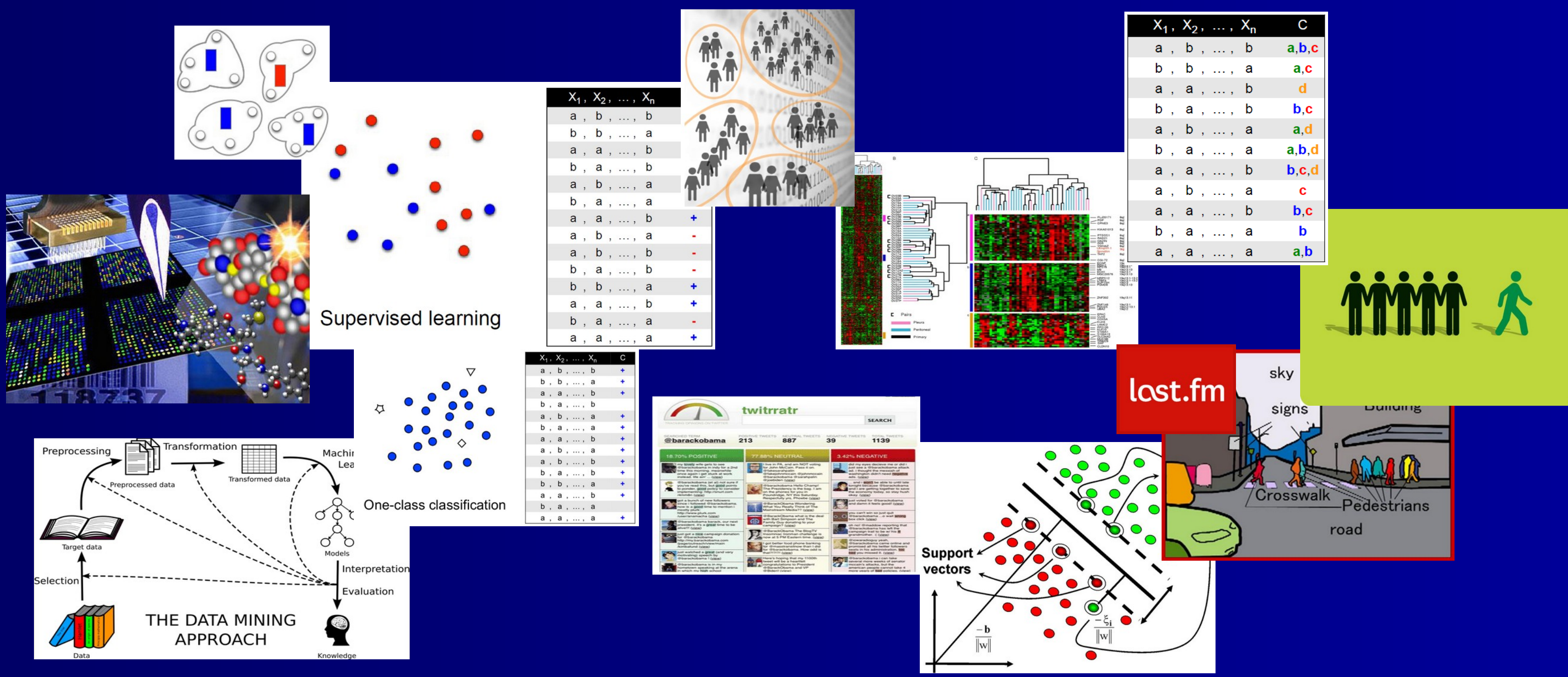
LEARNING UNDER WEAK SUPERVISION

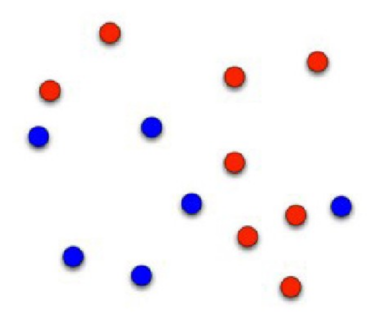


DATAI-UNAV, September'2022

SUPERVISED CLASSIFICATION

- Given a collection of records-samples [*training set*]
  - Each record → contains a set of *attributes-features-predictors*
  - Each record → belongs to a *class, our variable of interest, to be predicted*

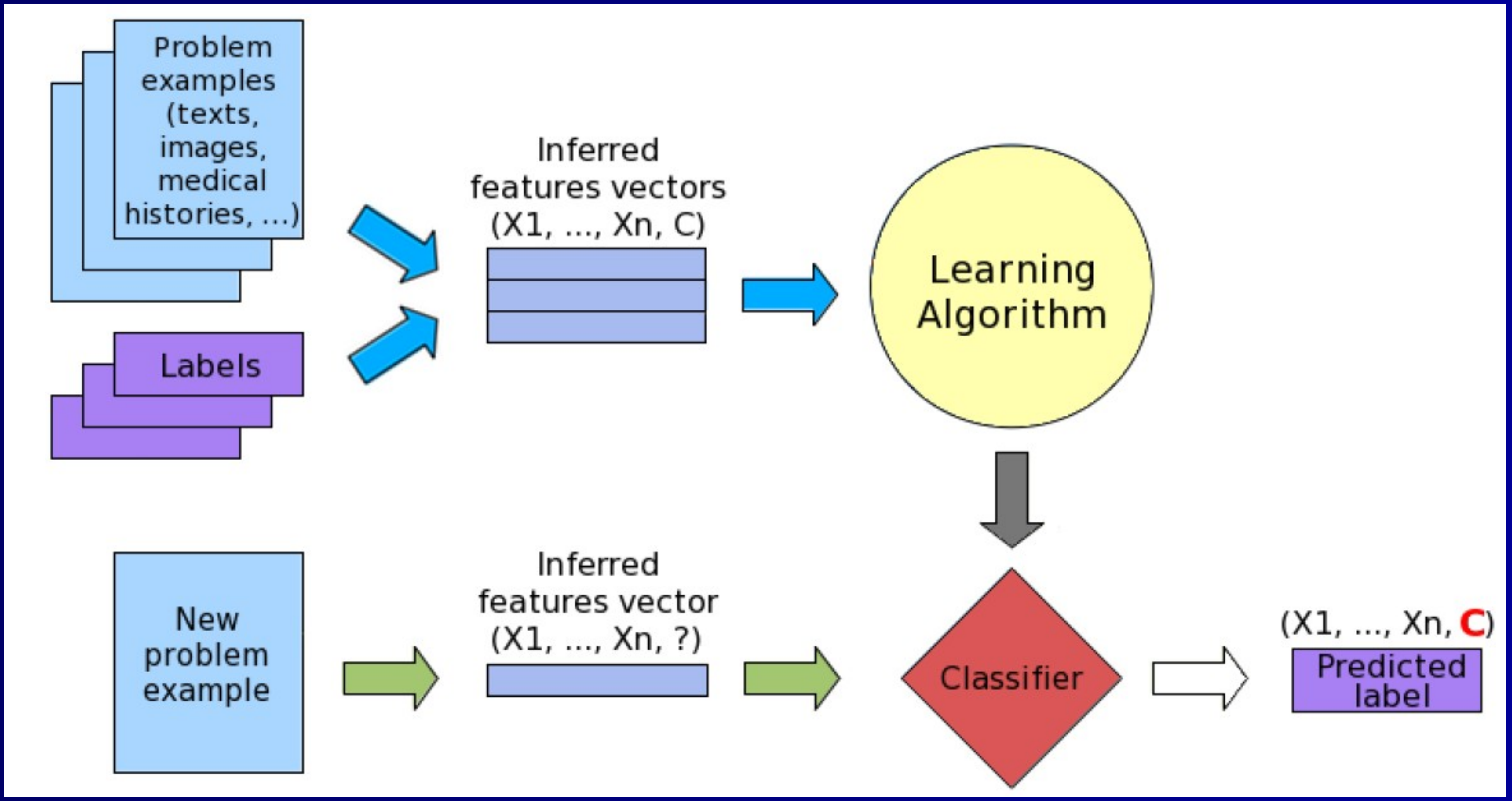
- Full supervision in training time



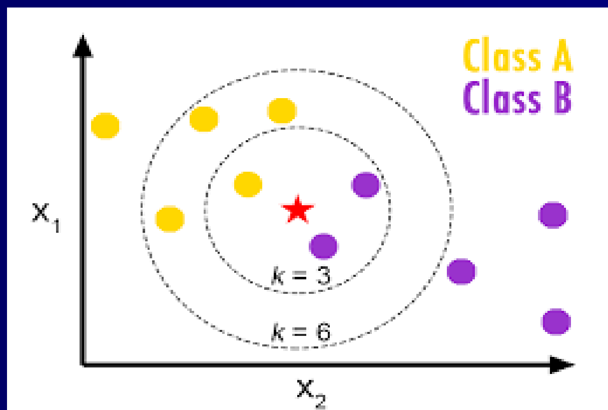
Supervised learning

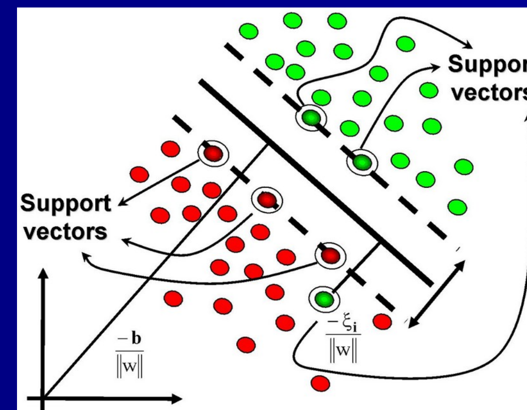
X_1 , X_2 , … , X_n	C
a , b , … , b	+
b , b , … , a	-
a , a , … , b	-
b , a , … , b	+
a , b , … , a	-
b , a , … , a	-
a , a , … , b	+
a , b , … , a	-
a , b , … , b	-
b , a , … , b	-
b , b , … , a	+
a , a , … , b	+
b , a , … , a	-
a , a , … , a	+

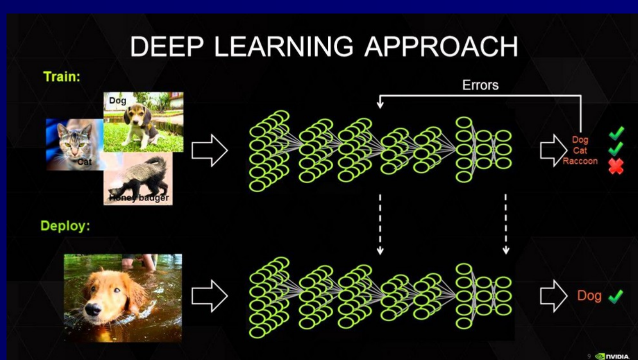
SUPERVISED CLASSIFICATION
- standard scenario -

