Computational Intelligence: Current State and Challenges

Pedro Larrañaga

pedro.larranaga@fi.upm.es
http://cig.fi.upm.es/

Computational Intelligence Group
Departamento de Inteligencia Artificial
Universidad Politécnica de Madrid

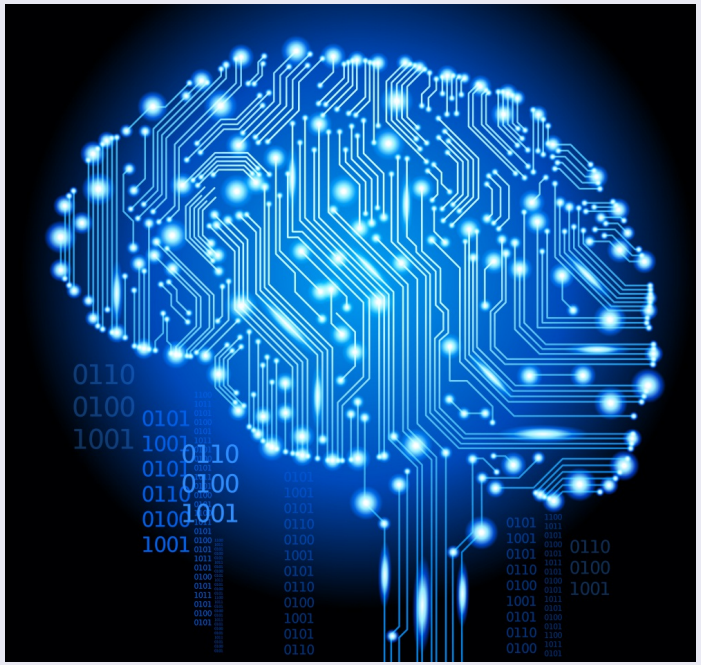
EVIA 2014 · A Coruña, September 3, 2014

Outline

Outline

Computational Intelligence

Towards a Definition

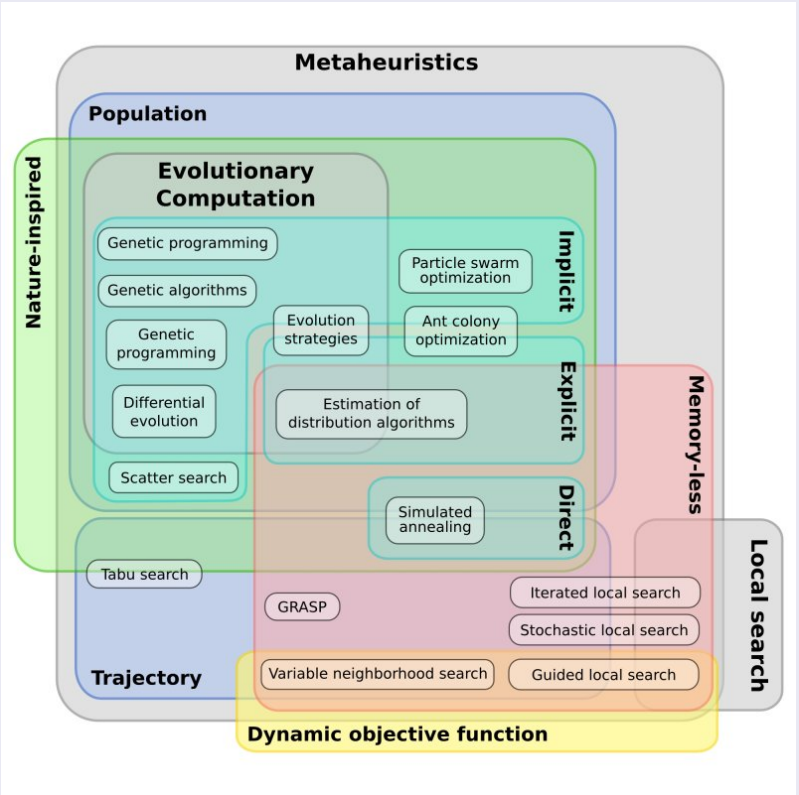


- Set of nature-inspired computational methodologies and approaches
- Address complex real-world problems
- Traditional approaches, i.e., explicit statistical modeling, are ineffective or infeasible
- Many such real-life problems are not considered to be well-posed problems mathematically

Computational Intelligence

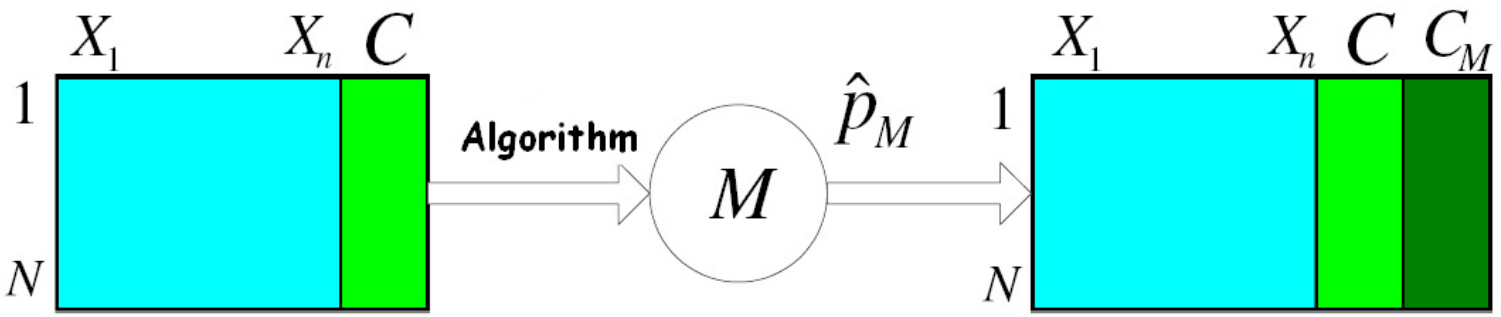
Machine Learning + Metaheuristic-based Optimization

- Machine learning as the field of study that gives computers the ability to learn without being explicitly programmed (Samuel, 1959): supervised classification, clustering, associations
- Metaheuristic-based optimization



Computational Intelligence. Machine Learning

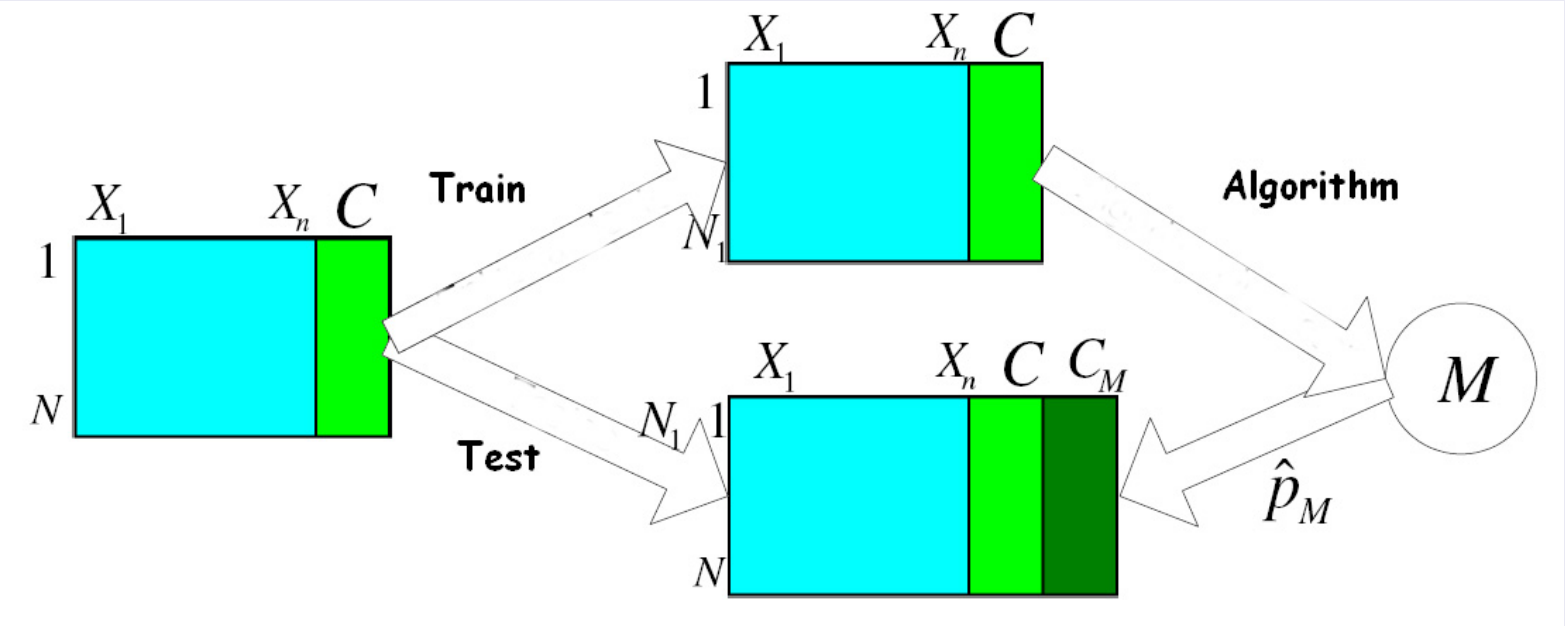
Supervised classification. Estimation methods. No honest



$$\hat{p}_M = \frac{1}{N} \sum_{i=1}^{N} \delta(c^{(i)} = c_M^{(i)})$$

Computational Intelligence. Machine Learning

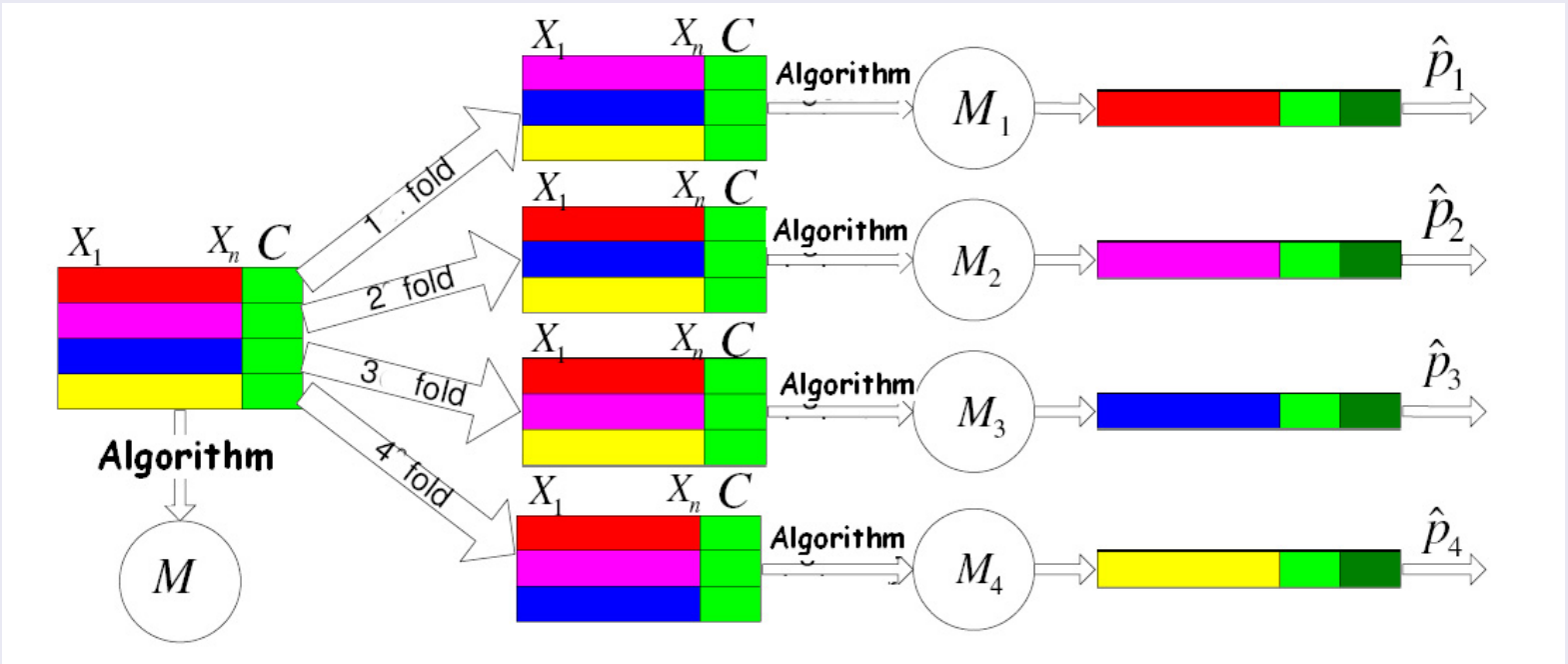
Supervised classification. Estimation methods. Train and test



$$\hat{p}_M = \frac{1}{N - N_1} \sum_{i=1}^{N-N_1} \delta(c^{(N_1+i)} = c_M^{(N_1+i)})$$

Computational Intelligence. Machine Learning

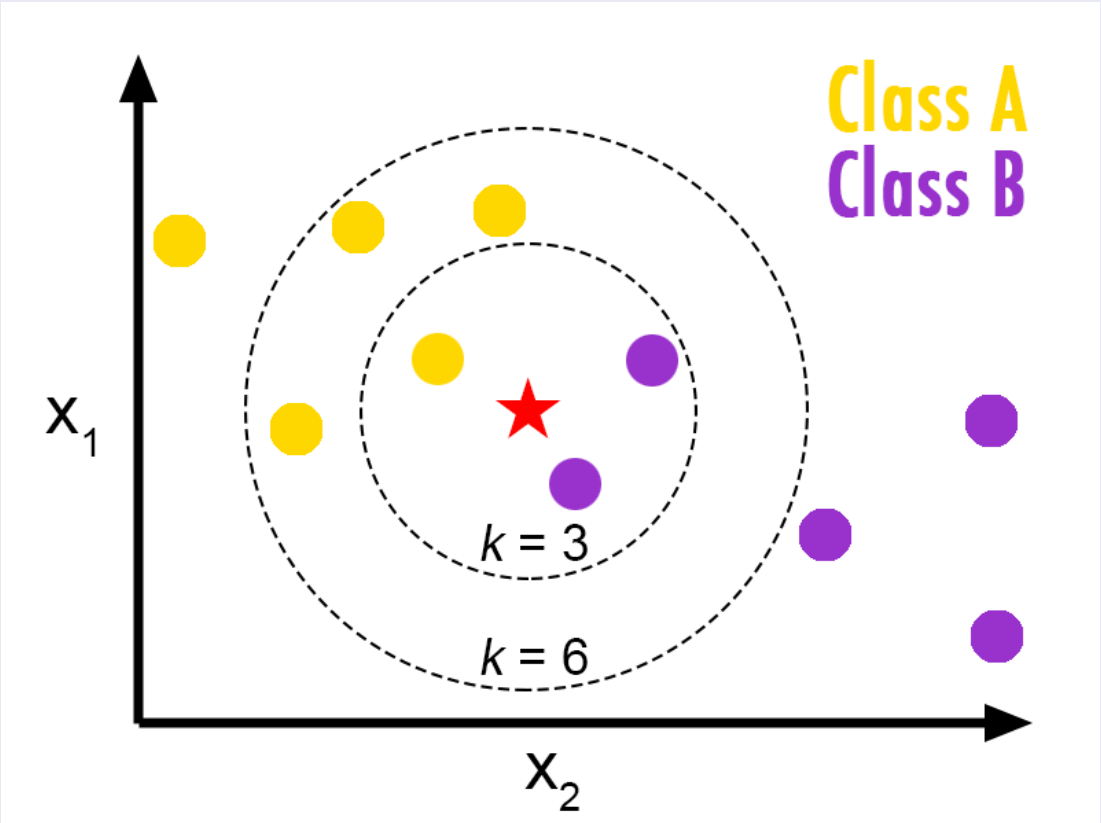
Supervised classification. Estimation methods. k–fold cross validation



