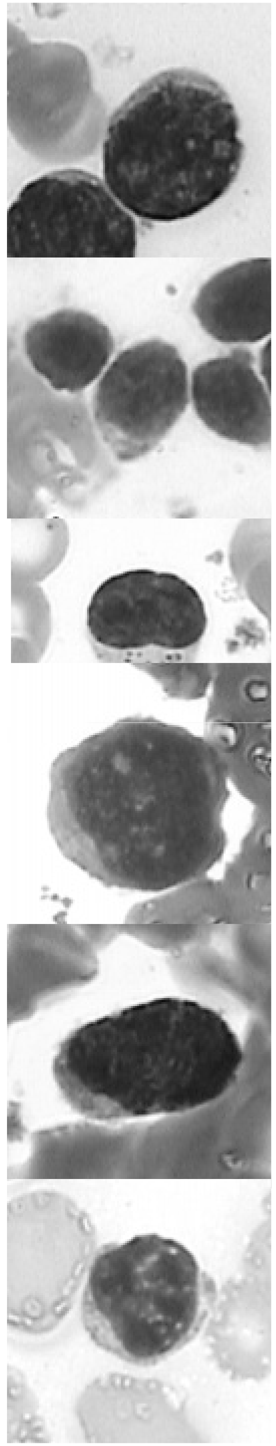
Measure	Region	Ref.	PSMS	E1	E2	E3
			M2 vs M3 vs M5			
Avg. AUC	C	78.66	92.37	**93.94**	**93.94**	93.82
Accuracy	C	66.13	**83.37**	81.32	79.76	78.79
Avg. AUC	N&C	92.80	92.36	**93.94**	93.92	93.28
Accuracy	N&C	**84.87**	81.84	81.87	82.34	79.34
			L1 vs L2 vs M2 vs M3 vs M5			
Avg. AUC	C	84.03	91.13	**93.78**	93.76	83.40
Accuracy	C	55.86	72.86	75.83	**76.06**	73.92
Avg. AUC	N&C	92.33	90.62	**94.21**	94.09	86.09
Accuracy	N&C	**77.48**	71.72	74.50	75.65	74.03



H. J. Escalante, J. A. Gonzalez, M. Montes-y-Gomez, P. Gómez, C. Reta, C. A. Reyes, A. Rosales. **Acute Leukemia Classiffcation with Ensemble Particle Swarm Model Selection,** Artificial Intelligence in Medicine, Available online 15 April 2012.

Models selected with EPSMS

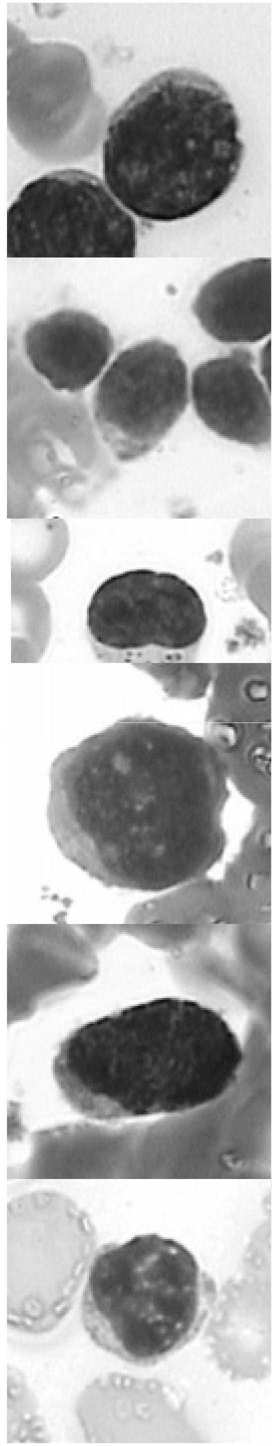
ID	P/FS	Preprocessing	Feature selection	Classification
Best avg. EPSMS				
1	FS	standardize(1), shift-scale(0)	Pearson(103)	lssvm(c=1;d=1;$\gamma = 0.4315$;sh=0.6828;b=1)
2	P	normalize(1), standardize(1), shift-scale(1)	Ftest(4)	logitboost(u=10;sh=0.33925;de=1)
3	P	normalize(1), shift-scale(1), standardize(1)	Ftest(4)	rf(u=100;m=1;b=1)
4	P	normalize(1), shift-scale(1), standardize(1)	relief(65)	lssvm(c=0;d=2;$\gamma = 2.8358$;sh=2;b=1)
5	-	normalize(1), shift-scale(1), standardize(1)	-	lssvm(c=1;d=1;$\gamma = 2.0133$;sh=0.92317;b=0)
6	P	standardize(0), shift-scale(1)	Ftest(17)	rf(u=10;m=4;b=1)