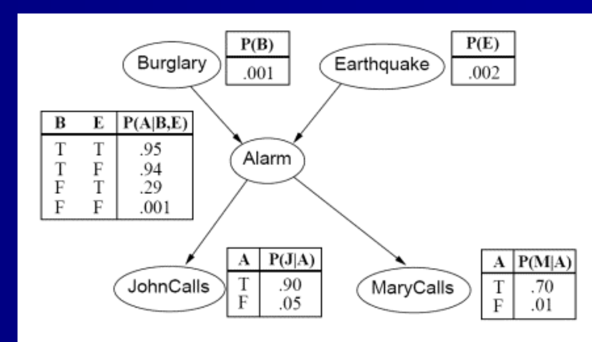


SUPERVISED CLASSIFICATION
- models -

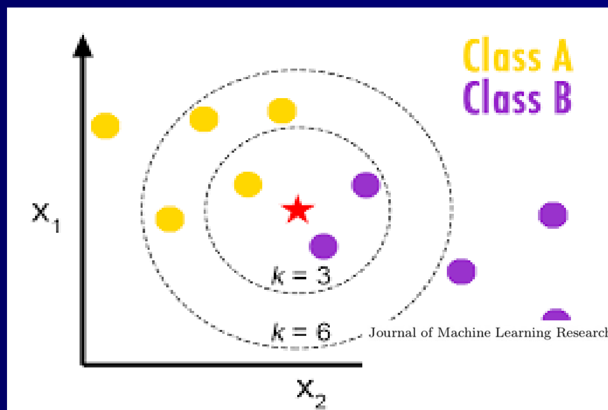


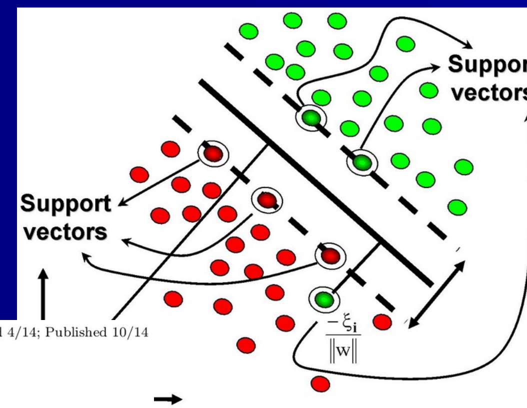






SUPERVISED CLASSIFICATION
- models -



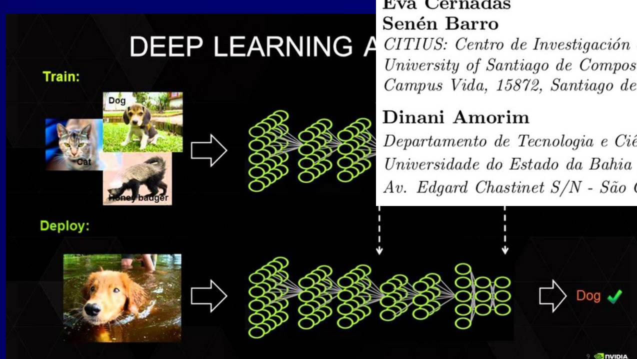


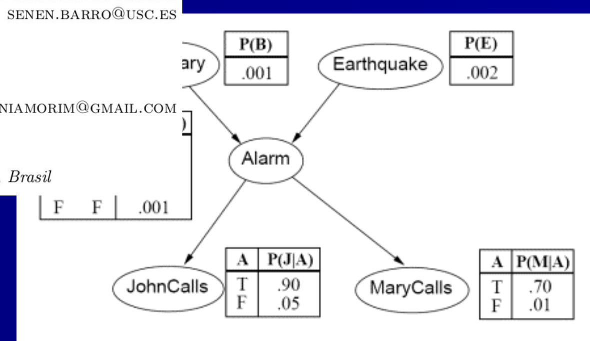
Journal of Machine Learning Research 15 (2014) 3133-3181 Submitted 11/13; Revised 4/14; Published 10/14

Do we Need Hundreds of Classifiers to Solve Real World Classification Problems?

Manuel Fernández-Delgado MANUEL.FERNANDEZ.DELGADO@USC.ES
Eva Cernadas EVA.CERNADAS@USC.ES
Senén Barro SENEN.BARRO@USC.ES
CITIUS: Centro de Investigación en Tecnoloxías da Información da USC
University of Santiago de Compostela
Campus Vida, 15872, Santiago de Compostela, Spain

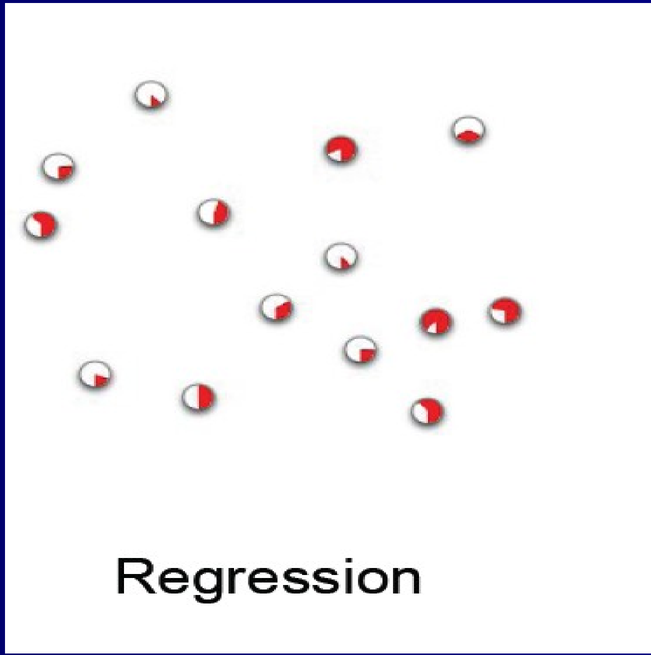
Dinani Amorim DINANIAMORIM@GMAIL.COM
Departamento de Tecnologia e Ciências Sociais- DTCS
Universidade do Estado da Bahia
Av. Edgard Chastinet S/N - São Geraldo - Juazeiro-BA, CEP: 48.305-680, Brasil





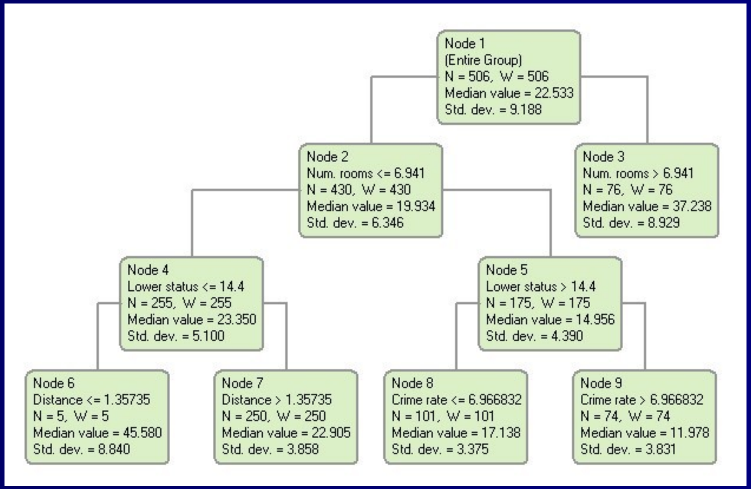
REGRESSION

- The *variable of interest* to be predicted → *quantitative*
- Full supervision in training time

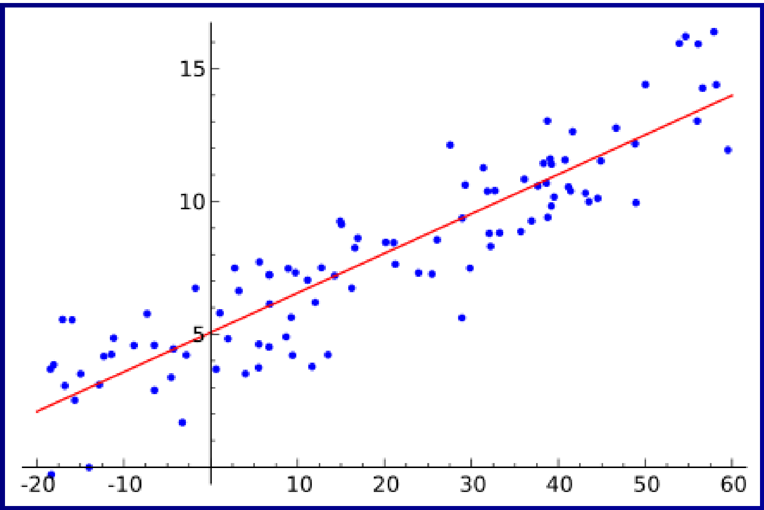


X_1	X_2	...	X_n	C
a	b	...	b	1.1
b	b	...	a	0.1
a	a	...	b	6.3
b	a	...	b	2.0
a	b	...	a	3.3
b	a	...	a	2.1
a	a	...	b	5.5
a	b	...	a	4.0
a	b	...	b	0.5
b	a	...	b	5.6
b	b	...	a	4.3
a	a	...	b	2.2
b	a	...	a	1.2
a	a	...	a	3.1

REGRESSION: models

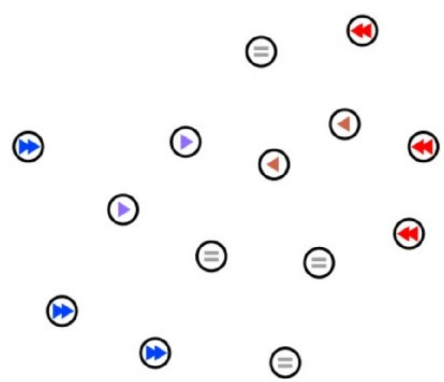


$$y_i = \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i = \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta} + \varepsilon_i, \qquad i = 1, \ldots, n,$$



ORDINAL CLASSIFICATION

- The *variable of interest* to be predicted → *discrete, but ordered*
- Full supervision in training time

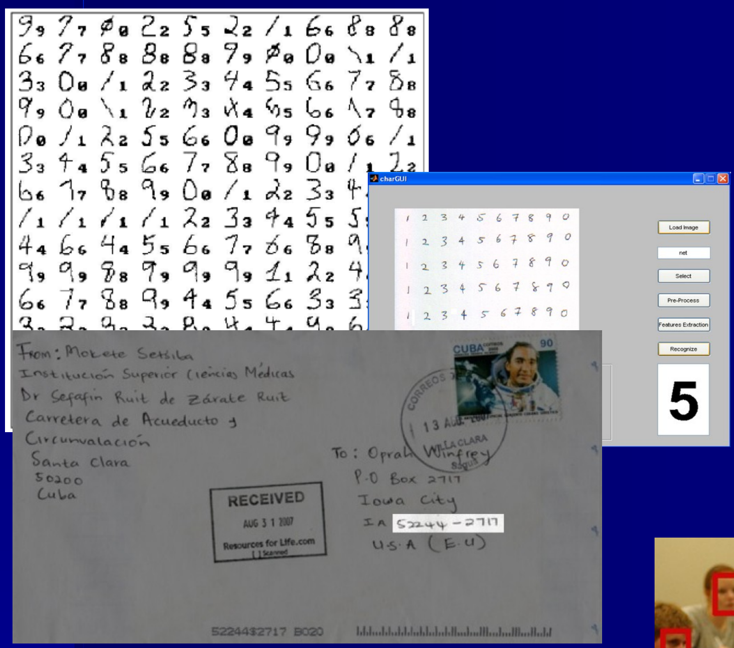


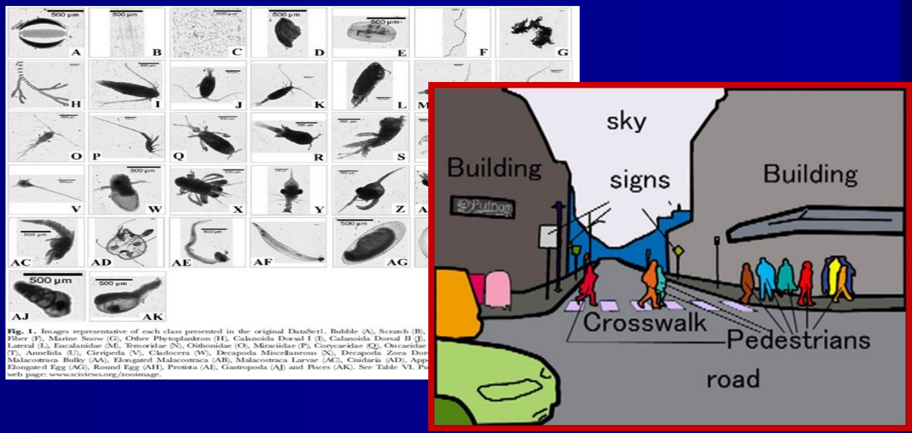
Ordinal learning

X_1 , X_2 , … , X_n	C
a , b , … , b	<<
b , b , … , a	>
a , a , … , b	=
b , a , … , b	<
a , b , … , a	>>
b , a , … , a	>
a , a , … , b	<
a , b , … , a	=
a , b , … , b	=
b , a , … , b	<
b , b , … , a	>>
a , a , … , b	>
b , a , … , a	<<
a , a , … , a	=