$$\hat{p}_M = \frac{1}{k} \sum_{i=1}^{k} \hat{p}_i$$

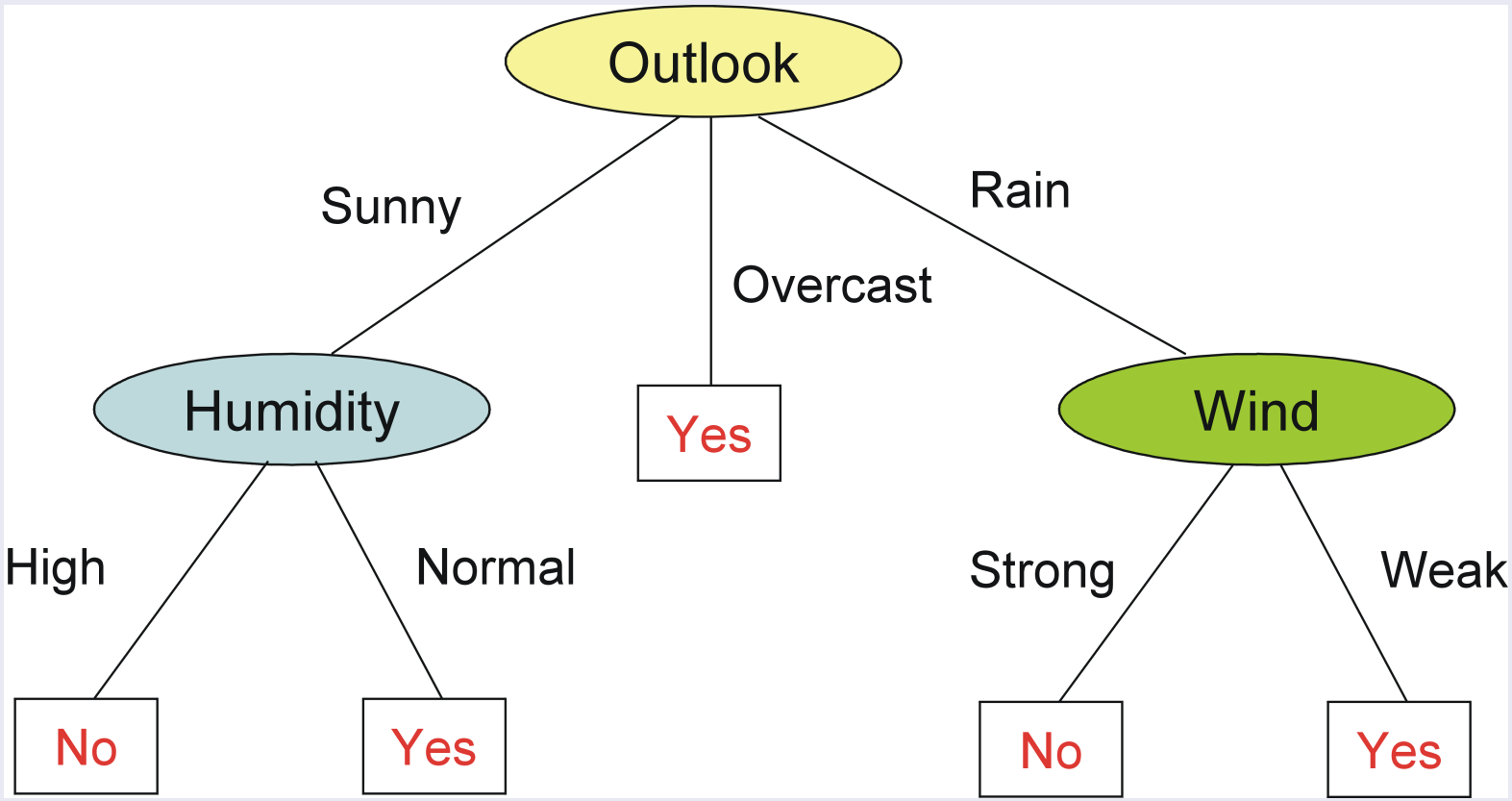
Computational Intelligence. Machine Learning

Supervised classification. k-NEAREST NEIGHBORS



Computational Intelligence. Machine Learning

Supervised classification. CLASSIFICATION TREE



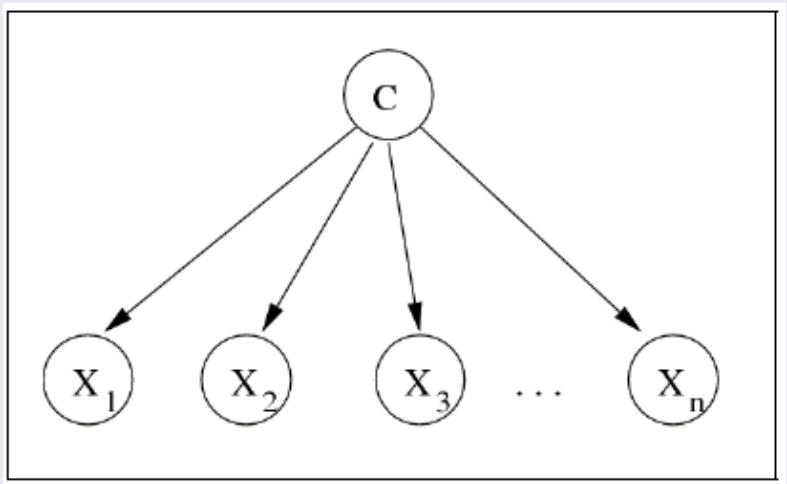
Computational Intelligence. Machine Learning

Supervised classification. NAIVE BAYES

Predictor variables are conditionally independent given C

$$P(c|x_1,...,x_n) \propto P(C=c)\prod_{i=1}^{n} P(X_i=x_i|C=c)$$

$$\Rightarrow \qquad c^* = arg \;\; max_c P(C=c)\prod_{i=1}^{n} P(X_i=x_i|C=c)$$

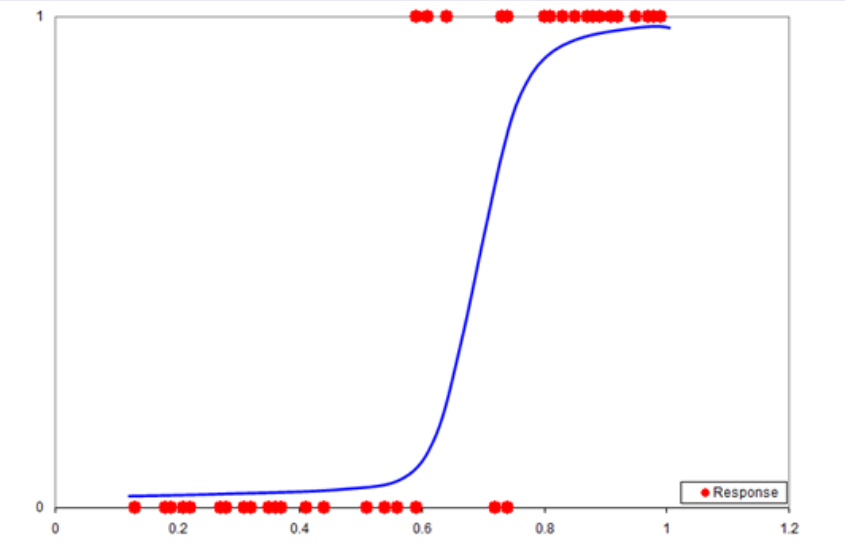


Computational Intelligence. Machine Learning

Supervised classification. LOGISTIC REGRESSION

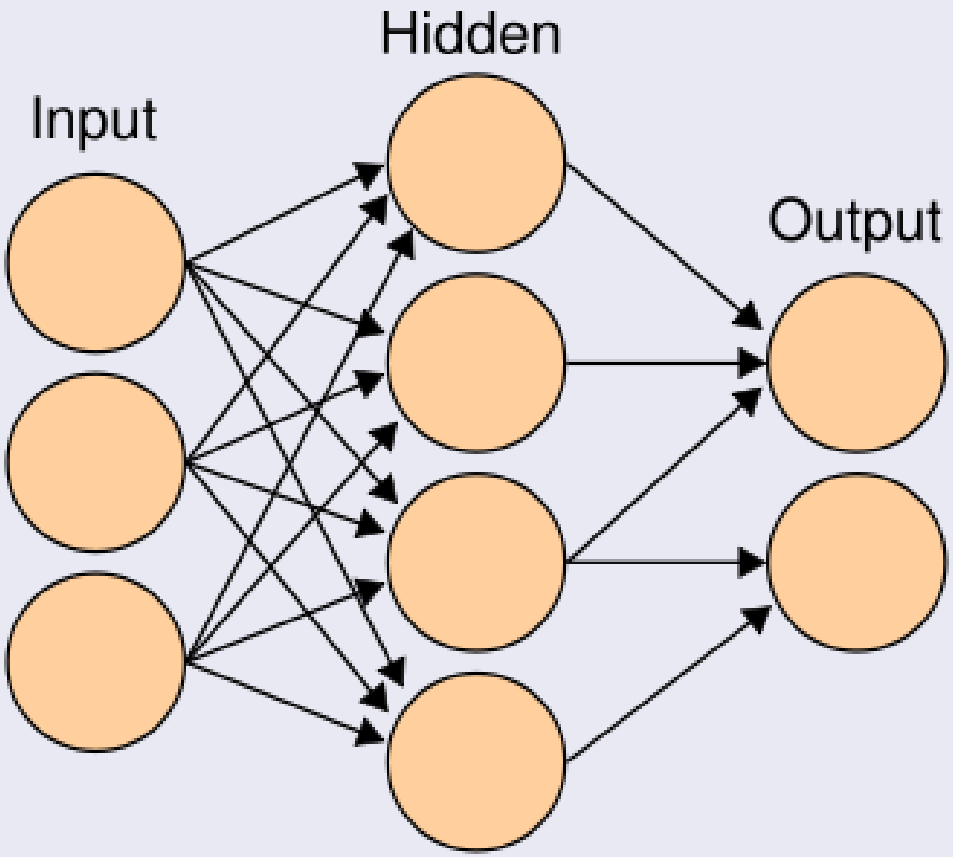
$$\pi_j = P(C = 1|\mathbf{x}_j) = \frac{1}{1+e^{-(\beta_0+\beta_1 x_{j1}+\cdots+\beta_n x_{jn})}}$$

$$\Rightarrow\ 1 - \pi_j = P(C = 0|\mathbf{x}_j) = \frac{1}{1+e^{(\beta_0+\beta_1 x_{j1}+\cdots+\beta_n x_{jn})}}$$



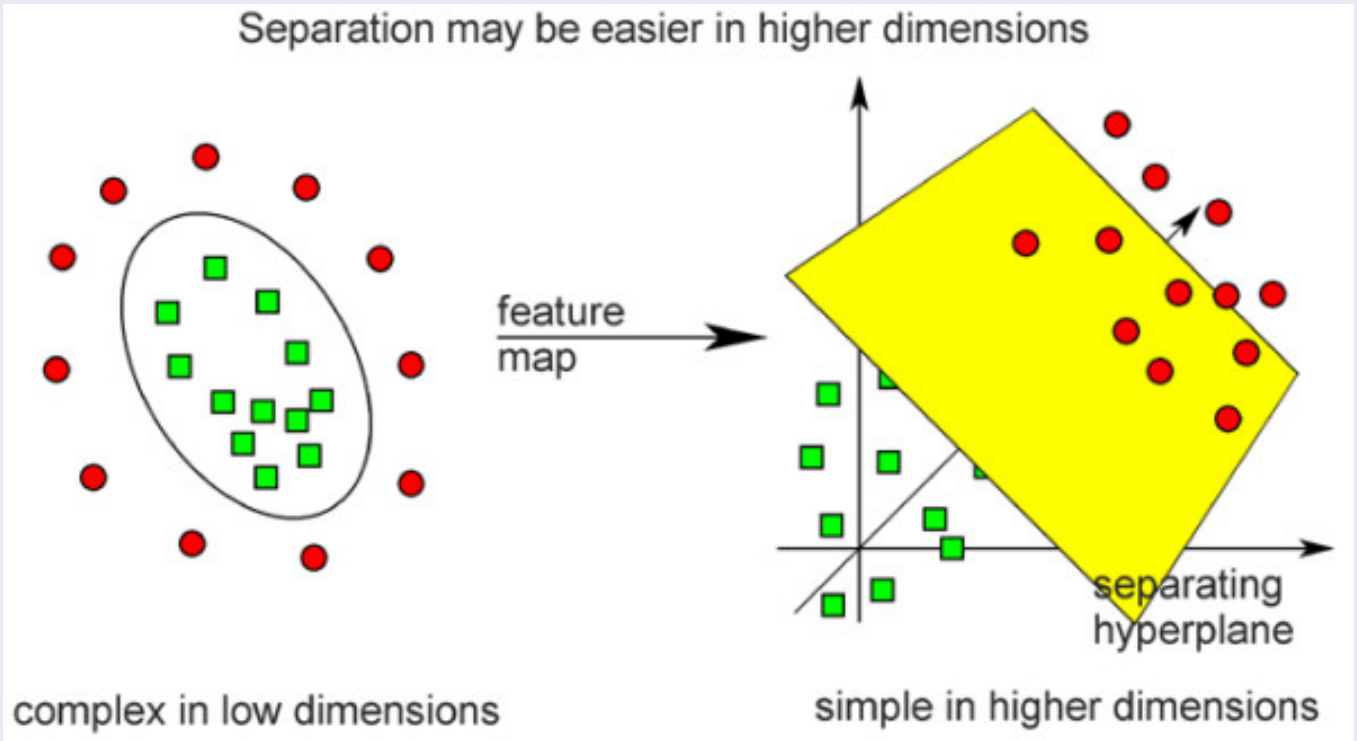
Computational Intelligence. Machine Learning

Supervised classification. NEURAL NETWORK



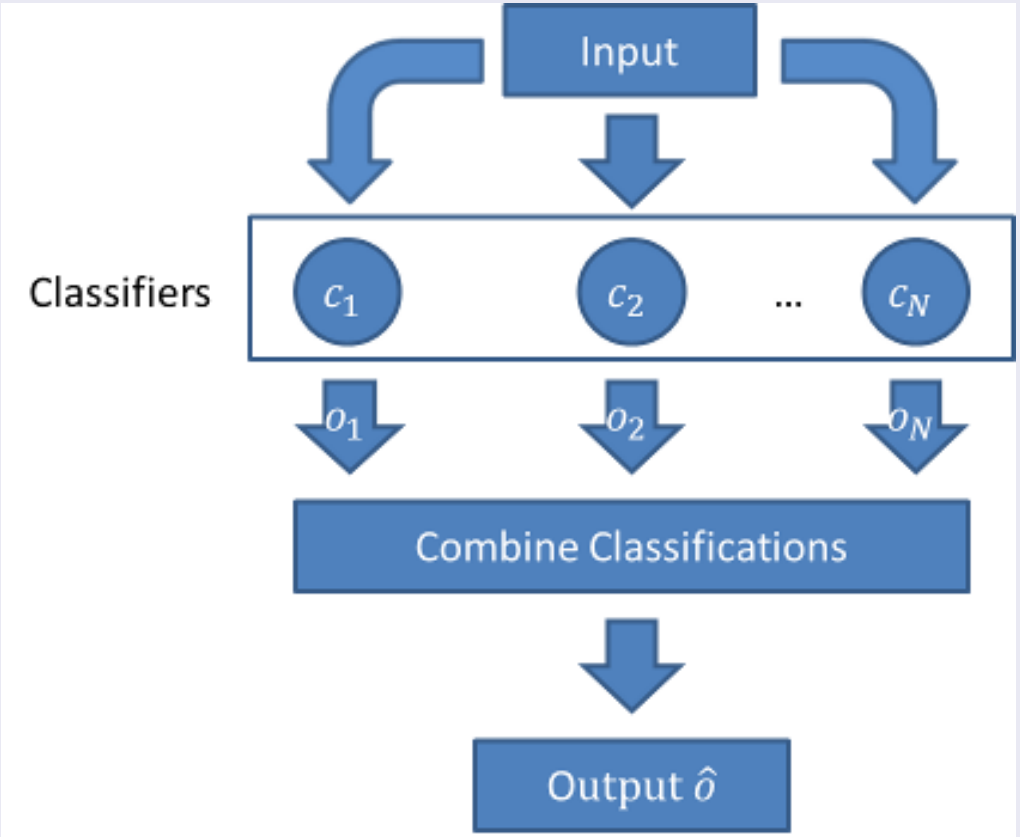
Computational Intelligence. Machine Learning