H. J. Escalante, J. A. Gonzalez, M. Montes-y-Gomez, P. Gómez, C. Reta, C. A. Reyes, A. Rosales. **Acute Leukemia Classiffcation with Ensemble Particle Swarm Model Selection,** Artificial Intelligence in Medicine, Available online 15 April 2012.

Models selected with EPSMS

Worst avg. EPSMS				
1	-	normalize(1)	-	rf(u=100;m=1;b=1)
2	-	normalize(1), shift-scale(0)	-	logitboost(u=101;1.77;d=1)
3	-	normalize(1), shift-scale(1), standardize(1)	-	logitboost(u=110;sh=2;de=2)
4	P	normalize(1), shift-scale(1)	Ftest(16)	neural(u=25;sh=1.42;b=1;e=10)
5	P	normalize(1), shift-scale(1)	gs(40)	neural(u=25;sh=1.14;b=1;e=10)
6	-	standardize(0)	-	rf(u=100;m=2;b=1)



H. J. Escalante, J. A. Gonzalez, M. Montes-y-Gomez, P. Gómez, C. Reta, C. A. Reyes, A. Rosales. **Acute Leukemia Classiffcation with Ensemble Particle Swarm Model Selection,** Artificial Intelligence in Medicine, Available online 15 April 2012.

Lessons learned

- Ensembles generated with EPSMS outperformed individual classifiers; including those selected with PSMS
- Models evaluated by PSMS are diverse to each other and accurate
- More stable predictions are obtained with the ensemble version of PSMS

Summary

- Full model selection is a broad view of the model selection problem in supervised learning
- There is an increasing interest for this type of methods
- The search space is huge and multiple minima may exist
- There is no guarantee of avoiding overfitting
- Yet, we showed evidence that it is possible to attempt to automate the cycle of design of pattern classification systems

Summary

- PSMS / EPSMS an automatic tool for the selection of classification models for any classification task

- PSMS / EPSMS has been successfully applied in several domains

- **Disclaimer:** We do not expect PSMS / EPSMS to perform well in every application it is tested, although we recommend it as a first option when building a classifier

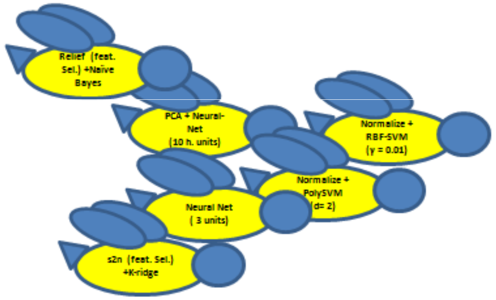
Summary

- For multiclass classification straight OVA strategies did not work

- Alternative methods that do not use the output of classifiers are a good option to explore as future work

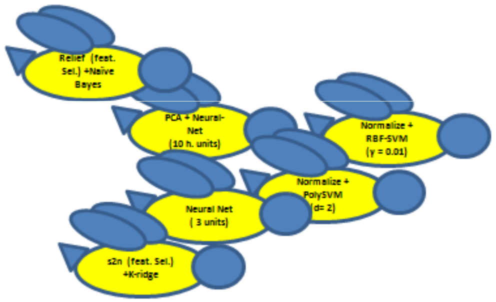
Research opportunities

- Multi-Swarm PSMS for building ensembles
- Multiclass PSMS/EPSMS:
  - Bi-level optimization (i.e., individual and multiclass performance)
  - Learning to fuse the outputs (e.g., using genetic programming)
- Meta-ensembles