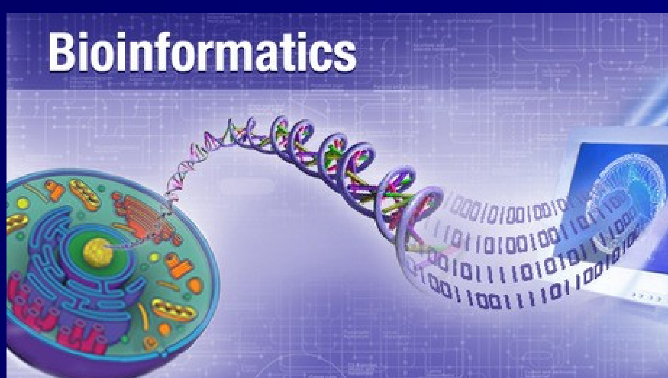
SUPERVISED CLASSIFICATION and REGRESSION: APPLICATIONS

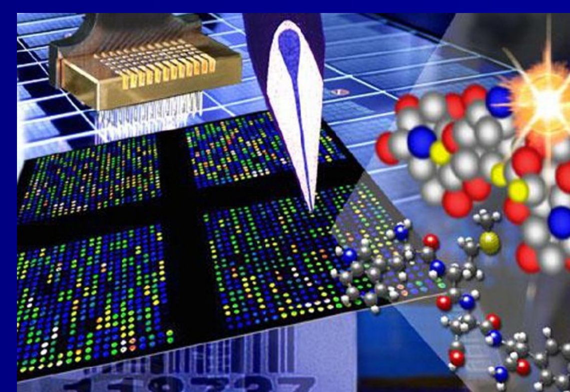
PATTERN RECOGNITION





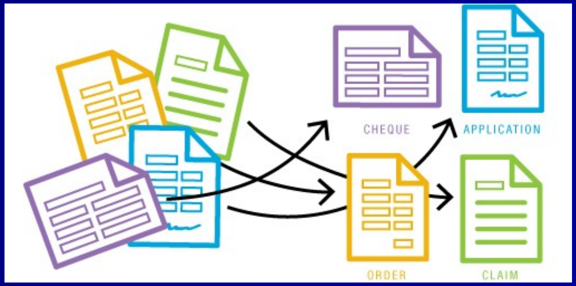
BIOINFORMATICS
DIAGNOSIS AND PROGNOSIS OF DISEASES
BIOMARKER DISCOVERY



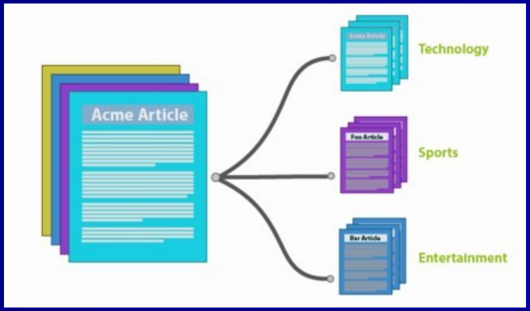


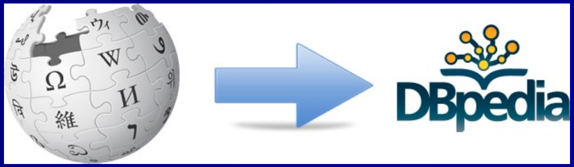
DOCUMENT CLASSIFICATION

- "Natural Language Processing" (NLP)



- Topic - category
- Level of difficulty
- Author's genre

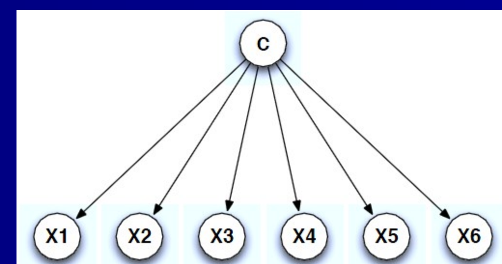


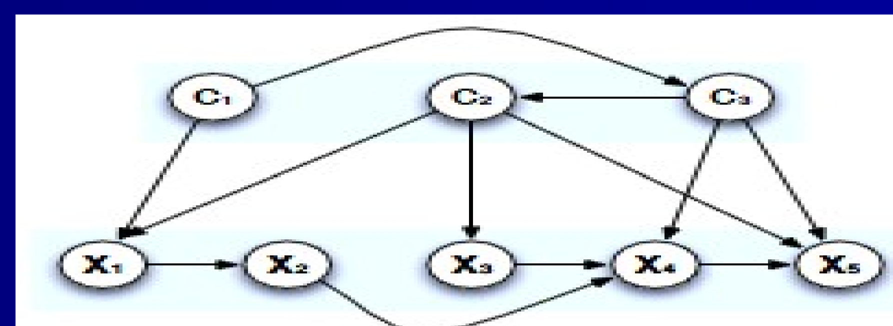


BEYOND SINGLE CLASS VARIABLE… MULTIDIMENSIONAL CLASSIFICATION

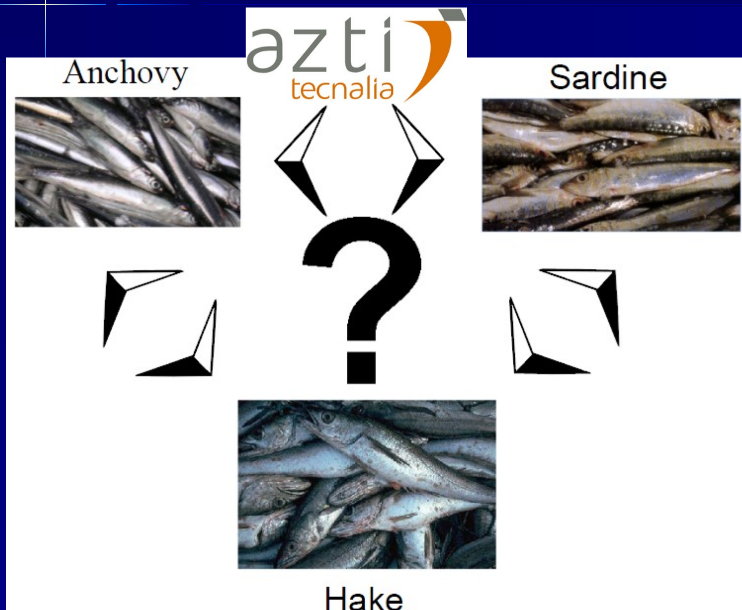
- Several class variables to be jointly predicted
- Learn relationships between class variables
- Full supervision in training time

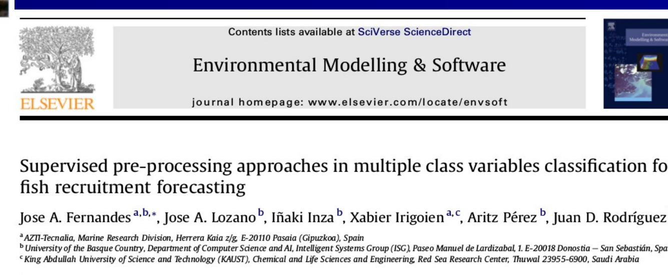
X_1	X_2	…	X_n	C_1	C_2	…	C_m
$x_1^{(1)}$	$x_2^{(1)}$	…	$x_n^{(1)}$	$c_1^{(1)}$	$c_2^{(1)}$	…	$c_m^{(1)}$
$x_1^{(2)}$	$x_2^{(2)}$	…	$x_n^{(2)}$	$c_1^{(2)}$	$c_2^{(2)}$	…	$c_m^{(2)}$
…	…	…	…	…	…	…	…
$x_1^{(N)}$	$x_2^{(N)}$	…	$x_n^{(N)}$	$c_1^{(N)}$	$c_2^{(N)}$	…	$c_m^{(N)}$





MULTIDIMENSIONAL CLASSIFICATION - APPLICATIONS -





Contents lists available at SciVerse ScienceDirect

Environmental Modelling & Software

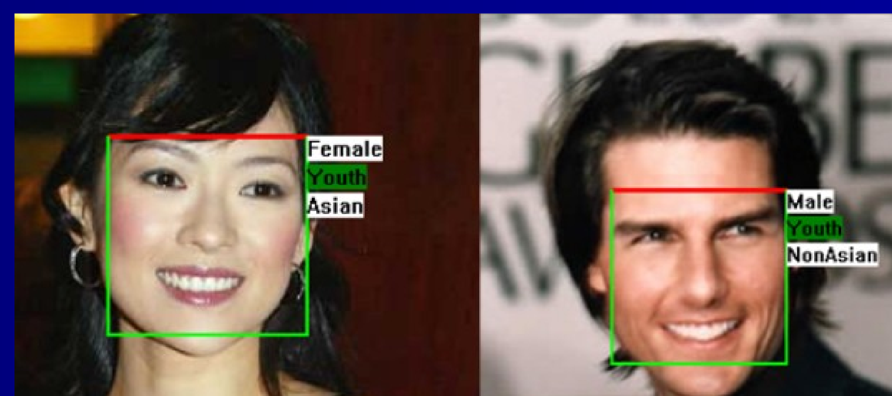
journal homepage: www.elsevier.com/locate/envsoft