Supervised classification. SUPPORT VECTOR MACHINE



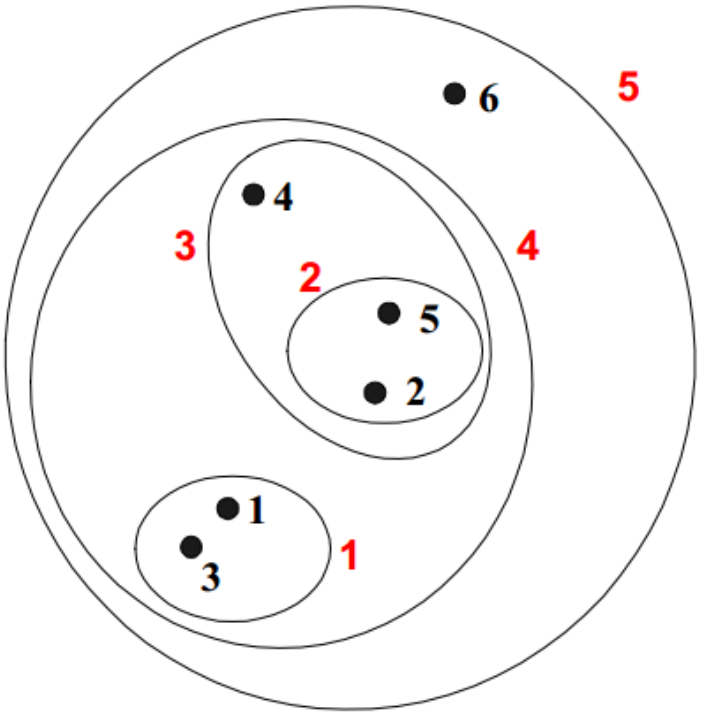
Computational Intelligence. Machine Learning

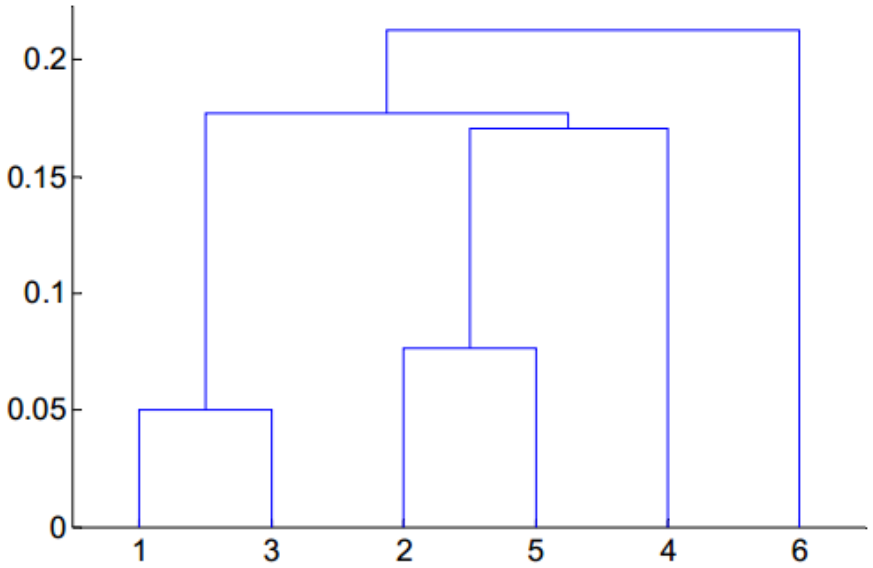
Supervised classification. METACLASSIFIERS



Computational Intelligence. Machine Learning

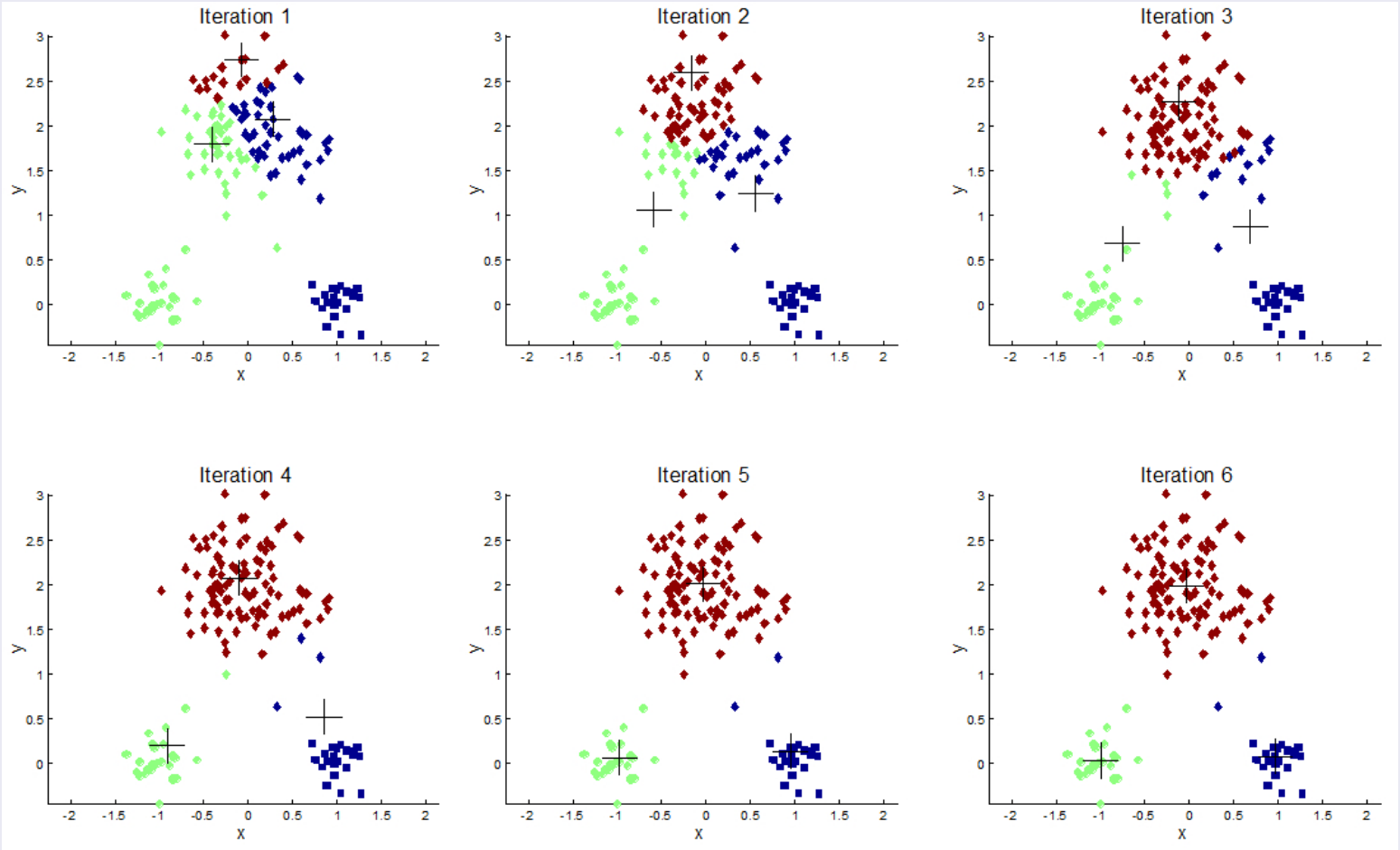
Clustering. HIERARCHICAL CLUSTERING





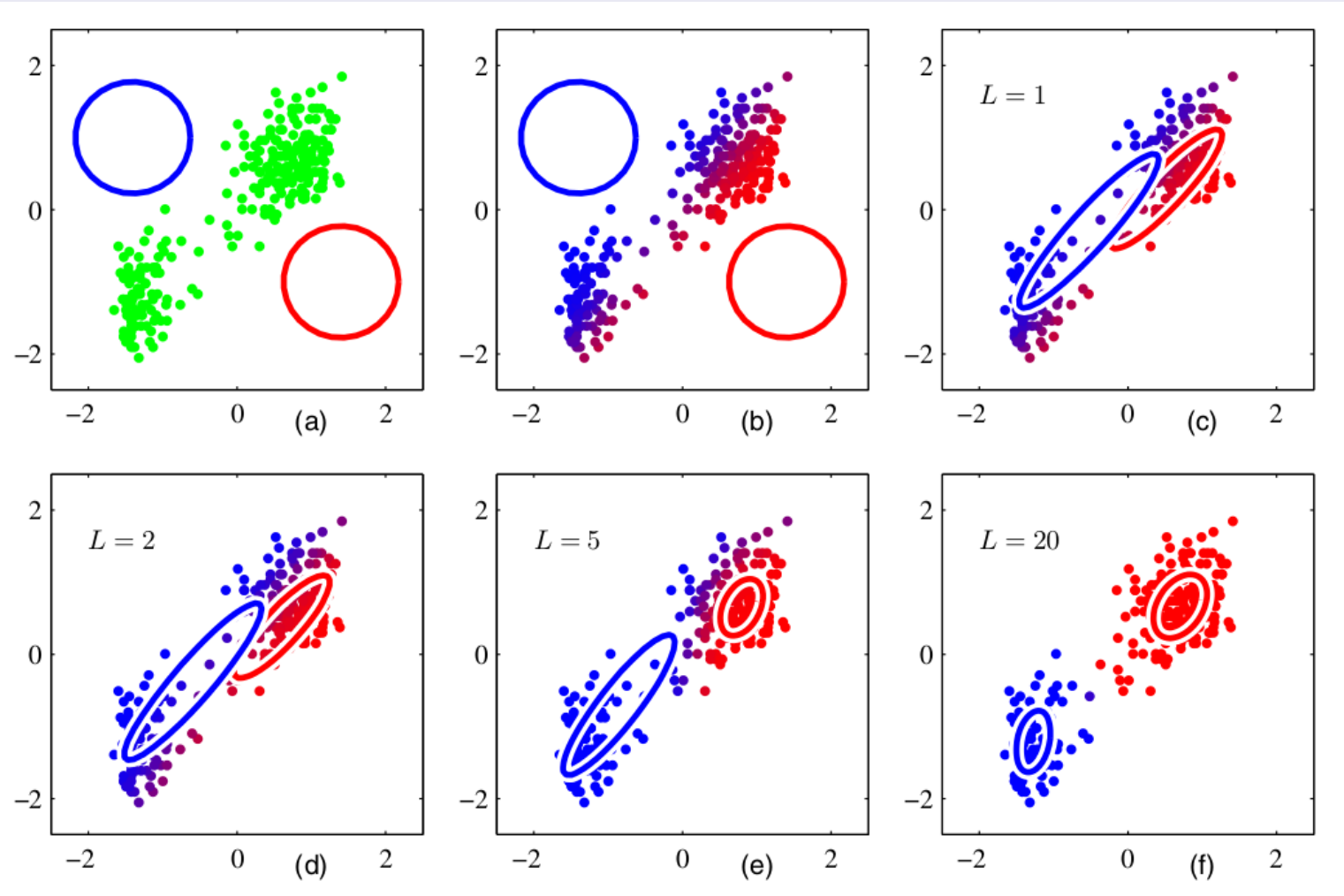
Computational Intelligence. Machine Learning

Clustering. PARTITIONAL CLUSTERING: k-means



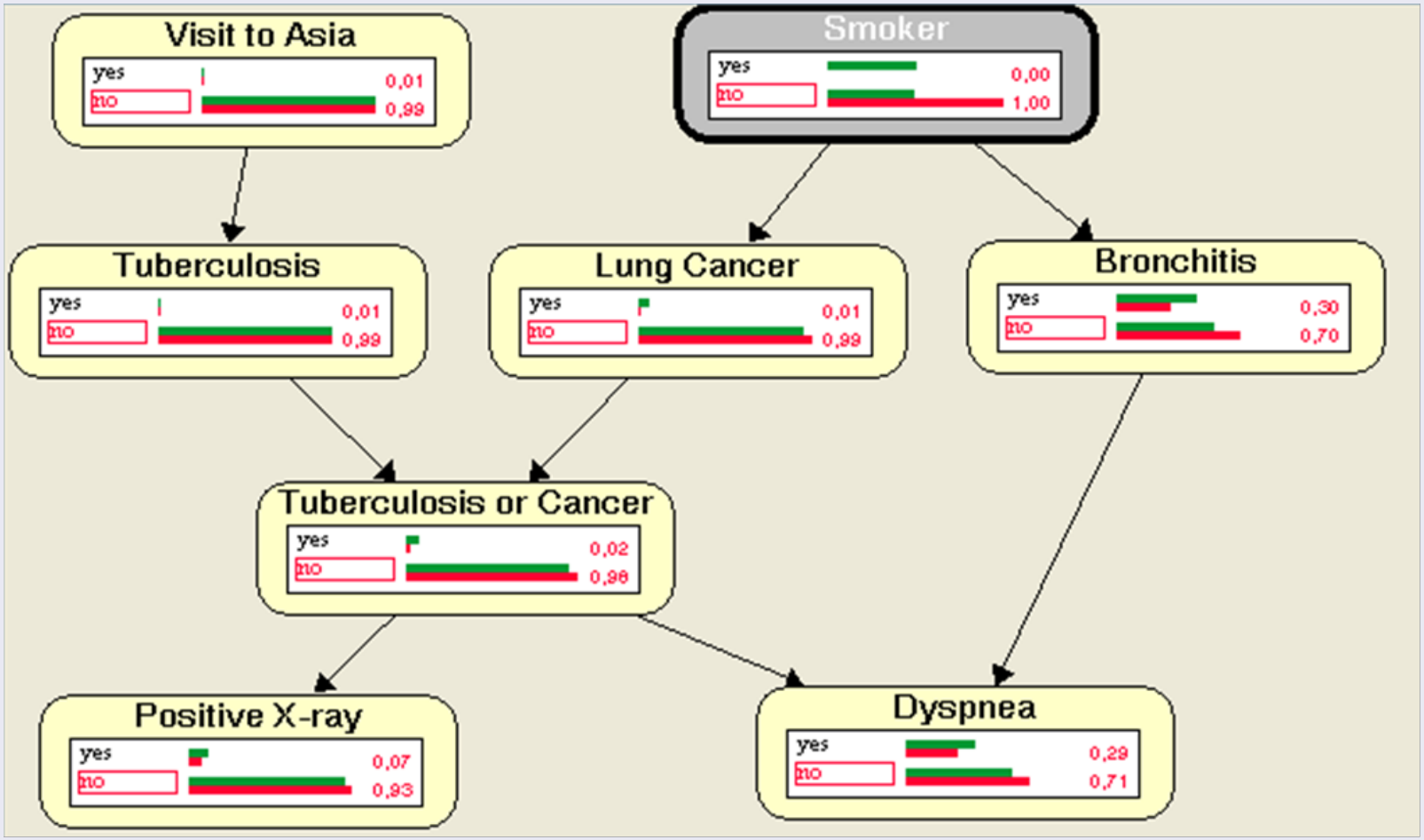
Computational Intelligence. Machine Learning

Clustering. PROBABILISTIC CLUSTERING: finite mixture models with EM



Computational Intelligence. Machine Learning

Associations. BAYESIAN NETWORKS



Evidence: "Smoker = no"

Outline

Successful Applications

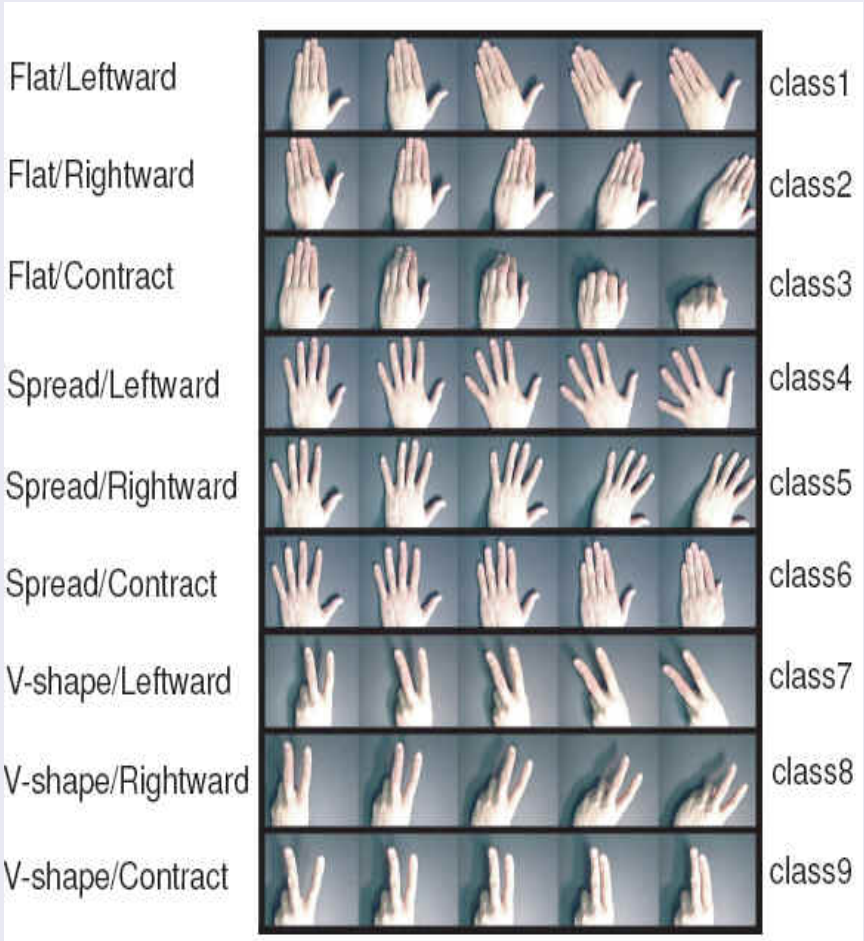
Chess

Successful Applications

JEOPARDY!