Research opportunities

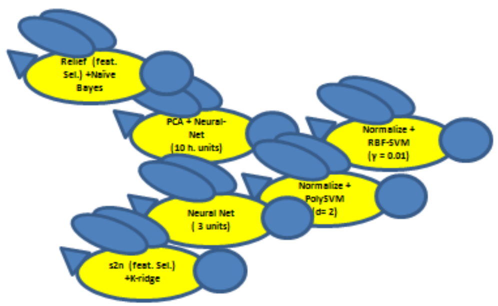
- Other search strategies for full model selection (e.g., Genetic programming, GRASP, Tabu search)



- Other toolboxes (e.g., weka)
- Meta-learning + PSMS
- Combination of different full model selection techniques

Applications

- Any classification problem where specific classifiers are required for each class



References

- Isabelle Guyon. **A Practical Guide to Model Selection.** In Jeremie Marie, editor, Machine Learning Summer School 2008,Springer Texts in Statistics, 2011. (slide from I.Guyon's)

- A. Statnikov, I. Tsamardinos, Y. Dosbayev, C.F. Aliferis. **GEMS: A System for Automated Cancer Diagnosis and Biomarker Discovery from Microarray Gene Expression Data**. International Journal of Medical Informatics, 2005 Aug;74(7-8):491-503.

- D. Gorissen, T. Dhaene, F. de Turck. **Evolutionary Model Type Selection for Global Surrogate Modeling.** *In Journal of Machine Learning Research,* 10(Jul):2039-2078, 2009

- Brazdil P., Giraud-Carrier C., Soares C., Vilalta R. **Metalearning: Applications to Data Mining.** Springer Verlag. ISBN: 978-3-540-73262-4, 2008.

- H. J. Escalante. **Towards a Particle Swarm Model Selection algorithm.** Multi-level inference workshop and model selection game, NIPS, Whistler, VA, BC, Canada, 2006.

- Isabelle Guyon, Amir Saffari, Hugo Jair Escalante, Gokan Bakir, and Gavin Cawley, **CLOP: a Matlab Learning Object Package**. NIPS 2007 Demonstrations, Vancouver, British Columbia, Canada 2007.

- H. J. Escalante, E. Sucar, M. Montes. **Particle Swarm Model Selection,** *In Journal of Machine Learning Research, 10(Feb):405--440, 2009.*

- Hugo Jair Escalante, Jesus A. Gonzalez, Manuel Montes-y-Gomez, Pilar Gómez, Carolina Reta, Carlos A. Reyes, Alejandro Rosales. **Acute Leukemia Classiffcation with Ensemble Particle Swarm Model Selection,** In press, Artificial Intelligence in Medicine, Available online 15 April 2012.

- Hugo Jair Escalante, Manuel Montes, and Luis Villaseñor. **Particle Swarm Model Selection for Authorship Verification.** LNCS 5856, pp. 563--570, Springer, 2009.


THANK YOU,
¿QUESTIONS?