ELSEVIER

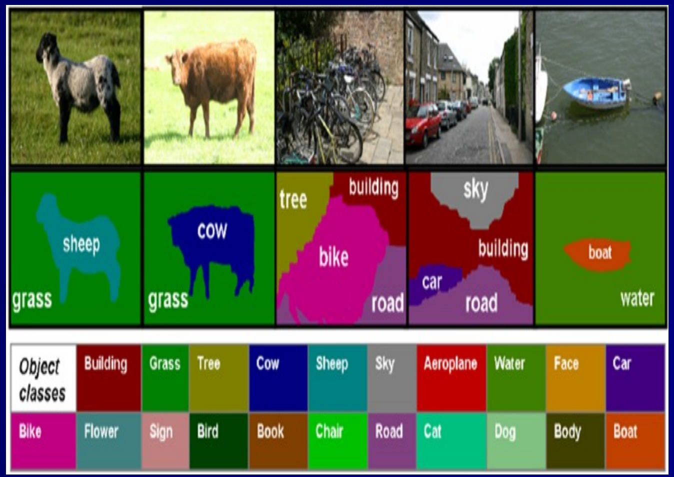
Supervised pre-processing approaches in multiple class variables classification for fish recruitment forecasting

Jose A. Fernandes [a,b,*], Jose A. Lozano [b], Iñaki Inza [b], Xabier Irigoien [a,c], Aritz Pérez [b], Juan D. Rodríguez [b]

[a] AZTI-Tecnalia, Marine Research Division, Herrera Kaia z/g, E-20110 Pasaia (Gipuzkoa), Spain
[b] University of the Basque Country, Department of Computer Science and AI, Intelligent Systems Group (ISG), Paseo Manuel de Lardizabal, 1, E-20018 Donostia – San Sebastián, Spain
[c] King Abdullah University of Science and Technology (KAUST), Chemical and Life Sciences and Engineering, Red Sea Research Center, Thuwal 23955-6900, Saudi Arabia



MULTILABEL CLASSIFICATION

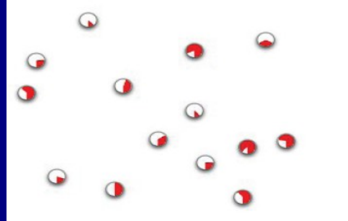


N.	Film	Year	Genre
1	Cadena perpetua	1994	Crime, Drama
2	El padrino	1972	Crime, Drama
3	El padrino. Parte II	1974	Crime, Drama
4	El bueno, el feo y el malo	1966	Adventure, Western
5	Pulp Fiction	1994	Crime, Thriller
6	12 hombres sin piedad	1957	Drama
7	La lista de Schindler	1993	Biography, Drama, History, War
8	El caballero oscuro	2008	Action, Crime, Drama, Thriller
9	El señor de los anillos: El ret. . .	2003	Action, Adventure, Drama, Fantasy
10	El club de la lucha	1999	Drama

X	y1	y2	y3	y4
x1	0	1	1	0
x2	1	0	0	0
x3	0	1	0	0
x4	0	1	1	0
x5	1	1	1	1
x6	0	1	0	0

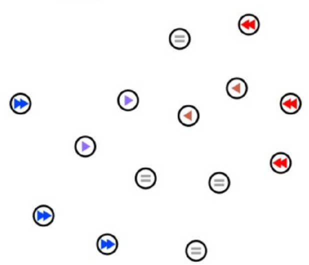
FULL SUPERVISION

X_1	X_2	...	X_n	C_1	C_2	...	C_m
$x_1^{(1)}$	$x_2^{(1)}$	...	$x_n^{(1)}$	$c_1^{(1)}$	$c_2^{(1)}$	...	$c_m^{(1)}$
$x_1^{(2)}$	$x_2^{(2)}$	...	$x_n^{(2)}$	$c_1^{(2)}$	$c_2^{(2)}$	...	$c_m^{(2)}$
...	...	...	...	...	...	...	...
$x_1^{(N)}$	$x_2^{(N)}$	...	$x_n^{(N)}$	$c_1^{(N)}$	$c_2^{(N)}$	...	$c_m^{(N)}$



Regression

X_1, X_2, ..., X_n	C
a , b , ... , b	1.1
b , b , ... , a	0.1
a , a , ... , b	6.3
b , a , ... , b	2.0
a , b , ... , a	3.3
b , a , ... , a	2.1
a , a , ... , b	5.5
a , b , ... , a	4.0
a , b , ... , b	0.5
b , a , ... , b	5.6
b , b , ... , a	4.3
a , a , ... , b	2.2
b , a , ... , a	1.2
a , a , ... , a	3.1



Ordinal learning

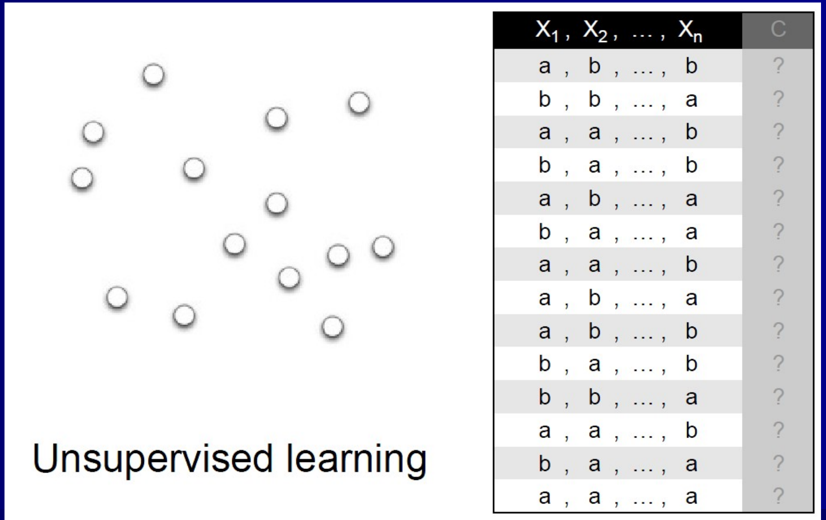
X_1, X_2, ..., X_n	C
a , b , ... , b	<<
b , b , ... , a	>
a , a , ... , b	=
b , a , ... , b	<
a , b , ... , a	>>
b , a , ... , a	>
a , a , ... , b	<
a , b , ... , a	=
a , b , ... , b	=
b , a , ... , b	<
b , b , ... , a	>>
a , a , ... , b	>
b , a , ... , a	<<
a , a , ... , a	=

X	y1	y2	y3	y4
x1	0	1	1	0
x2	1	0	0	0
x3	0	1	0	0
x4	0	1	1	0
x5	1	1	1	1
x6	0	1	0	0

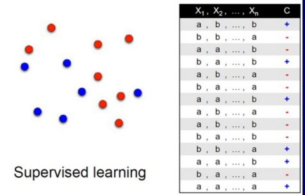
UNSUPERVISED CLASSIFICATION CLUSTERING

- Given a collection of records-samples (*training set*)
  - Each record → a set of *attributes-features-~~predictors~~*
  - *No "target feature" (class)* which supervises the learning process

- Groups of cases:
  - Large intra-group ~ homogeneity
  - Large inter-groups ~ heterogeneity

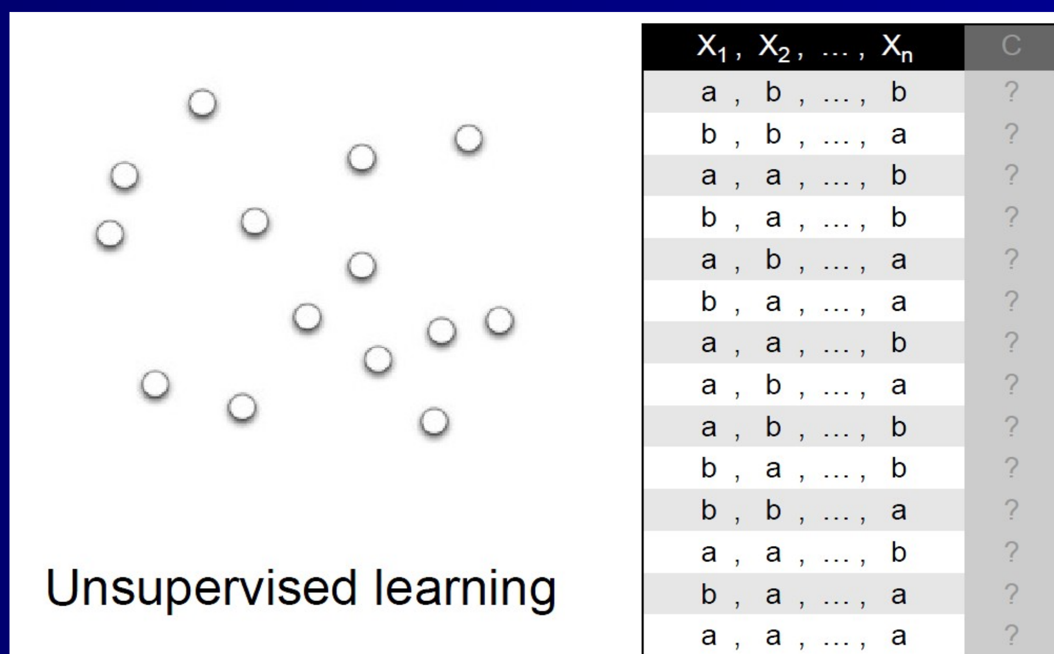


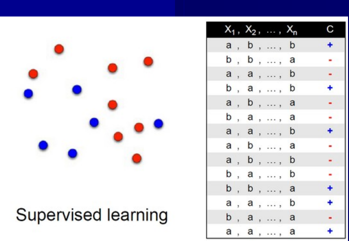
X_1, X_2, …, X_n	C
a , b , … , b	?
b , b , … , a	?
a , a , … , b	?
b , a , … , b	?
a , b , … , a	?
b , a , … , a	?
a , a , … , b	?
a , b , … , a	?
a , b , … , b	?
b , a , … , b	?
b , b , … , a	?
a , a , … , b	?
b , a , … , a	?
a , a , … , a	?