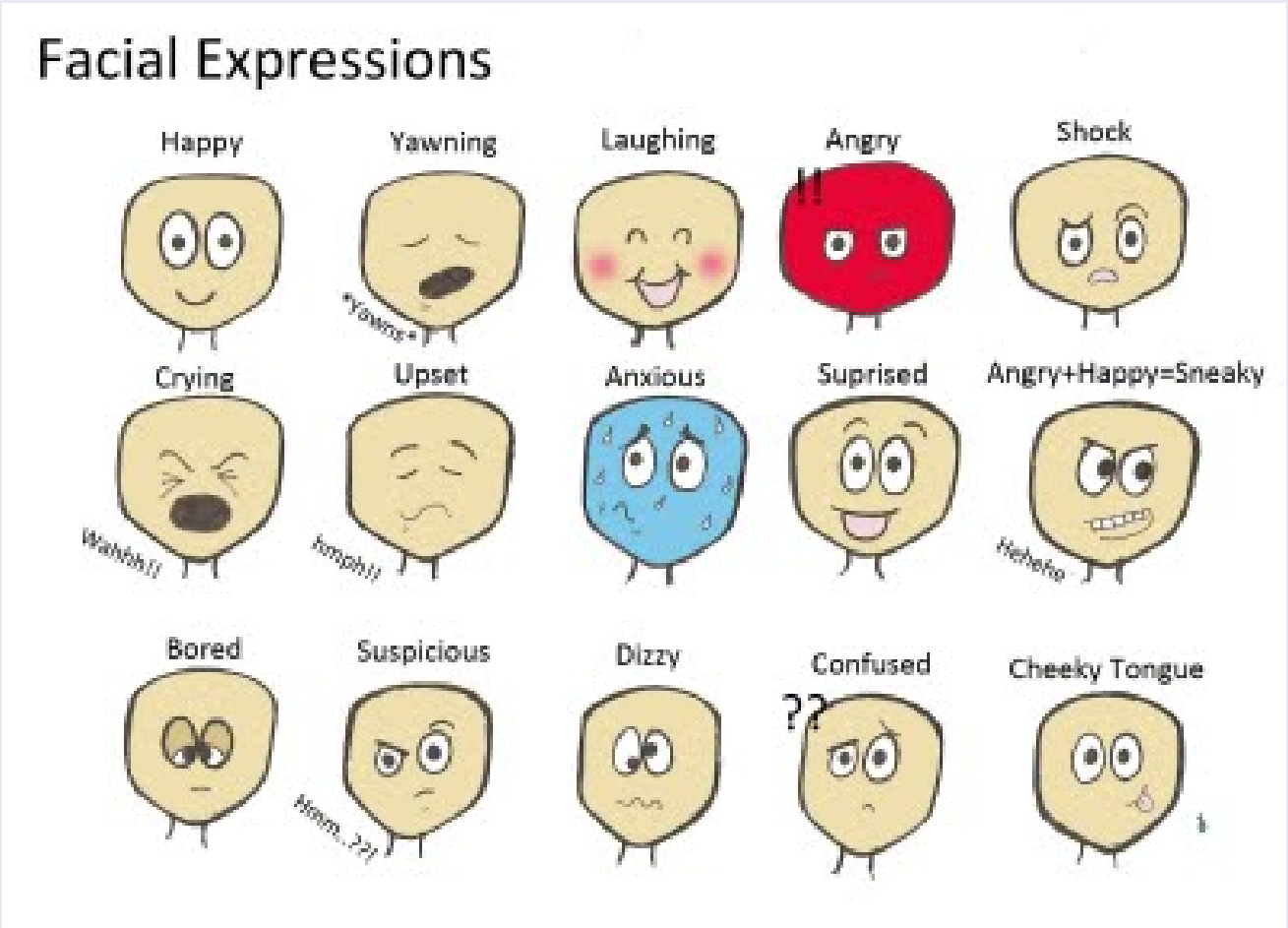
Successful Applications

Gesture Recognition



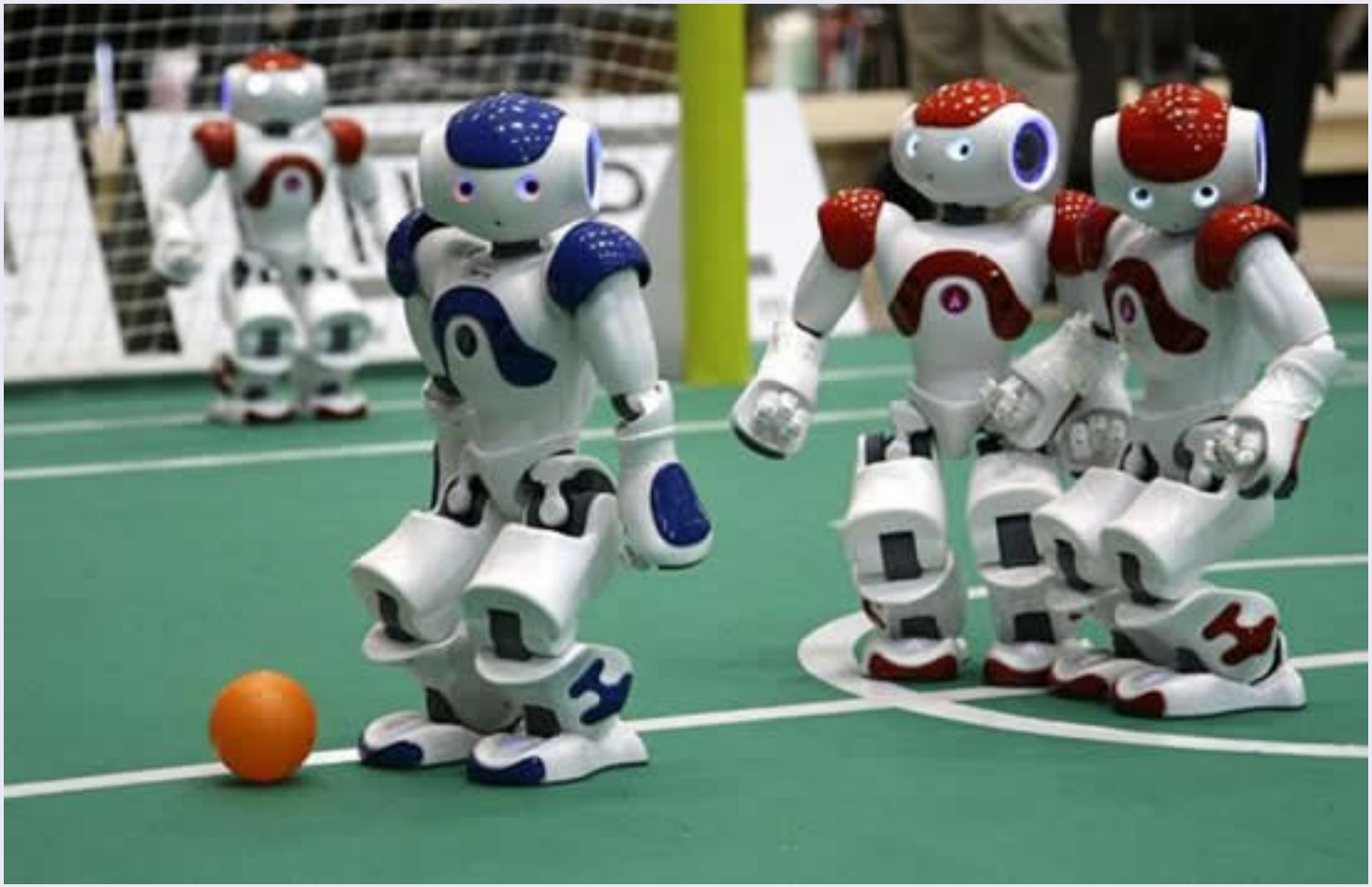
Successful Applications

Facial Expressions



Successful Applications

Robot Soccer



Successful Applications

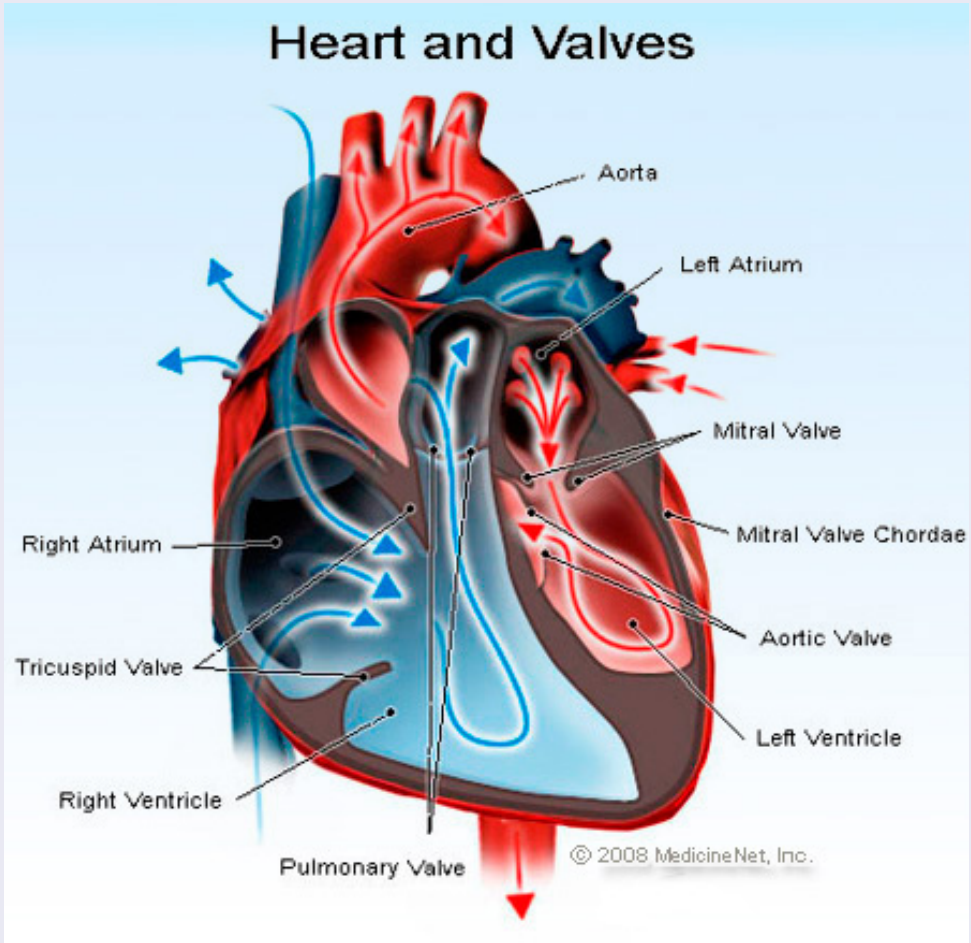
Brain Computer Interface

Successful Applications

Autonomous Car

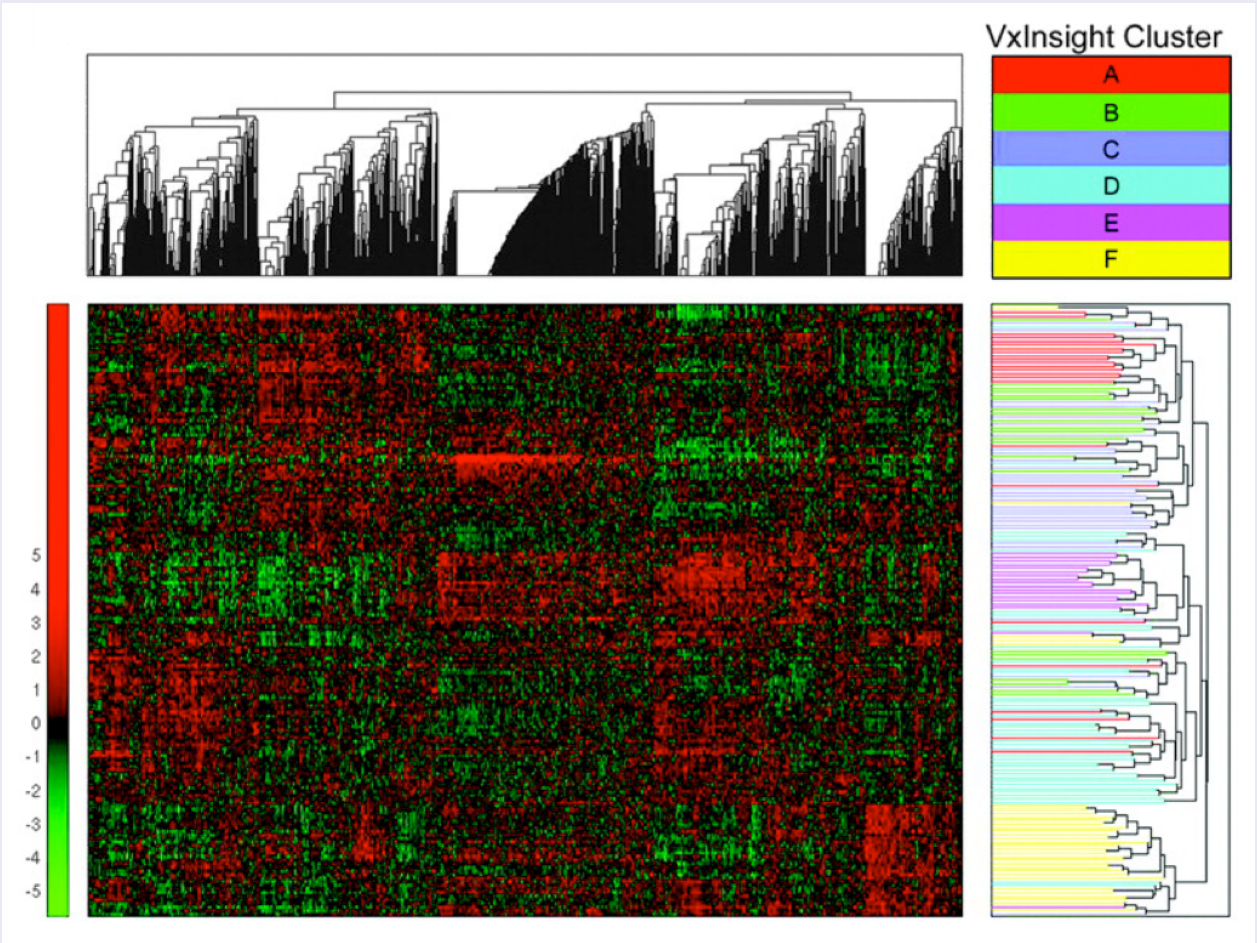
Successful Applications

Medical Diagnosis



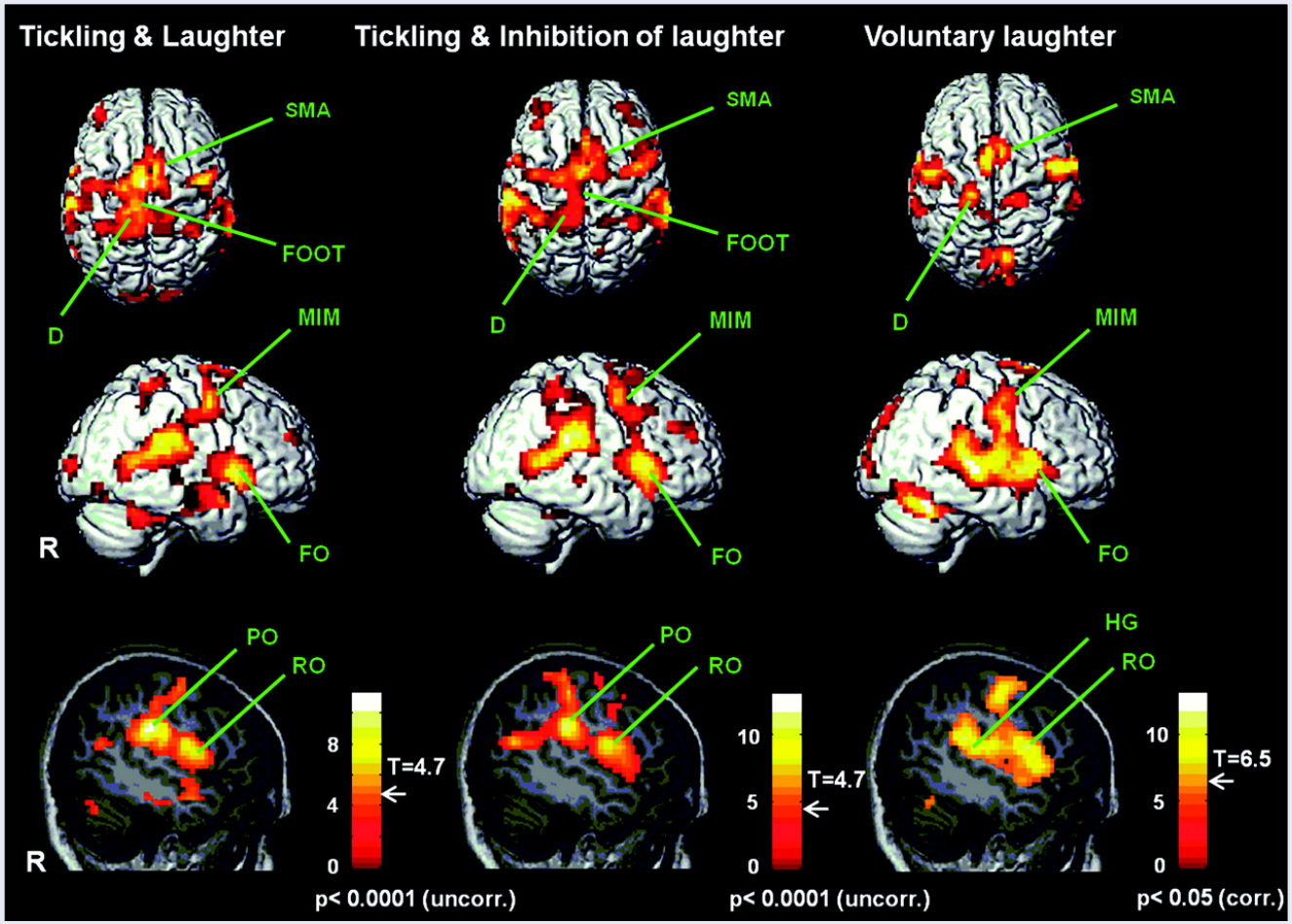
Successful Applications

Bioinformatics



Successful Applications

Neuroscience



Top 10 Algorithms in Data Mining

Knowledge Information Systems (2008) 14:1-37

1. C4.5 (Quinlan, 1993)
2. The *k*-means algorithm (Lloyd, 1957)
3. Support vector machines (Vapnik, 1995)
4. The Apriori algorithm (Agraval and Srikant, 1994)