INSTITUTO NACIONAL DE ASTROFÍSICA, ÓPTICA Y ELECTRÓNICA

Acerca del INAOE

INSTITUTO NACIONAL DE ASTROFÍSICA, ÓPTICA Y ELECTRÓNICA

Rankings

CONACYT 40 años

SCIMAGO LAB

SIR

SIR Global Mexico 2013 - Rank: Output 2007-2011

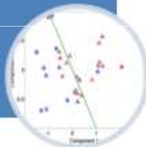
Powered by Scopus

WR	RR	CR	Organization	Sector	Country	Region	O	% IC	NI	% Q1	Spec	% Exc	% Lead	% EwL
122 ↓	3 →	1 →	Universidad Nacional Autonoma de Mexico	HE	MEX	LA	19736 ↑	39.8 ↑	0.78 →	39.7 ↑	0.54 ↓	7.26 ↓	59.71 ↓	2.95 ↓
367 ↑	11 ↓	2 →	Consejo Nacional de Ciencia y Tecnologia*	GO	MEX	LA	9390 ↑	40.32 ↑	0.76 ↓	33.79 ↑	0.66 ↓	7.4 ↓	58.46 ↓	3.07 ↓
486 →	13 →	3 →	Centro de Investigacion y de Estudios Avanzados del IPN	HE	MEX	LA	7235 ↑	39.85 ↑	1.02 ↑	36.83 ↑	0.64 →	9.88 ↑	55.01 ↓	3.19 ↓
636 ↑	20 →	4 →	Instituto Politecnico Nacional	HE	MEX	LA	5608 ↑	28.99 ↑	0.63 ↑	27.3 ↑	0.63 →	5.13 ↑	55.9 ↓	2.01 ↑
857 ↓	31 ↓	5 →	Universidad Autonoma Metropolitana	HE	MEX	LA	4033 ↑	27.3 ↑	0.66 ↑	31.61 ↑	0.6 ↑	5.91 ↓	55.64 ↓	2.53 ↓
961 ↓	36 ↓	6 →	Instituto Mexicano del Seguro Social	HL	MEX	LA	3513 ↓	20.41 ↑	0.75 ↑	22.37 ↑	0.8 ↑	5.03 ↑	60.97 ↓	1.15 ↓
1387 ↑	56 →	7 →	Universidad de Guadalajara	HE	MEX	LA	2109 ↑	33.24 ↑	0.58 ↑	27.98 ↑	0.62 ↓	3.99 ↑	53.77 ↑	1.71 ↑
1473 ↑	61 ↑	8 ↑	Universidad Autonoma de Nuevo Leon	HE	MEX	LA	1942 ↑	33.42 ↓	0.61 →	26.47 ↑	0.66 →	6.15 ↑	59.11 ↑	3.63 ↑
1478 ↑	62 ↑	9 ↑	Benemerita Universidad Autonoma de Puebla	HE	MEX	LA	1934 ↑	35.52 ↑	1.02 ↑	28.39 ↑	0.67 ↓	8.86 ↑	49.02 ↓	1.63 ↓
1569 ↓	68 ↓	10 ↓	Instituto Nacional de Astrofisica Optica y Electronica (sub)	GO	MEX	LA	1769 ↑	44.38 ↑	0.83 ↑	19.79 ↑	0.93 →	7.9 ↑	56.7 ↑	1.89 ↑
1656 ↑	72 ↑	11 ↑	Universidad de Guanajuato	HE	MEX	LA	1625 ↑	42.65 ↓	0.76 ↑	32.43 ↑	0.69 ↓	7.23 ↑	54.95 ↓	2.56 ↓
1665 ↓	73 ↓	12 ↓	Instituto Tecnologico y de Estudios Superiores de Monterrey	HE	MEX	LA	1614 ↑	40.27 ↑	0.79 →	26.21 ↑	0.73 ↓	7.34 ↑	61.21 ↓	3.55 ↓
1723 ↑	79 ↑	13 ↑	Universidad Michoacana de San Nicolas de Hidalgo	HE	MEX	LA	1510 ↑	34.04 ↑	0.89 ↑	31.06 ↑	0.7 ↓	8.64 ↑	58.08 ↑	3.01 ↓
1724 ↑	80 ↑	14 ↑	Universidad Autonoma de San Luis Potosi	HE	MEX	LA	1508 ↑	40.78 ↑	1.1 ↑	39.85 ↑	0.66 ↓	11 ↑	52.06 ↓	2.3 ↓
1742 ↓	83 ↓	15 ↓	Instituto Nacional de Ciencias Medicas y Nutricion Salvador Zubiran	HL	MEX	LA	1486 ↑	28.06 ↑	0.99 ↑	37.62 ↑	0.87 ↑	7.77 ↑	54.91 ↓	2.63 ↑

INSTITUTO NACIONAL DE ASTROFÍSICA, ÓPTICA Y ELECTRÓNICA

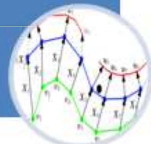
Coordinación de Ciencias Computacinoales

Aprendizaje Computacional y Reconocimiento de patrones



Cómputo reconfigurable y alto rendimiento

Cómputo y procesamiento ubícuo



Procesamiento de Bioseñales y computación médica

Robótica

Tecnologías del lenguaje

Visión por computadora