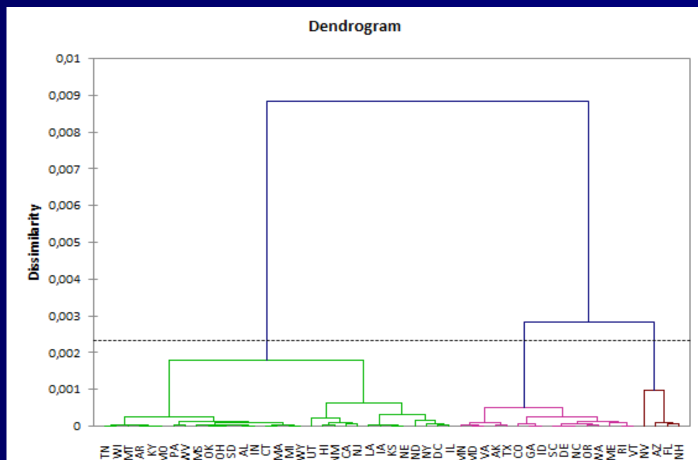
UNSUPERVISED CLASSIFICATION CLUSTERING

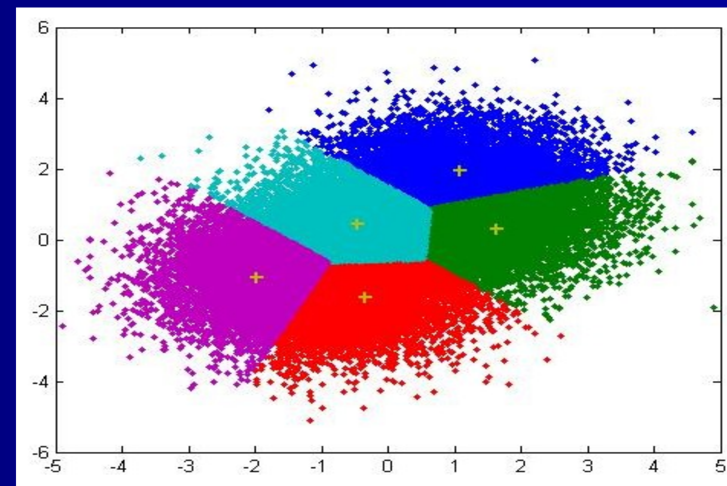
- Difficult evaluation-measure of these properties --› no recognition rate
- Number of groups → deciding before-hand → difficult decision
- “Distance”-”similarity” function

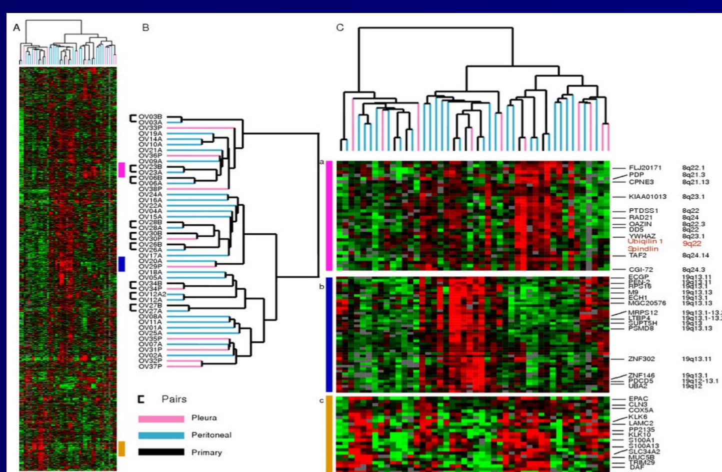


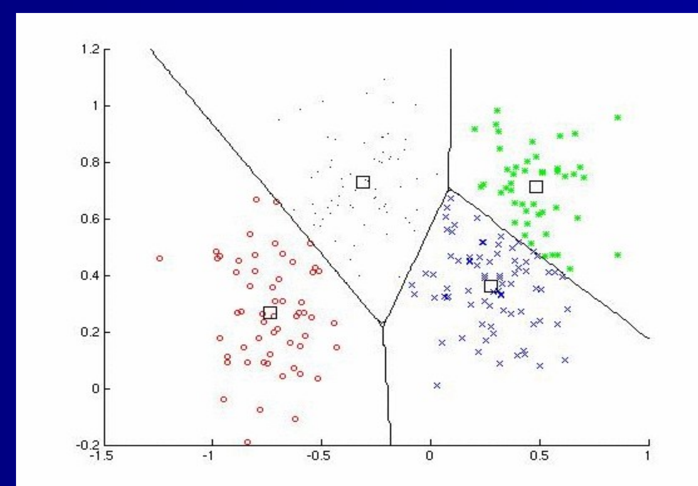


CLUSTERING: MODELS









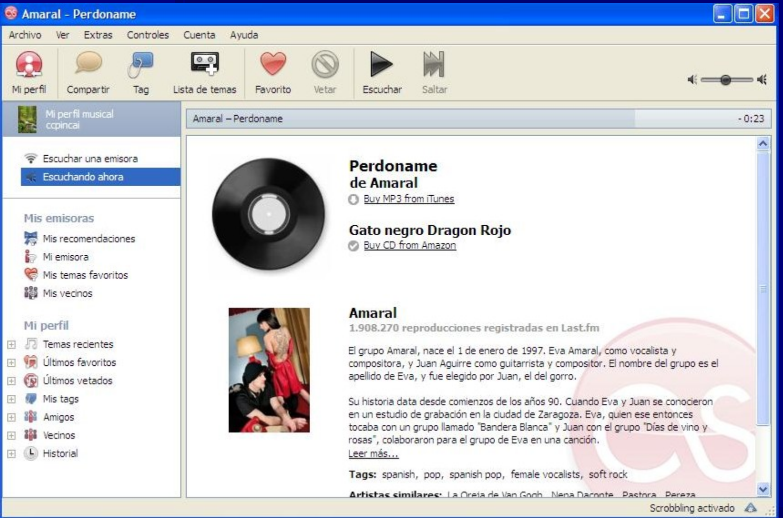
CLUSTERING: APPLICATIONS CUSTOMER SEGMENTATION

- Identify micro-markets and develop policies for each
- Targeted marketing
- Similar customers are grouped in the same cluster

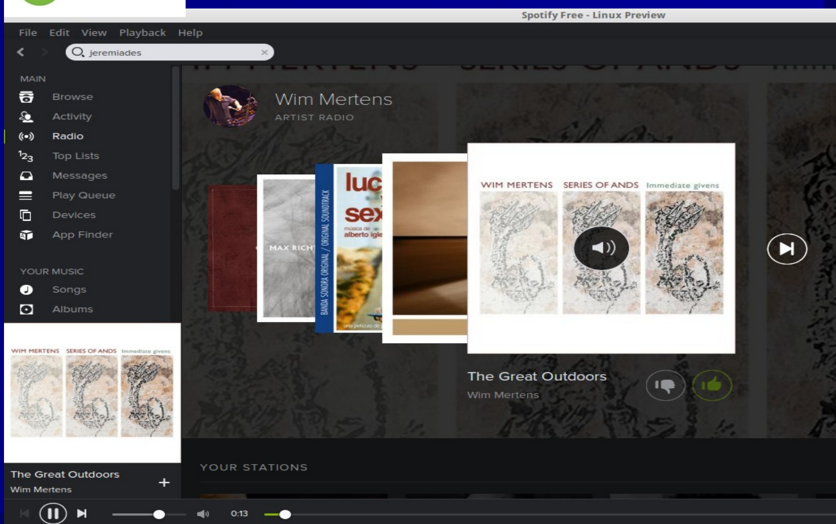
COLLABORATIVE FILTERING RECOMMENDER SYSTEMS

Movie	Alice (1)	Bob (2)	Carol (3)	Dave (4)
Love at last	5	5	0	0
Romance forever	5	?	?	0
Cute puppies of love	?	4	0	?
Nonstop car chases	0	0	5	4
Swords vs. karate	0	0	5	?

last.fm

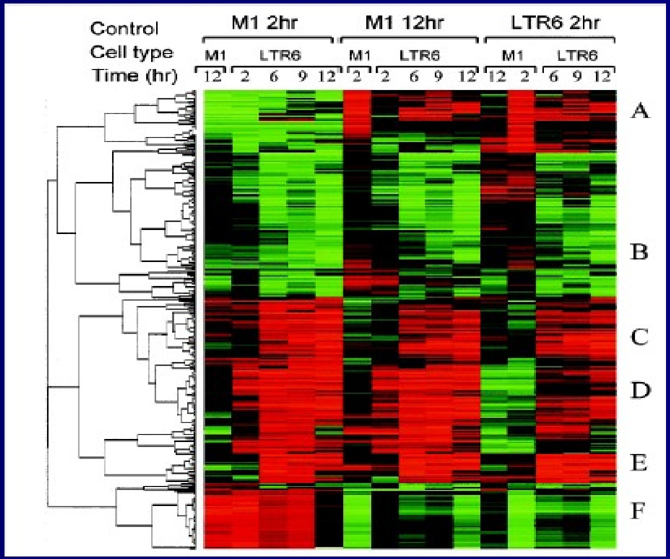


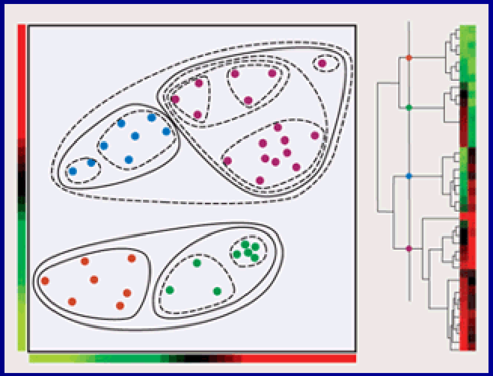
Spotify



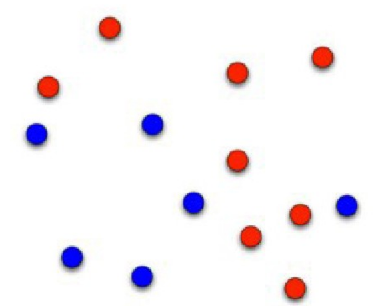
GENE EXPRESSION BI-CLUSTERING

- Find genes with similar expression profiles ~ a way to infer the function of genes whose function is unknown
- Biclustering… a classic concept in fashion again:
  - Finding a subgroup of samples with a similar pattern in a subgroup of variables (not in all the variables)



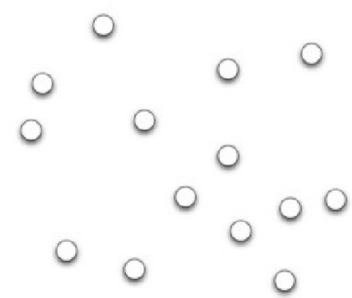


IS THERE SOMEONE IN THE MIDDLE?



X_1	X_2	…	X_n	C
a	b	…	b	+
b	b	…	a	-
a	a	…	b	-
b	a	…	b	+
a	b	…	a	-
b	a	…	a	-
a	a	…	b	+
a	b	…	a	-
a	b	…	b	-
b	a	…	b	-
b	b	…	a	+
a	a	…	b	+
b	a	…	a	-
a	a	…	a	+

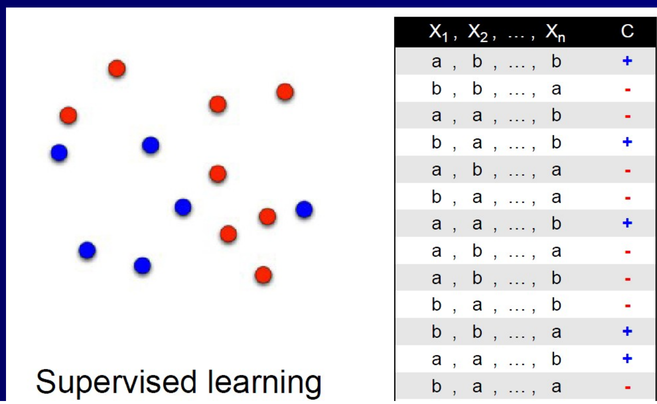
Supervised learning

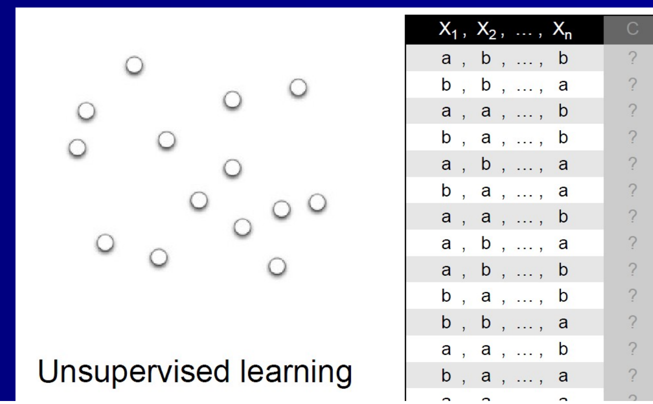


X_1	X_2	…	X_n	C
a	b	…	b	?
b	b	…	a	?
a	a	…	b	?
b	a	…	b	?
a	b	…	a	?
b	a	…	a	?
a	a	…	b	?
a	b	…	a	?
a	b	…	b	?
b	a	…	b	?
b	b	…	a	?
a	a	…	b	?
b	a	…	a	?
a	a	…	a	?