5. The EM algorithm (Dempster et al., 1977)
6. PageRank algorithm (Brin and Page, 1998)
7. AdaBoost (Freund and Schapire, 1995)
8. *k*-nearest neighbors (Fix and Hodges, 1951)
9. Naive Bayes (Minsky, 1961)
10. CART: Classification and Regression Trees (Breiman et al., 1984)

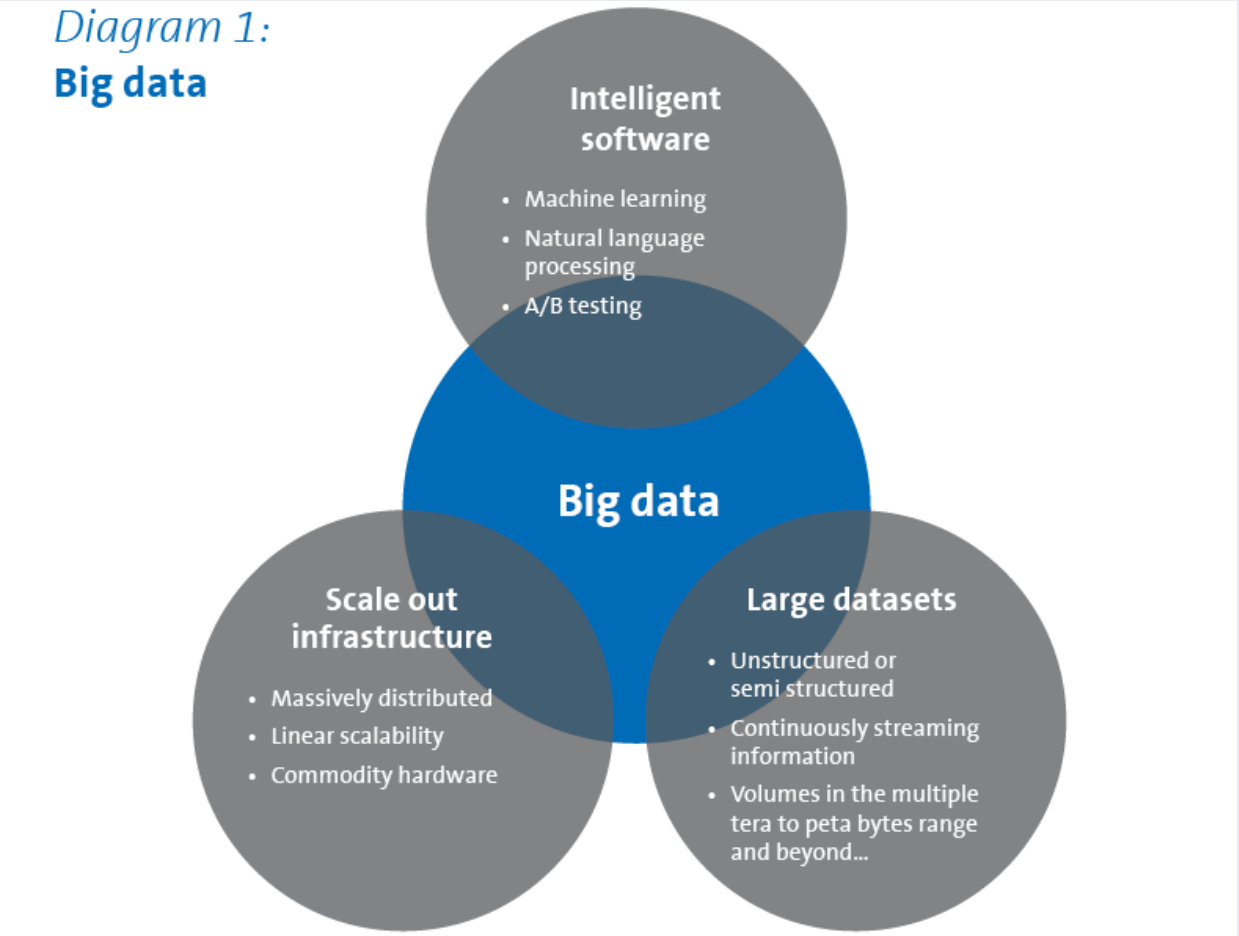
Outline

10 Challenging Problems

10 Challenging Problems

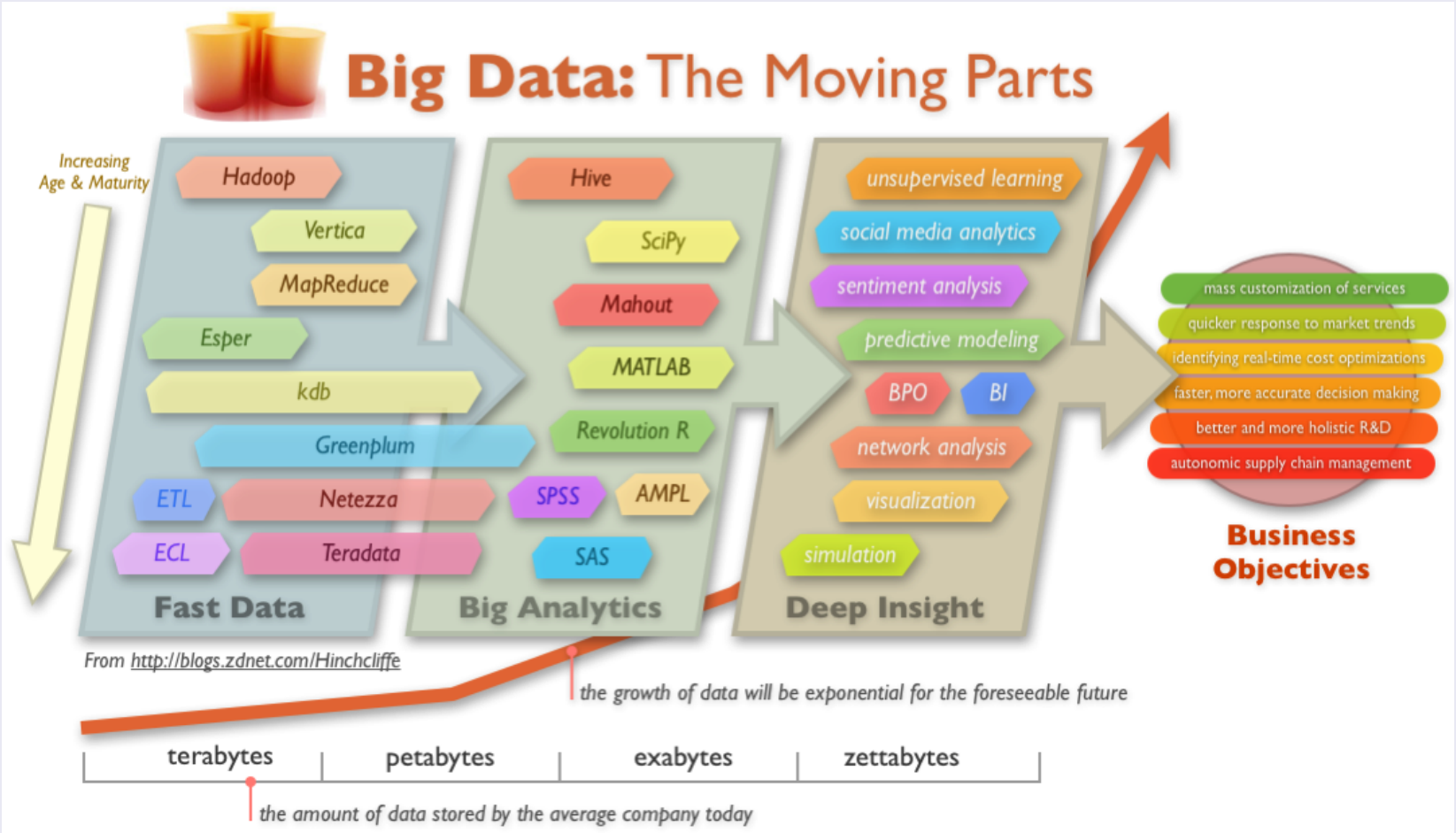
10 Challenges

1. Big Data: The five V's: volume, variety, velocity, viability and value



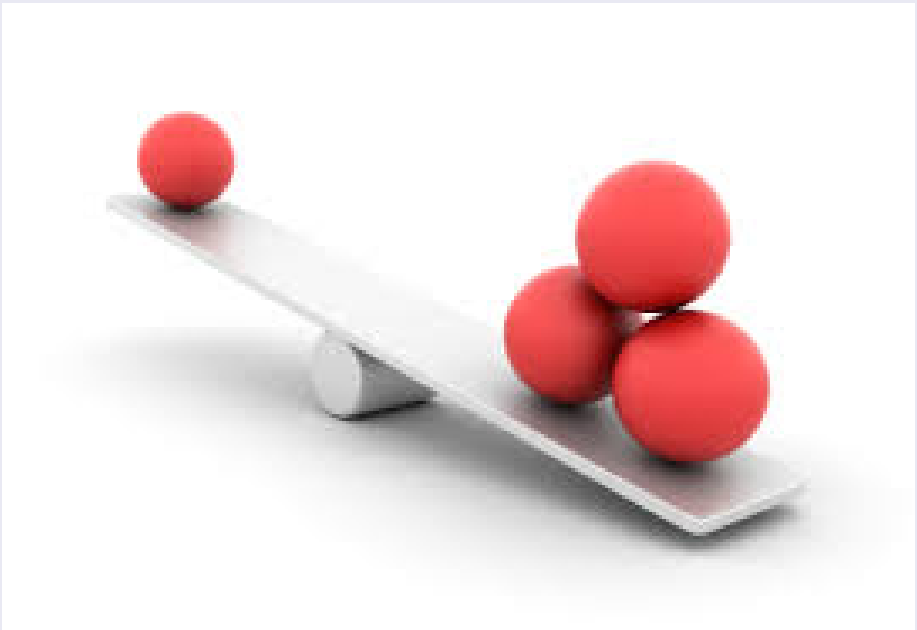
10 Challenges

1. Big Data. Adaptation of machine learning algorithms



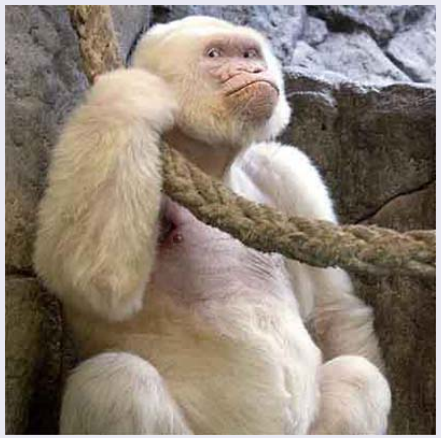
10 Challenges

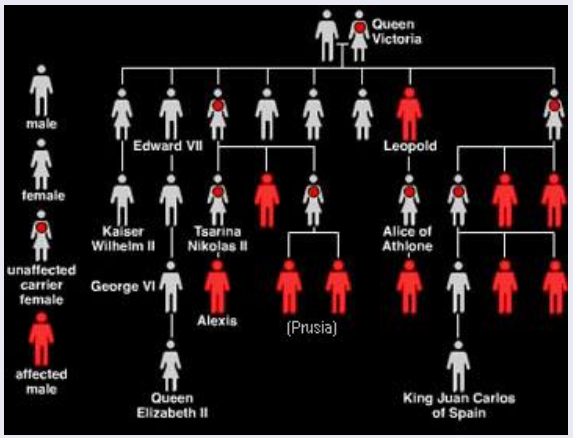
2. Variants of Supervised Classification. Class imbalance