Unsupervised learning

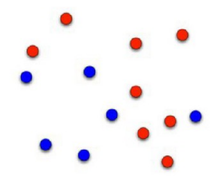
IS THERE SOMEONE IN THE MIDDLE?





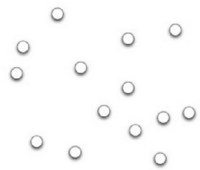
- Hidden big data. Large quantities of useful data are in fact useless because they are untagged, file-based, and unstructured. The 2012 IDC study on big data [117] explained that, in 2012, 23% (643 exabytes) of the digital universe would be useful for big data if tagged and analyzed. However, at that time only 3% of the potentially useful data was tagged, and even less was analyzed. The figures have probably gotten worse in recent years. The Open Data and Semantic Web movements have emerged, in part, to make us aware and improve on this situation. No comments

IS THERE SOMEONE IN THE MIDDLE?



Supervised learning

X_1 , X_2 , … , X_n	C
a , b , … , b	+
b , b , … , a	-
a , a , … , b	-
b , a , … , b	+
a , b , … , a	-
b , a , … , a	-
a , a , … , b	+
a , b , … , a	-
a , b , … , b	-
b , a , … , b	-
b , b , … , a	+
a , a , … , b	+
b , a , … , a	-
a , a , … , a	+



Unsupervised learning

X_1 , X_2 , … , X_n	C
a , b , … , b	?
b , b , … , a	?
a , a , … , b	?
b , a , … , b	?
a , b , … , a	?
b , a , … , a	?
a , a , … , b	?
a , b , … , a	?
a , b , … , b	?
b , a , … , b	?
b , b , … , a	?
a , a , … , b	?
b , a , … , a	?
a , a , … , a	?

Pattern Recognition Letters 69 (2016) 49–55

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

ELSEVIER

Weak supervision and other non-standard classification problems: A taxonomy☆

Jerónimo Hernández-González*, Iñaki Inza, Jose A. Lozano

Intelligent Systems Group, University of the Basque Country UPV/EHU, P. Manuel Lardizabal 1, 20018 Donostia, Spain

ARTICLE INFO

Article history:
Received 10 May 2015
Available online 24 October 2015

Keywords:
Weakly supervised classification
Partially supervised classification
Degrees of supervision

ABSTRACT

In recent years, different researchers in the machine learning community have presented new classification frameworks which go beyond the standard supervised classification in different aspects. Specifically, a wide spectrum of novel frameworks that use partially labeled data in the construction of classifiers has been studied. With the objective of drawing up a description of the state-of-the-art, three identifying characteristics of these novel frameworks have been considered: (1) the relationship between instances and labels of a problem, which may be beyond the *one-instance one-label* standard, (2) the possible provision of partial class information for the training examples, and (3) the possible provision of partial class information also for the examples in the prediction stage. These three ideas have been formulated as axes of a comprehensive taxonomy that organizes the state-of-the-art. The proposed organization allows us both to understand similarities/differences among the different classification problems already presented in the literature as well as to discover unexplored frameworks that might be seen as further challenges and research opportunities. A representative set of state-of-the-art problems has been used to illustrate the novel taxonomy and support the discussion.

International Journal of Approximate Reasoning 150 (2022) 258–272

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

ELSEVIER

On the relative value of weak information of supervision for learning generative models: An empirical study

Jerónimo Hernández-González [a,*], Aritz Pérez [b]

[a] *Departament de Matemàtiques i Informàtica, Universitat de Barcelona (UB), Gran Via de les Corts Catalanes 585, Barcelona, Spain*
[b] *Basque Center for Applied Mathematics, Al. Mazarredo 14, Bilbao, Spain*

ARTICLE INFO

Article history:
Received 2 July 2022
Received in revised form 8 August 2022
Accepted 22 August 2022
Available online 31 August 2022

Keywords:
Weak supervision
Model learning
Generative models
Empirical study

ABSTRACT

Weakly supervised learning is aimed to learn predictive models from partially supervised data, an easy-to-collect alternative to the costly standard full supervision. During the last decade, the research community has striven to show that learning reliable models in specific weakly supervised problems is possible. We present an empirical study that analyzes the value of weak information of supervision throughout its entire spectrum, from none to full supervision. Its contribution is assessed under the realistic assumption that a small subset of fully supervised data is available. Particularized in the problem of learning with candidate sets, we adapt Cozman and Cohen [1] key study to learning from weakly supervised data. Standard learning techniques are used to infer generative models from this type of supervision with both synthetic and real data. Empirical results suggest that weakly labeled data is helpful in realistic scenarios, where fully labeled data is scarce, and its contribution is directly related to both the amount of information of supervision and how meaningful this information is.

THE TERM
– "WEAKLY SUPERVISED LEARNING" –
- RESEARCH OPORTUNITIES -

- GoogleScholar – number of "search results":
  - Since 2015 → 24,600
  - Since 2018 → 21,200
  - Since 2020 → 18,900

WILDCAT: Weakly Supervised Learning of Deep ConvNets for Image Classification, Pointwise Localization and Segmentation

Thibaut Durand[(1)]*, Taylor Mordan[(1,2)]*, Nicolas Thome[(3)], Matthieu Cord[(1)]
(1) Sorbonne Universités, UPMC Univ Paris 06, CNRS, LIP6 UMR 7606, 4 place Jussieu, 75005 Paris
(2) Thales Optronique S.A.S., 2 Avenue Gay Lussac, 78990 Élancourt, France
(3) CEDRIC - Conservatoire National des Arts et Métiers, 292 rue St Martin, 75003 Paris, France
{thibaut.durand, taylor.mordan, nicolas.thome, matthieu.cord}@lip6.fr

A Brief Introduction to Weakly Supervised Learning

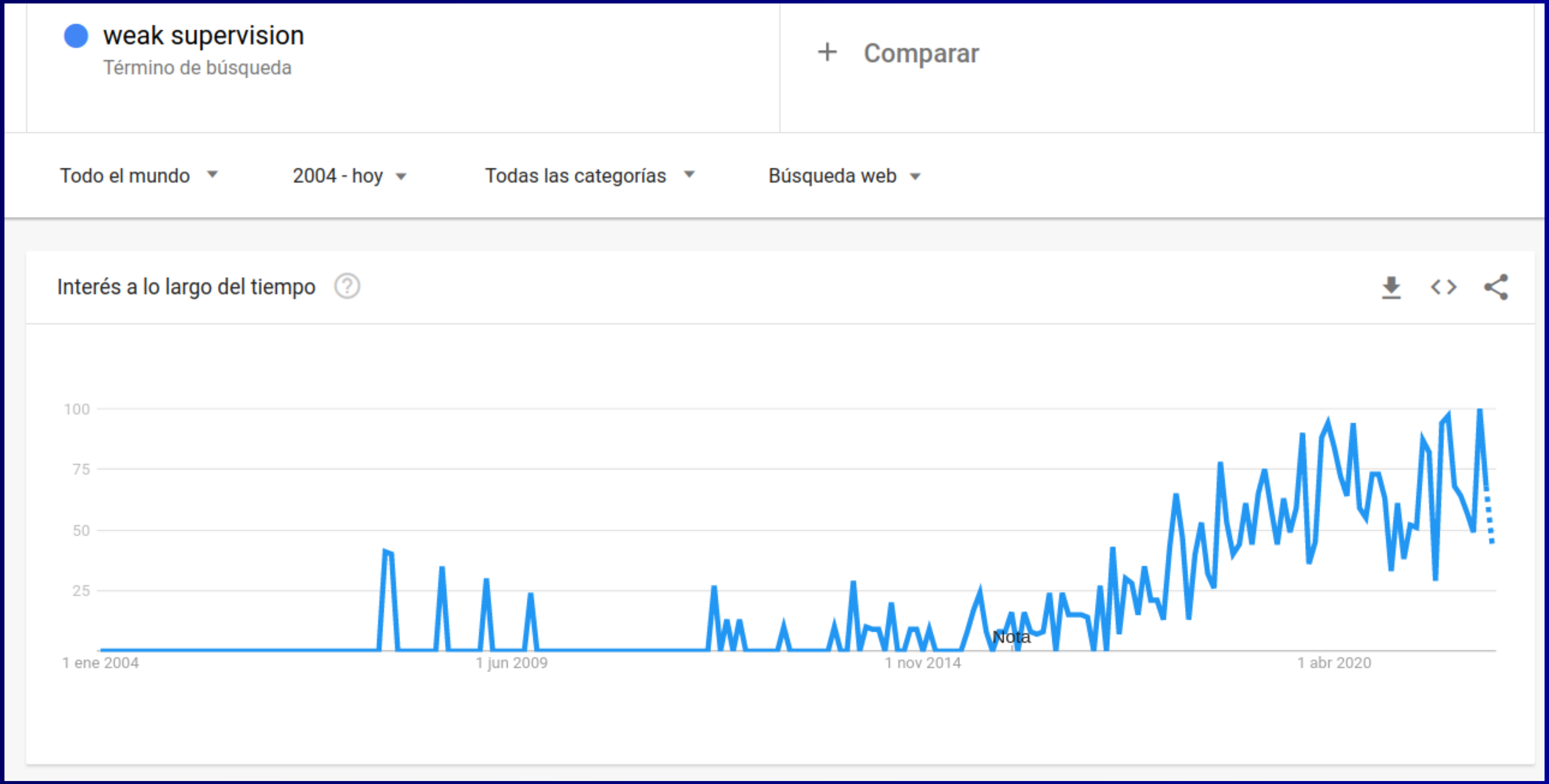
Zhi-Hua Zhou
National Key Laboratory for Novel Software Technology
Nanjing University, Nanjing 210023, China
zhouzh@nju.edu.cn

Jain *et al. BMC Bioinformatics* 2015, **17**(Suppl 1):1
DOI 10.1186/s12859-015-0844-1
BMC Bioinformatics
PROCEEDINGS Open Access

Weakly supervised learning of biomedical information extraction from curated data

Suvir Jain[1†], Kashyap R.[1†], Tsung-Ting Kuo[2], Shitij Bhargava[1], Gordon Lin[1] and Chun-Nan Hsu[2*]

THE TERM – GOOGLE TRENDS

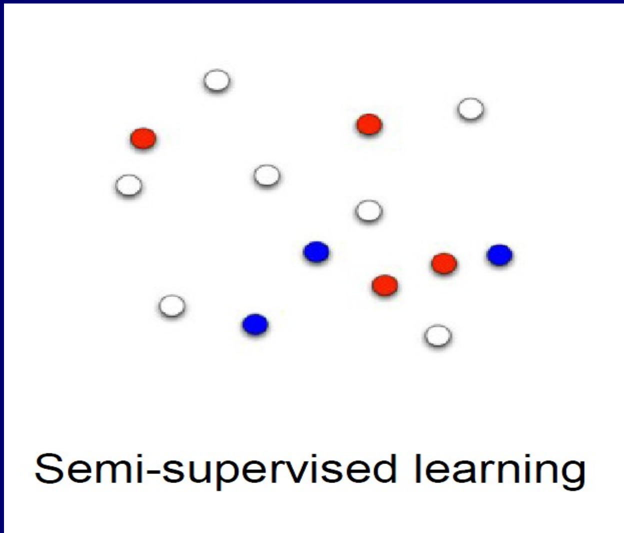


SEMI SUPERVISED LEARNING

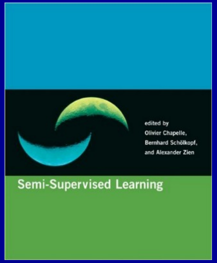
- Most of the samples do not show a class value. Why?
  - Categorization: human-time consuming task
  - No knowledge to categorize the samples

- Objective → learn a supervised model

- Can a learning process which takes advantage of unlabeled samples, construct a better supervised classification model?