10 Challenges

2. Variants of Supervised Classification. Positive Labels





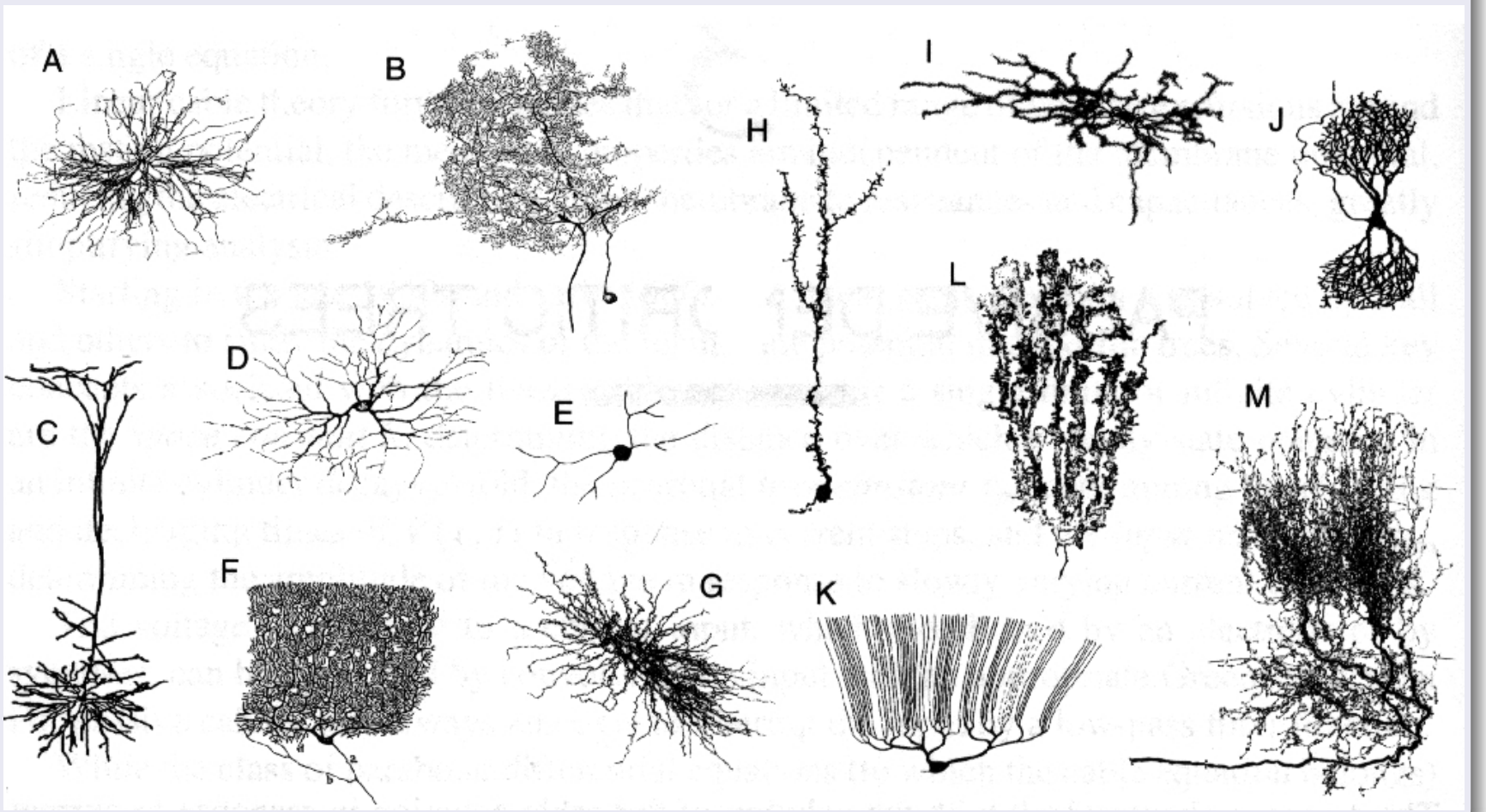
10 Challenges

2. Variants of Supervised Classification. Positive Labels

genes	X_1	...	X_n	C
$\boldsymbol{g}^{(1)}$	$x_1^{(1)}$	...	$x_n^{(1)}$	$+$
$\boldsymbol{g}^{(2)}$	$x_1^{(2)}$	...	$x_n^{(2)}$	$+$
$\boldsymbol{g}^{(3)}$	$x_1^{(3)}$	...	$x_n^{(3)}$	?
$\boldsymbol{g}^{(4)}$	$x_1^{(4)}$	...	$x_n^{(4)}$	$+$
...	...	...	.....	
$\boldsymbol{g}^{(N-1)}$	$x_1^{(N-1)}$	...	$x_n^{(N-1)}$	?
$\boldsymbol{g}^{(N)}$	$x_1^{(N)}$	...	$x_n^{(N)}$	?

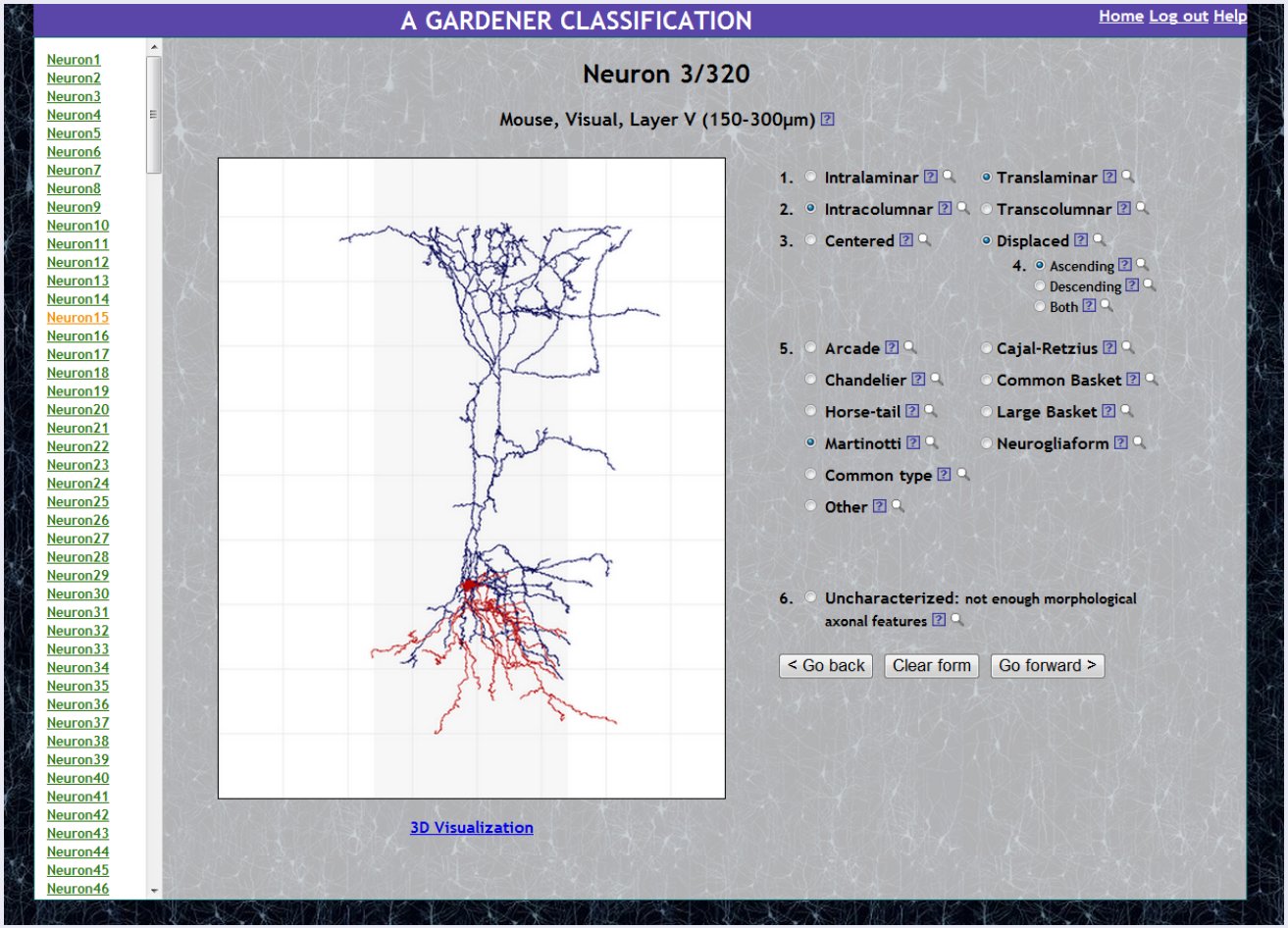
10 Challenges

2. Variants of Supervised Classification. Semi-supervised



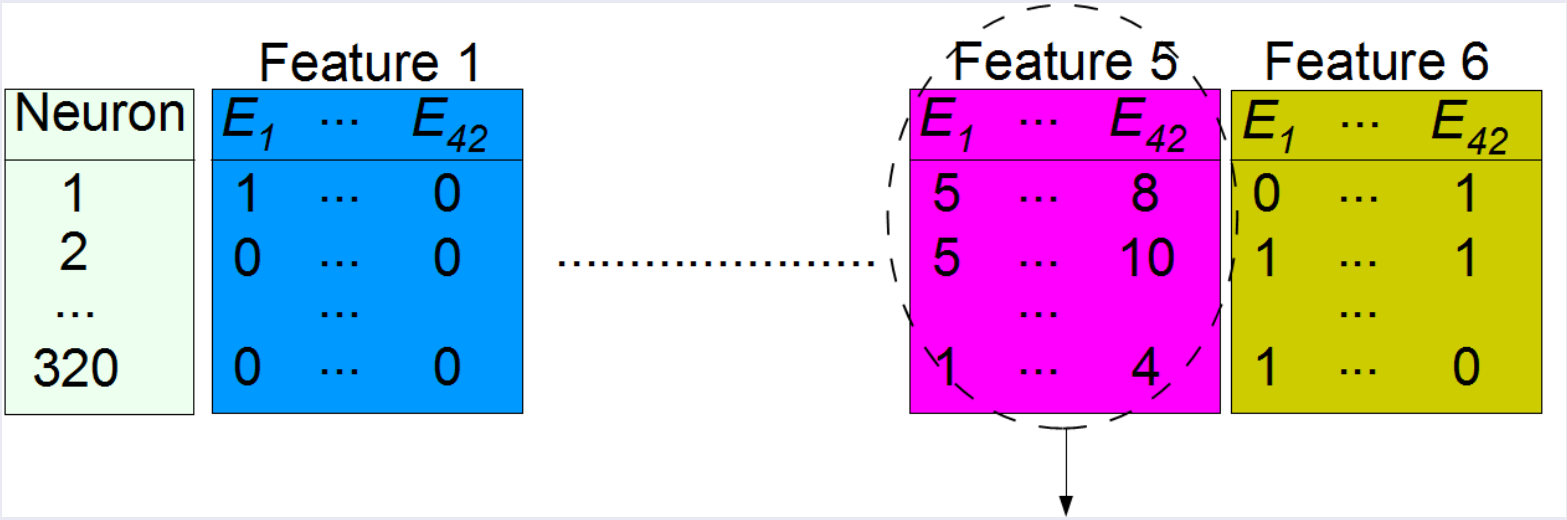
10 Challenges

2. Variants of Supervised Classification. Semi-supervised



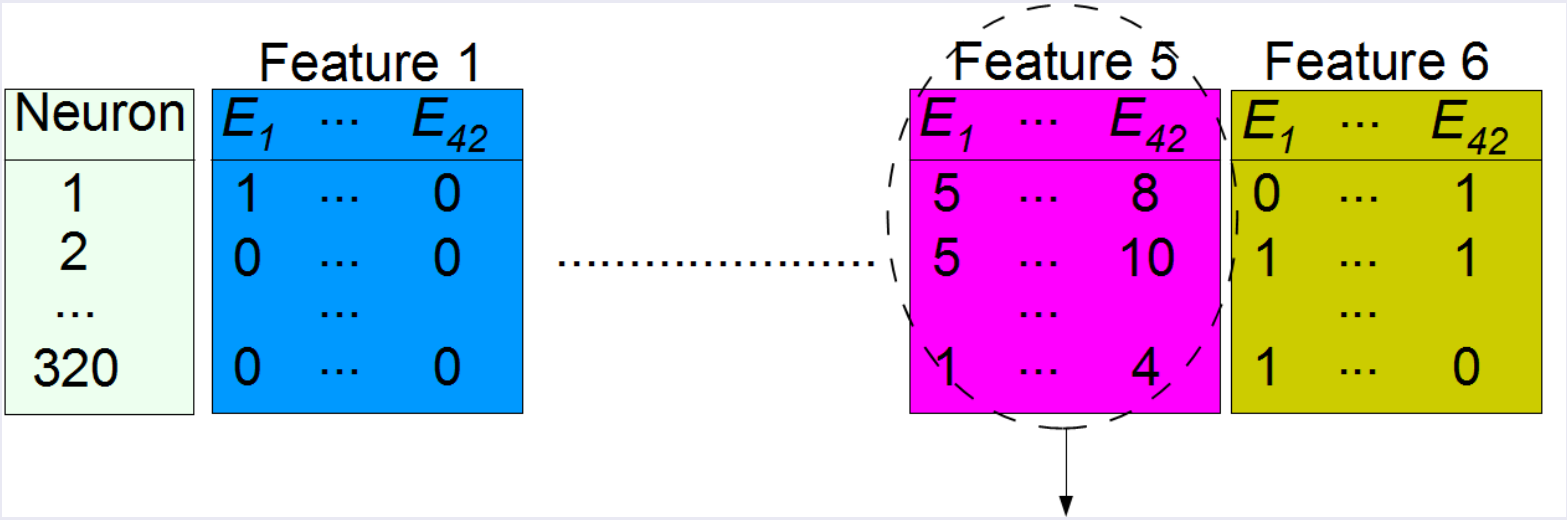
10 Challenges

2. Variants of Supervised Classification. Semi-supervised



10 Challenges

2. Variants of Supervised Classification. Semi-supervised



128 Axon, 86 Dendrite, 10 Soma ≥ 26 votes

Neuron	X_1	X_2	...	X_{224}	C
1	8.4	7.1	...	2.1	Martinotti (27)
2	2.0	5.4	...	4.0	?
3	7.3	2.0	...	1.1	?
...		.....			...
240	3.7	2.8	...	5.2	Basket (32)

10 Challenges

2. Variants of Supervised Classification. Probabilistic labels

	morpholog. variables			classif. 42 experts			class
cell	X_1	... ...	X_{2885}	E_1	... ...	E_{42}	C
1	10.1	... ...	6.6	trans	... ...	intra	38-4
2	3.7	... ...	7.7	intra	... ...	trans	24-18
3	5.9	... ...	9.2	intra	... ...	intra	35-7
4	11.2	... ...	10.1	intra	... ...	intra	10-32
...	...	... ...	...	...	... ...	...	...
...	...	... ...	...	...	... ...	...	...
...	...	... ...	...	...	... ...	...	...
240	13.6	... ...	5.7	intra	... ...	intra	3-39

10 Challenges

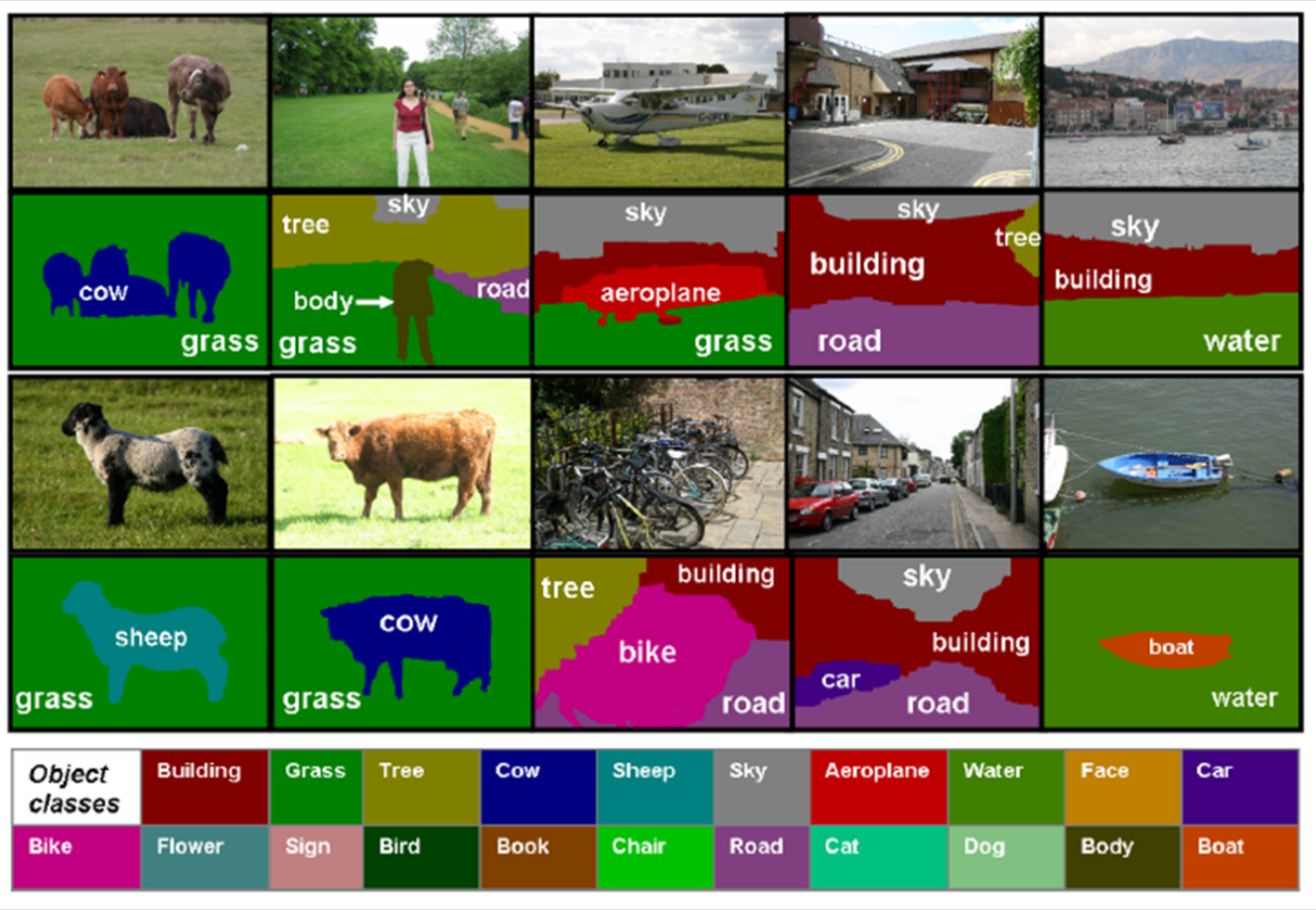
3. Multitarget Prediction. Multilabel Classification

X_1	X_2	X_3	X_4	X_5	C
3.2	1.4	4.7	7.5	3.7	1
2.8	6.3	1.6	4.7	2.7	0
7.7	6.2	4.1	3.3	7.7	1
9.2	0.4	2.8	0.5	3.9	0
5.5	5.3	4.9	0.6	6.6	1

X_1	X_2	X_3	X_4	X_5	C_1	C_2	C_3	C_4
3.2	1.4	4.7	7.5	3.7	1	0	1	1
2.8	6.3	1.6	4.7	2.7	0	0	1	0
7.7	6.2	4.1	3.3	7.7	1	0	1	1
9.2	0.4	2.8	0.5	3.9	0	1	0	0
5.5	5.3	4.9	0.6	6.6	1	1	0	1

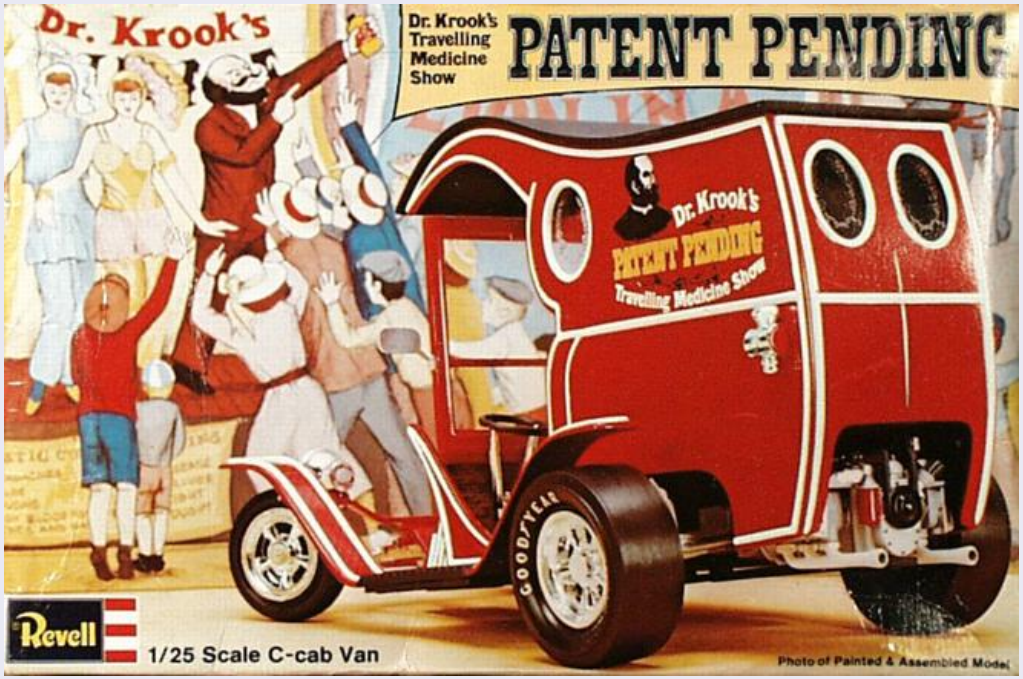
10 Challenges

3. Multitarget Prediction. Multilabel Classification



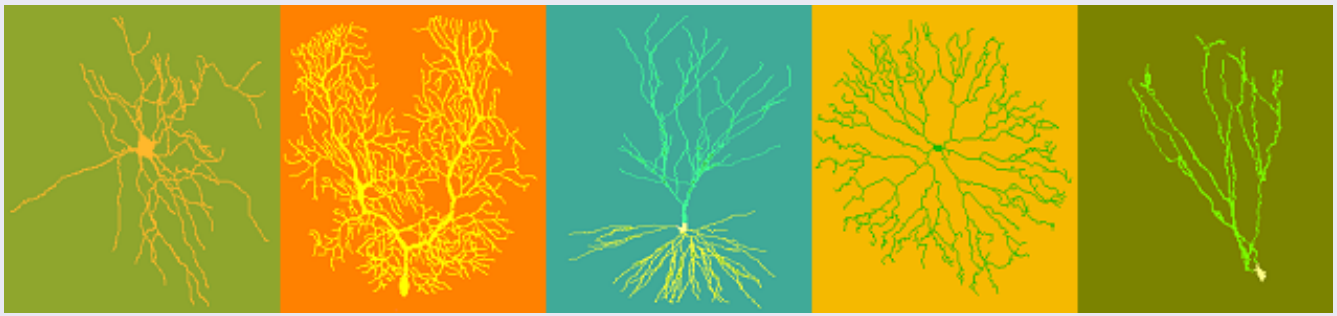
10 Challenges

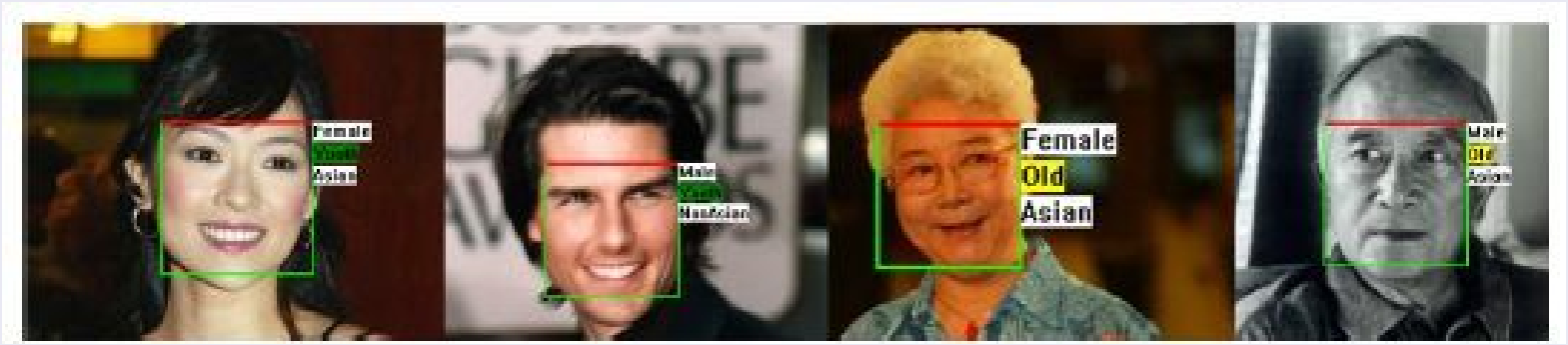
3. Multitarget Prediction. Multilabel Classification



10 Challenges

3. Multitarget Prediction. Multidimensional Classification





10 Challenges

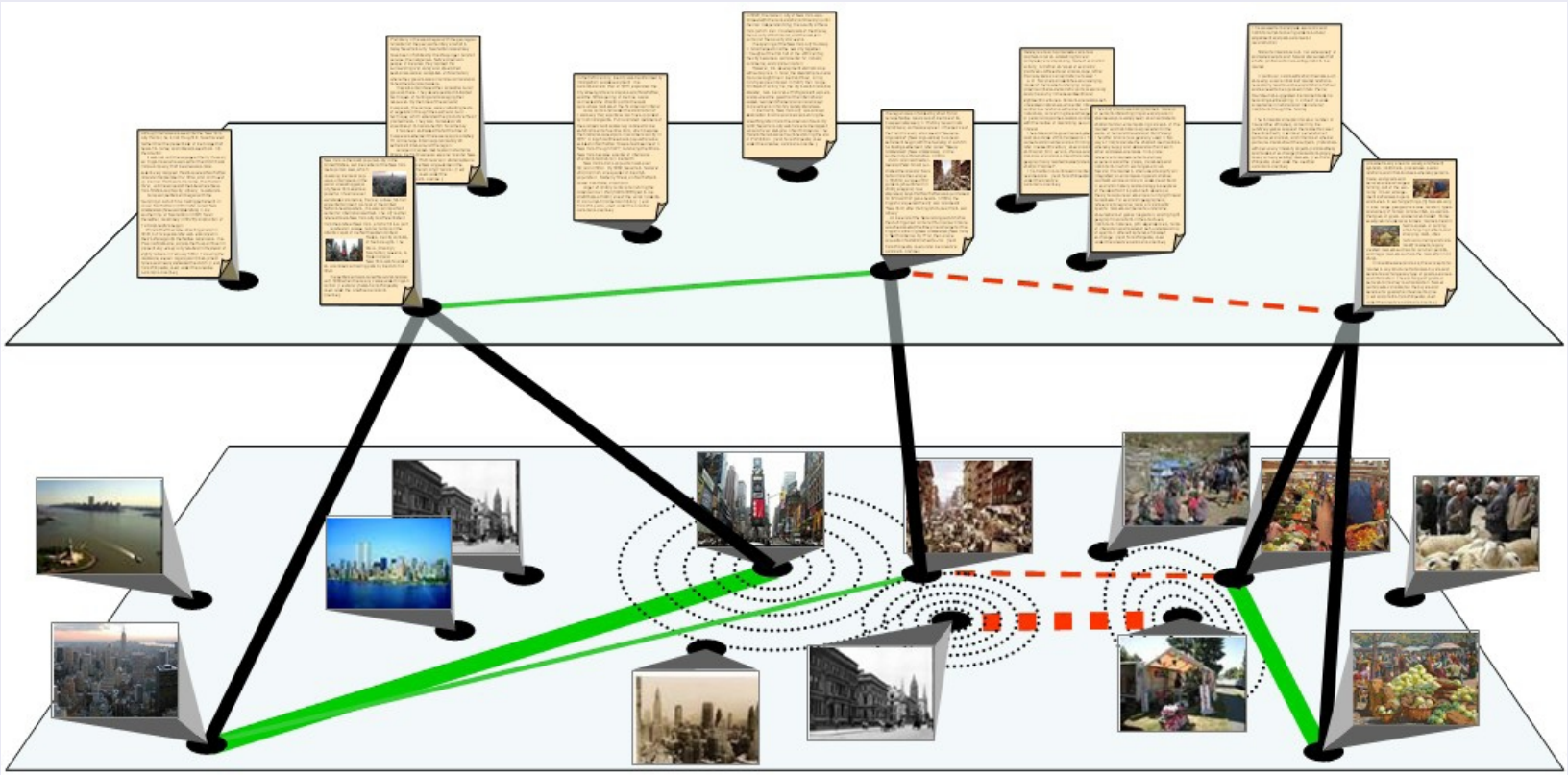
3. Multitarget Prediction. Multioutput Regression

X_1	X_2	X_3	X_4	X_5	Y
3.2	1.4	4.7	7.5	3.7	1.7
2.8	6.3	1.6	4.7	2.7	0.4
7.7	6.2	4.1	3.3	7.7	1.9
9.2	0.4	2.8	0.5	3.9	0.2
5.5	5.3	4.9	0.6	6.6	1.7

X_1	X_2	X_3	X_4	X_5	Y_1	Y_2	Y_3	Y_4
3.2	1.4	4.7	7.5	3.7	1.8	0.2	1.8	1.2
2.8	6.3	1.6	4.7	2.7	0.3	0.4	1.1	0.1
7.7	6.2	4.1	3.3	7.7	1.3	0.4	1.1	1.9
9.2	0.4	2.8	0.5	3.9	0.7	1.1	0.1	0.9
5.5	5.3	4.9	0.6	6.6	1.1	1.2	0.7	1.2

10 Challenges

4. Clustering. Multipartition. Subspace Clustering



10 Challenges

5. Regularization

lasso: $\hat{\beta} = \text{argmin}_{\beta} ||\boldsymbol{y} - \boldsymbol{X}\beta||_2^2 + \lambda||\beta||_1$

adaptive lasso: $\hat{\beta} = \text{argmin}_{\beta} ||\boldsymbol{y} - \boldsymbol{X}\beta||_2^2 + \lambda \sum_{j=1}^{p} \boldsymbol{w}_j|\beta_j|$