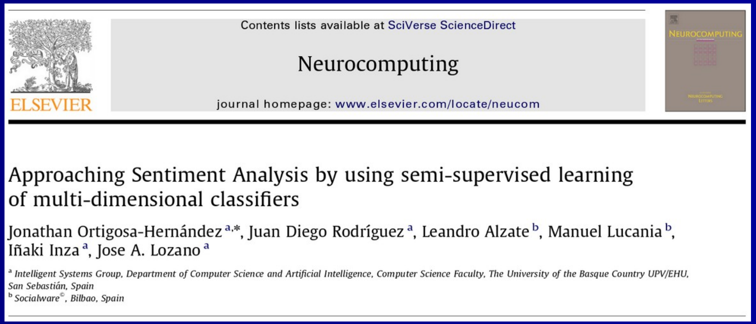


X_1	X_2	...	X_n	C
a	b	...	b	+
b	b	...	a	-
a	a	...	b	-
b	a	...	b	+
a	b	...	a	-
b	a	...	a	-
a	a	...	b	+
a	b	...	a	?
a	b	...	b	?
b	a	...	b	?
b	b	...	a	?
a	a	...	b	?
b	a	...	a	?
a	a	...	a	?



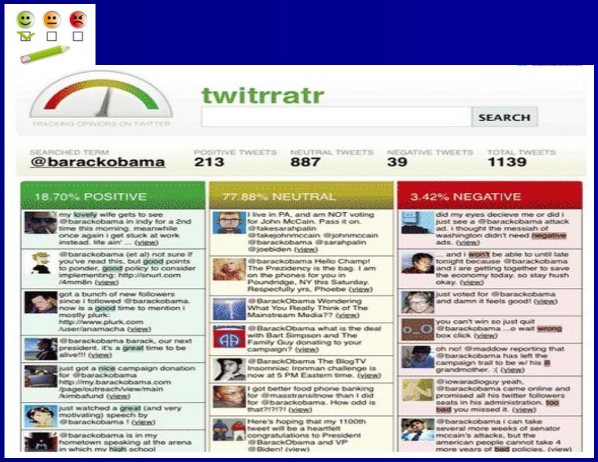
SEMI SUPERVISED LEARNING SENTIMENT ANALYSIS

- Companies: reputation
- Opinions about its products:
  - social networks
  - blogs
  - forums...
- Automatically classify the written opinion: {+, -, neutral}
- NLP: “Natural Language Processing”


Contents lists available at SciVerse ScienceDirect
Neurocomputing
journal homepage: www.elsevier.com/locate/neucom

Approaching Sentiment Analysis by using semi-supervised learning of multi-dimensional classifiers

Jonathan Ortigosa-Hernández [a,*], Juan Diego Rodríguez [a], Leandro Alzate [b], Manuel Lucania [b], Iñaki Inza [a], Jose A. Lozano [a]
[a] Intelligent Systems Group, Department of Computer Science and Artificial Intelligence, Computer Science Faculty, The University of the Basque Country UPV/EHU, San Sebastián, Spain
[b] Socialware©, Bilbao, Spain



SEMI SUPERVISED LEARNING

Machine Learning, 39, 103–134, 2000.

Text Classification from Labeled and Unlabeled Documents using EM

KAMAL NIGAM knigam@cs.cmu.edu
School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA

ANDREW KACHITES MCCALLUM mccallum@justresearch.com
Just Research, 4616 Henry Street, Pittsburgh, PA 15213, USA; School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA

SEBASTIAN THRUN thrun@cs.cmu.edu
TOM MITCHELL tom.mitchell@cmu.edu
School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Int. J. Mach. Learn. & Cyber. (2017) 8:355–370
DOI 10.1007/s13042-015-0328-7

ORIGINAL ARTICLE

Semi-supervised self-training for decision tree classifiers

Jafar Tanha · Maarten van Someren · Hamideh Afsarmanesh

Combining Labeled and Unlabeled Data with Co-Training*†

Avrim Blum
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213-3891
avrim+@cs.cmu.edu

Tom Mitchell
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213-3891
mitchell+@cs.cmu.edu

AMBIGUOUS TRAINING DATA

- “Positive” + “Negative” + Unlabeled
- Annotated by expert

- Unlabeled → “Ambiguous” samples
- Difficult to label by expert

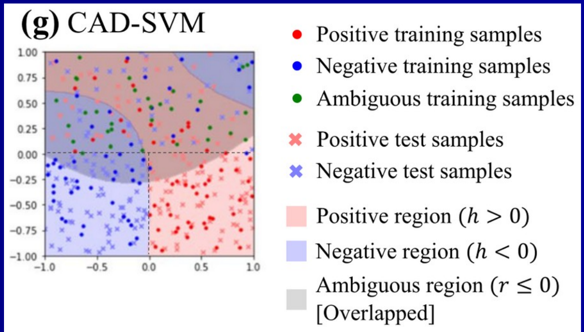
- “Ambiguous” → “Unlabelled” + Semi-Supervised learning ????
- Not semi-supervised scenario !!
- Unlabeled instances → not uniformly distributed !!

X_1 , X_2 , ... , X_n	C
a , b , ... , b	+
b , b , ... , a	-
a , a , ... , b	-
b , a , ... , b	+
a , b , ... , a	-
b , a , ... , a	-
a , a , ... , b	+
a , b , ... , a	?
a , b , ... , b	?
b , a , ... , b	?
b , b , ... , a	?
a , a , ... , b	?
b , a , ... , a	?
a , a , ... , a	?

Machine Learning (2020) 109:2369–2388
https://doi.org/10.1007/s10994-020-05915-2

Binary classification with ambiguous training data

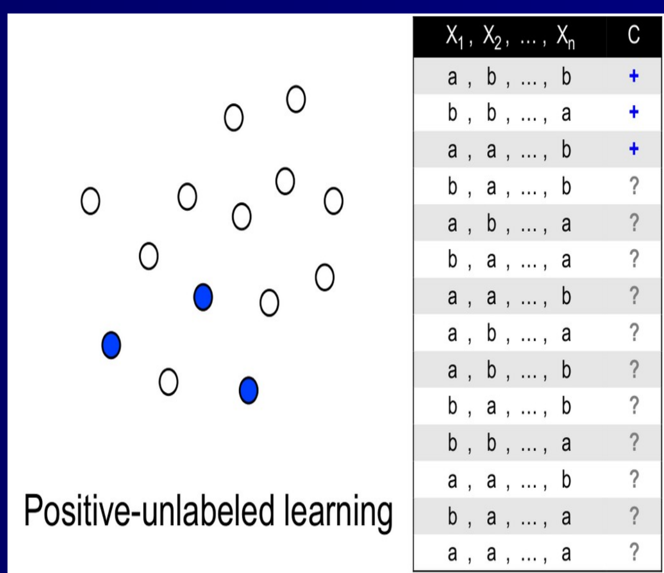
Naoya Otani[1] · Yosuke Otsubo[1] · Tetsuya Koike[1] · Masashi Sugiyama[2,3]



POSITIVE UNLABELED LEARNING

- More difficult than semi-supervised classification
- Prediction: “+” or “-”

- Application → prediction of genes related to cancer
- Web page visiting prediction → personalized ads



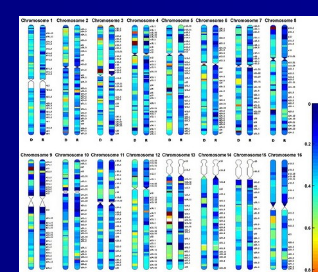
$X_1, X_2, \ldots, X_n$	C
a , b , ... , b	+
b , b , ... , a	+
a , a , ... , b	+
b , a , ... , b	?
a , b , ... , a	?
b , a , ... , a	?
a , a , ... , b	?
a , b , ... , a	?
a , b , ... , b	?
b , a , ... , b	?
b , b , ... , a	?
a , a , ... , b	?
b , a , ... , a	?
a , a , ... , a	?


Published online 18 August 2008 — Nucleic Acids Research, 2008, Vol. 36, No. 18 e115
doi:10.1093/nar/gkn482

Prioritization of candidate cancer genes—an aid to oncogenomic studies

Simon J. Furney[1], Borja Calvo[2], Pedro Larrañaga[3], Jose A. Lozano[2] and Nuria Lopez-Bigas[1,*]

[1]Research Unit on Biomedical Informatics, Experimental and Health Science Department, Universitat Pompeu Fabra, Barcelona 08080, [2]Intelligent Systems Group, Department of Computer Science and Artificial Intelligence, University of the Basque Country, Donostia-San Sebastián 20018 and [3]Department of Artificial Intelligence, Technical University of Madrid, Boadilla del Monte 28660, Spain

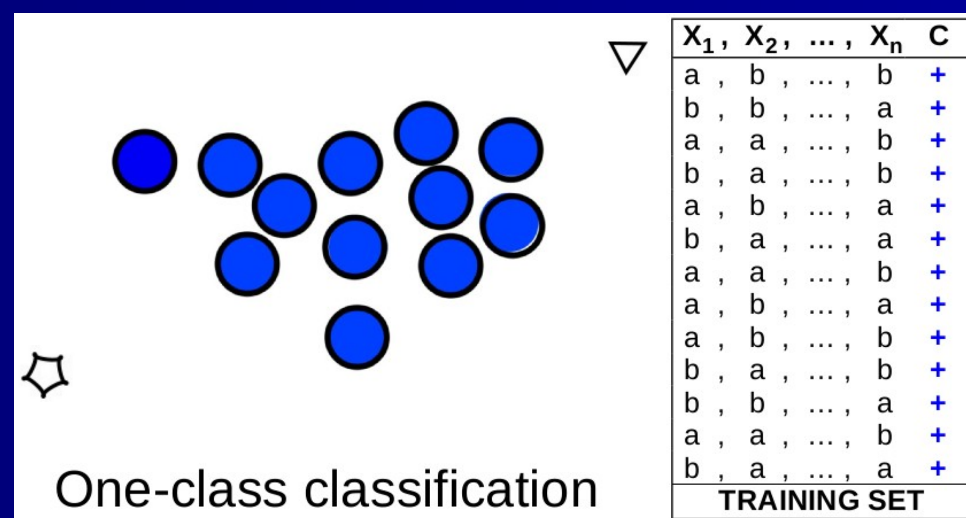


ONE CLASS CLASSIFICATION – OUTLIER DETECTION –

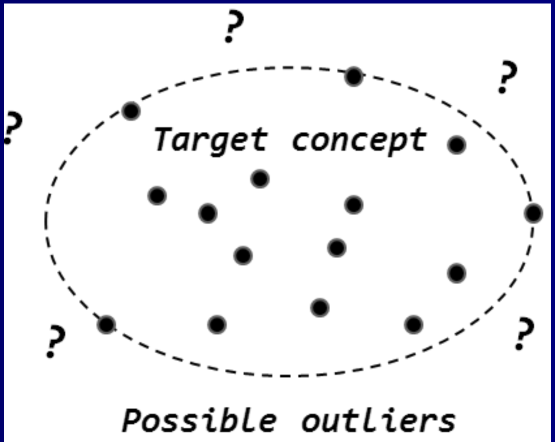
- One category: forms a representative sample
- Only "normal behaviour" samples in training time