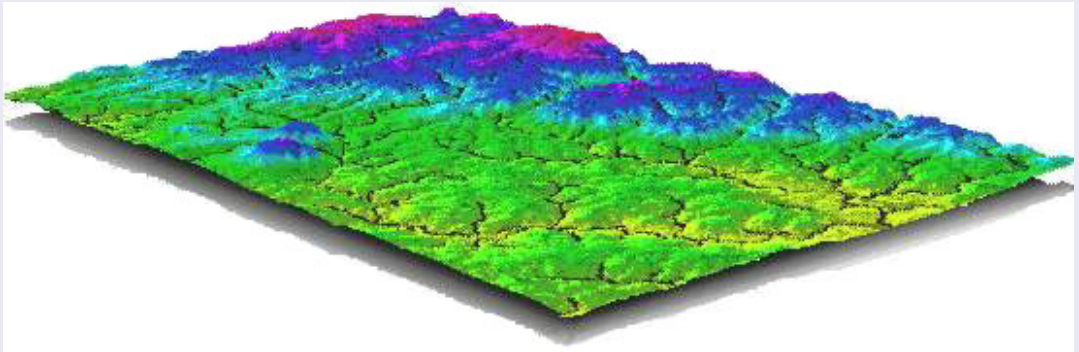
Dantzig: $\hat{\beta} = \text{argmin}_{\beta} ||\boldsymbol{y} - \boldsymbol{X}\beta||_\infty + \lambda||\beta||_1$

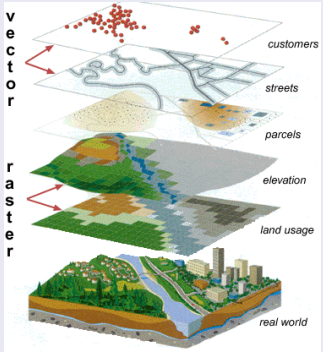
elastic net: $\hat{\beta} = \text{argmin}_{\beta} ||\boldsymbol{y} - \boldsymbol{X}\beta||_2^2 + \lambda_1||\beta||_1 + \lambda_2||\beta||_2^2$

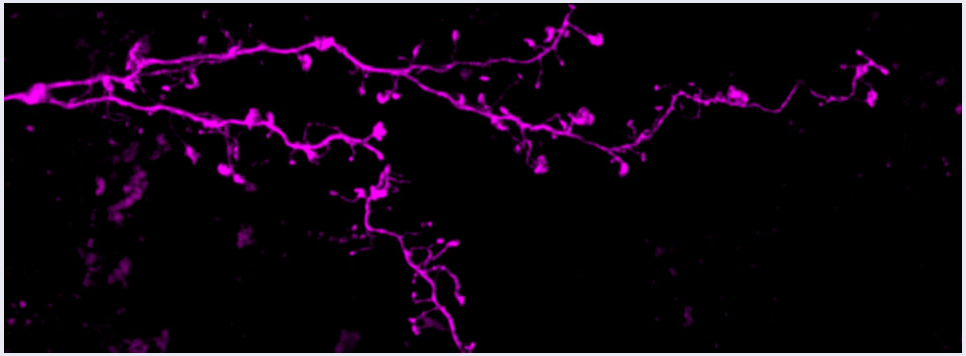
group lasso: $\hat{\beta} = \text{argmin}_{\beta} ||\boldsymbol{y} - \boldsymbol{X}\beta||_2^2 + \lambda \sum_{j=1}^{J} ||\beta_j||\boldsymbol{w}_j$

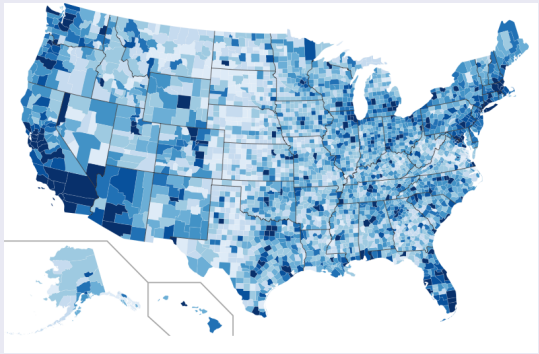
10 Challenges

6. Spatial Data



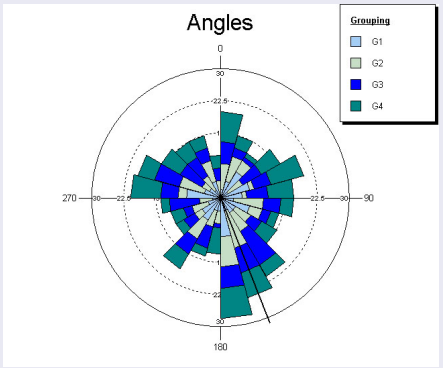


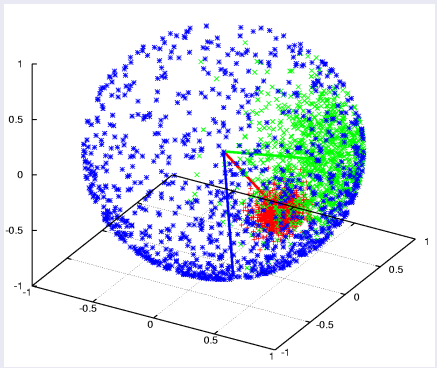


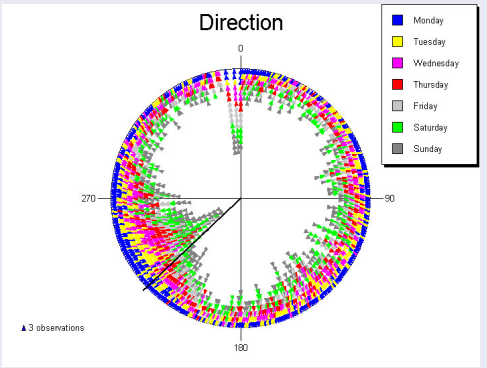


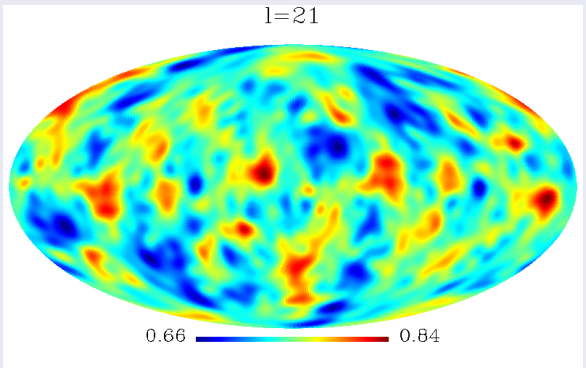
10 Challenges

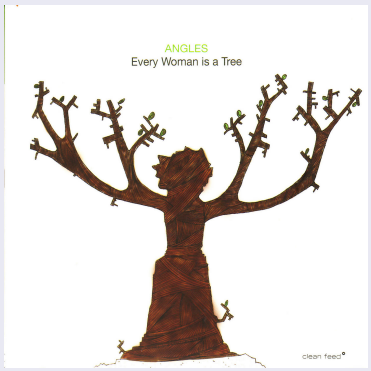
7. Directional Data

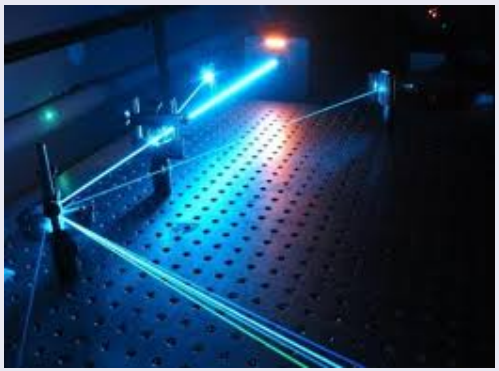






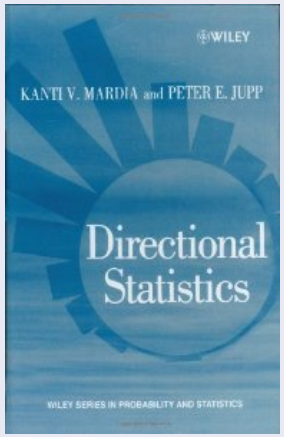


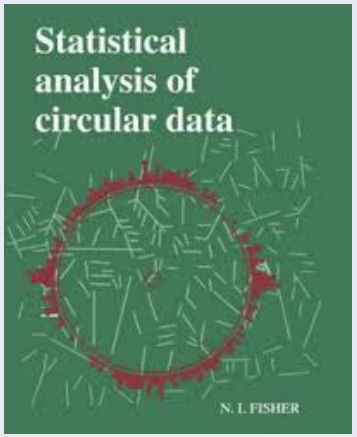


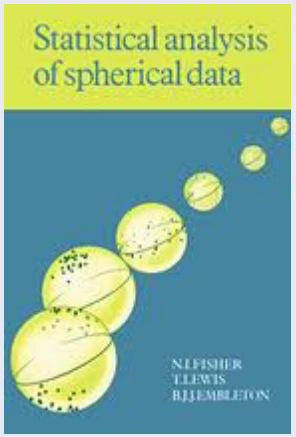


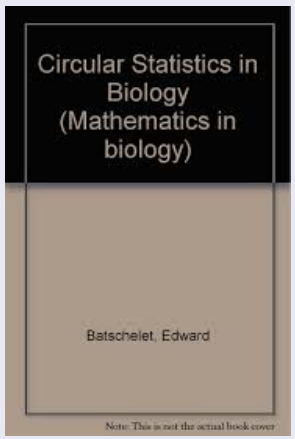
10 Challenges

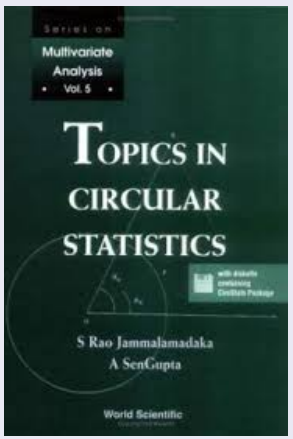
7. Directional Data. Books

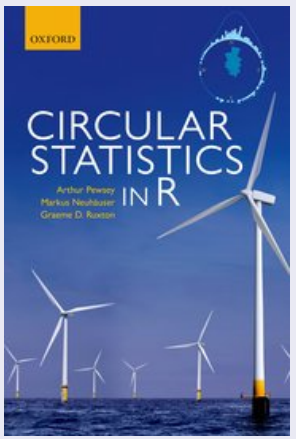










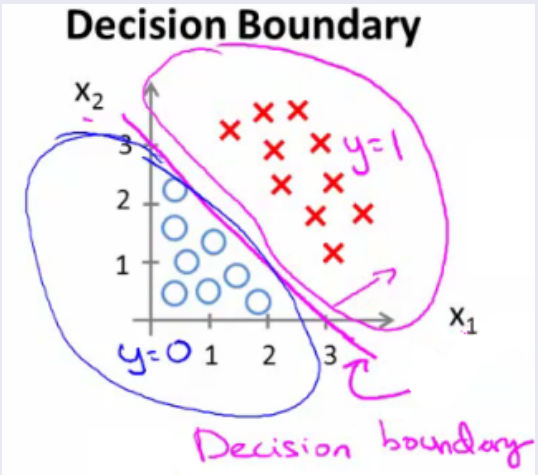


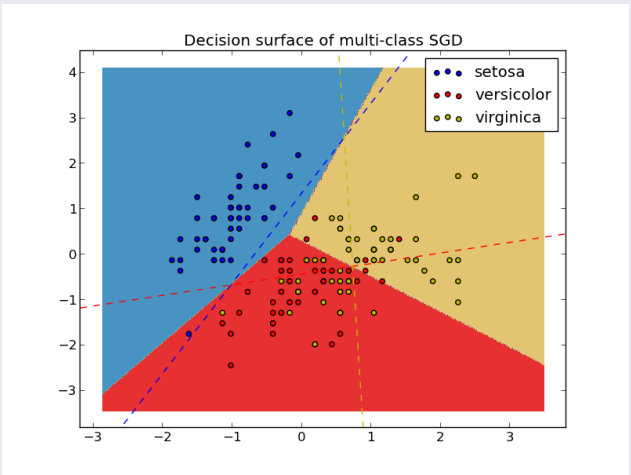
10 Challenges

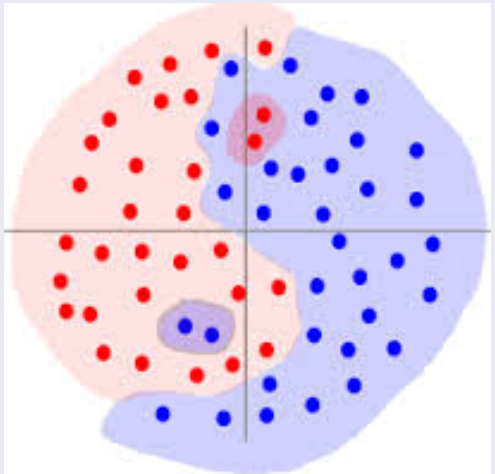
8. Data Streams. Concept Drift

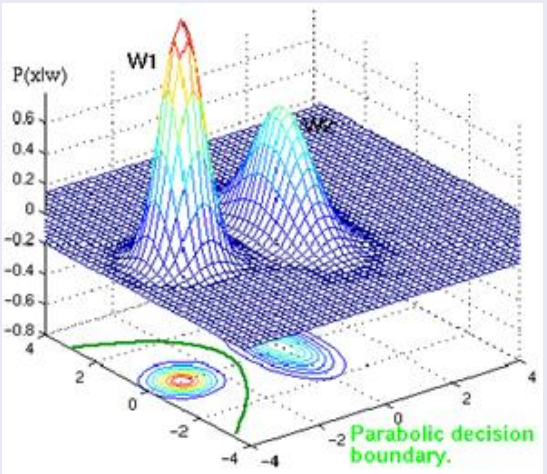
10 Challenges

9. Performance Measures. Decision Surfaces



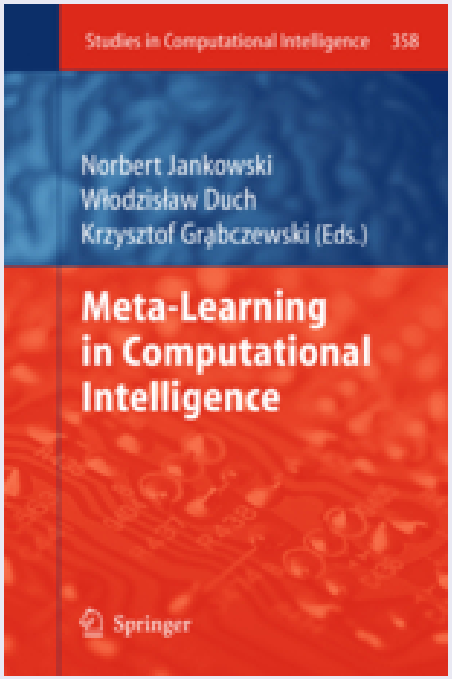


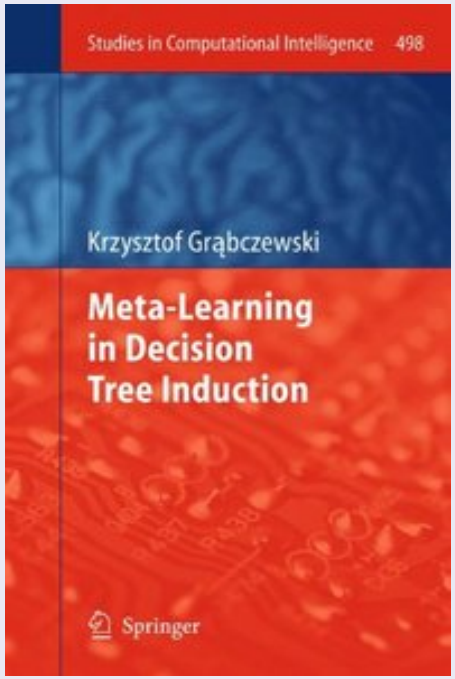




10 Challenges

10. Metalearning





Outline

Computational Intelligence

10 Challenges

1. Big data
2. Variants of supervised classification
3. Multitarget prediction
4. Clustering
5. Regularization
6. Spatial data
7. Directional data
8. Data streams
9. Performance measures
10. Metalearning

Computational Intelligence: Current State and Challenges

Pedro Larrañaga

pedro.larranaga@fi.upm.es
http://cig.fi.upm.es/

Computational Intelligence Group
Departamento de Inteligencia Artificial
Universidad Politécnica de Madrid

EVIA 2014 · A Coruña, September 3, 2014