- Training phase: model the "normal" behaviour
- Prediction phase → detect "deviations" from the "normal" model

- Model the "dominant" class + "isolate" outliers in "operation phase"

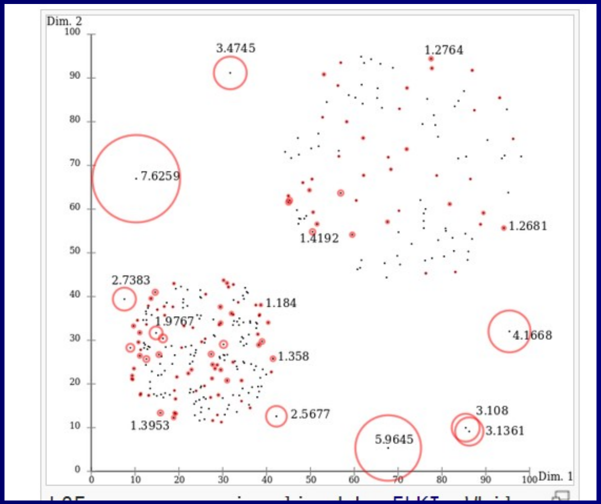


ONE CLASS CLASSIFICATION – OUTLIER DETECTION –



OneClass SVM
AutoEncoders

1-Class data

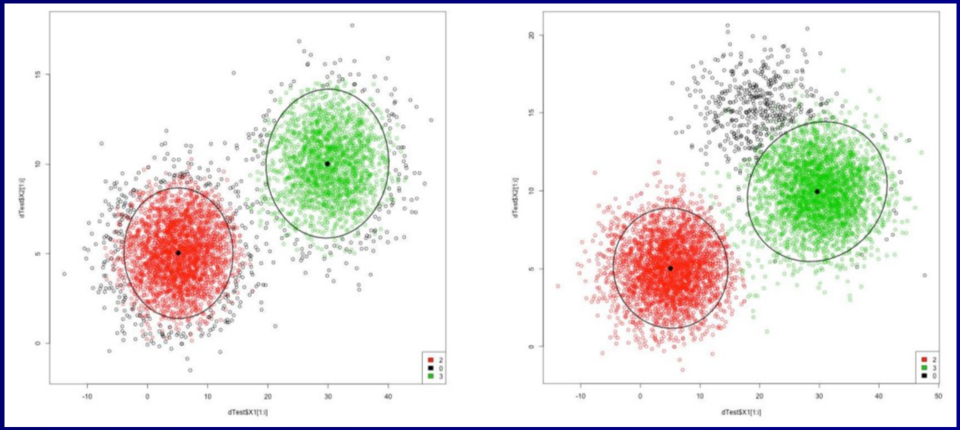


Local Outlier Factor
Isolation Forests

MultiClass data

NOVELTY DETECTION

- Initially labeled dataset → train a model
- Unlabeled samples arrive → in 2nd dataset – or streaming
- 2nd dataset → an "emergent" class appears?
- "Novel class"? → "detect + baptise"
- Separation + cohesion
- Re-train the model with the "baptised class" samples



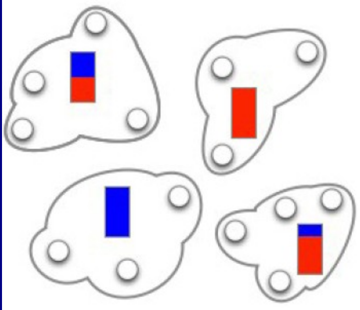
LEARNING with LABEL PROPORTIONS

X_1	X_2	…	X_n	C
a	b	…	b	0.5 0.5
b	b	…	a	
a	a	…	b	
b	a	…	b	
a	b	…	a	0.0 1.0
b	a	…	a	
a	b	…	a	
a	a	…	b	0.25 0.75
a	b	…	b	
b	a	…	b	
b	a	…	a	
a	a	…	b	1.0 0.0
b	b	…	a	
a	a	…	a	

Supervised Learning by Training on Aggregate Outputs

David R. Musicant, Robert Atlas, Janara M. Christensen, Jamie F. Olson, Jeffrey M. Rzeszotarski, Emma R. D. Turetsky

Abstract—Supervised learning is a classic data mining problem where one wishes to be able to predict an output value associated with a particular input vector. We present a new twist on this classic problem where, instead of having the training set contain an individual output value for each input vector, the output values in the training set are only given in aggregate over a number of input vectors. This new problem arose from a particular need in learning on mass spectrometry data, but could easily apply to situations when data has been aggregated in order to maintain privacy. We provide a formal description of this new problem for both classification and regression. We then examine how k-nearest neighbor, neural networks, support vector machines, and decision trees can be adapted for this problem.



LABEL PROPORTIONS – APPLICATIONS –

Embryo selection in Assisted Reproductive Technologies (ART)

Two steps:

- **Transfer**: step in which one or several embryos are placed into the uterus of the patient.
- **Implantation**: step in which pregnancy is established (by one or several embryos).

Application	MILlp problem
Transfered embryos	Dataset
Implanted or not	Class labels
ART process	Bag
Number of children	Label proportions

SMMR STATISTICAL METHODS IN MEDICAL RESEARCH

Article

Statistical Methods in Medical Research
0(0) 1–11
© The Author(s) 2016
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/0962280216651098
smm.sagepub.com
SAGE

Fitting the data from embryo implantation prediction: Learning from label proportions

Jerónimo Hernández-González,[1] Iñaki Inza,[1] Lorena Crisol-Ortíz,[2] María A Guembe,[2] María J Iñarra[2] and Jose A Lozano[1,3]

Information Sciences 481 (2019) 381–393

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

ELSEVIER

Aggregated outputs by linear models: An application on marine litter beaching prediction

Check for updates

Jerónimo Hernández-González[a,*], Iñaki Inza[a], Igor Granado[b], Oihane C. Basurko[b], Jose A. Fernandes[b], Jose A. Lozano[a,c]

[a] Department of Computer Science and Artificial Intelligence, University of the Basque Country UPV/EHU, Donostia, Spain
[b] Marine Research Division at AZTI-Tecnalia, Pasaia, Spain
[c] Basque Center for Applied Mathematics, Bilbao, Spain

2017 IEEE International Conference on Data Mining

A Probabilistic Approach for Learning with Label Proportions Applied to the US Presidential Election

Tao Sun[1], Dan Sheldon[1,2], Brendan O'Connor[1]
[1]College of Information and Computer Sciences, University of Massachusetts Amherst
[2]Department of Computer Science, Mount Holyoke College
Email: {taosun, sheldon, brenocon}@cs.umass.edu

Possible voters based on previous election results

- It involves any situation related with an election that can be organised as follows:

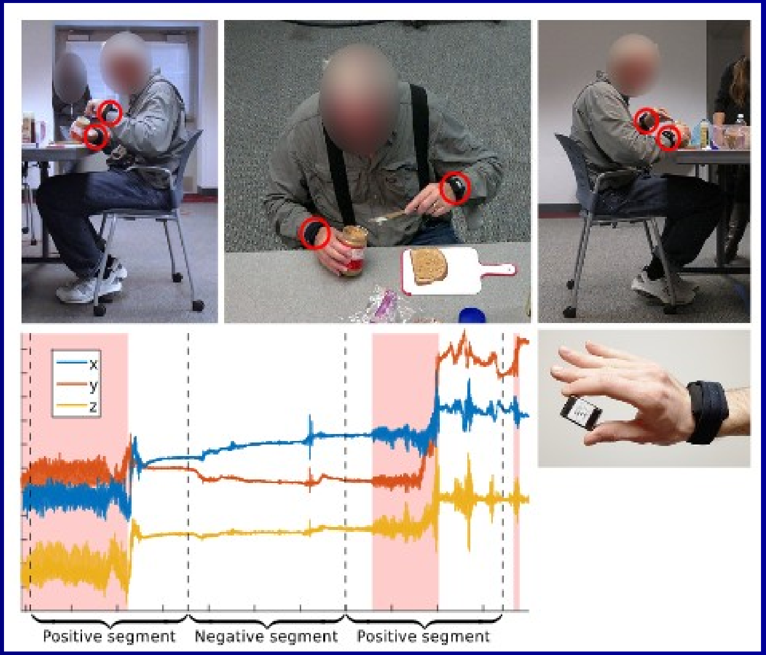
Application	MILlp problem
Census	Dataset
Candidates	Class labels
Polling station	Bag
Election results	Label proportions

STRATIFIED LEARNING

- Approximate proportion of each label is provided
- Labeling → intervals of labels' proportions

Weakly-supervised Learning for Parkinson's Disease Tremor Detection

Ada Zhang[1], Alexander Cebulla[2], Stanislav Panev[1], Jessica Hodgins[1], and Fernando De la Torre[1]
[1] Robotics Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA
[2] ETH Zurich, Switzerland



CLASSIFICATION WITH PARTIAL LABELS

Each instance comes annotated with several class labels but only one of them is valid.

Journal of Machine Learning Research 12 (2011) 1501-1536 Submitted 10/10; Revised 2/11; Published 5/11

Learning from Partial Labels

Timothee Cour TIMOTHEE.COUR@GMAIL.COM
NEC Laboratories America
10080 N Wolfe Rd # Sw3350
Cupertino, CA 95014, USA

Benjamin Sapp BENSAPP@CIS.UPENN.EDU
Ben Taskar TASKAR@SEAS.UPENN.EDU
Department of Computer and Information Science
University of Pennsylvania
3330 Walnut Street
Philadelphia, PA 19107, USA

X_1 , X_2 , … , X_n	C
a , b , … , b	a,b,c
b , b , … , a	a,c
a , a , … , b	d
b , a , … , b	b,c
a , b , … , a	a,d
b , a , … , a	a,b,d
a , a , … , b	b,c,d
a , b , … , a	c
a , a , … , b	b,c
b , a , … , a	b
a , a , … , a	a,b

CLASSIFICATION UNDER PARTIAL MULTI-LABEL

Each instance is assigned with
a candidate label sets,
which contains multiple relevant labels
and some noisy labels.

IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 44, NO. 7, JULY 2022

Partial Multi-Label Learning With Noisy Label Identification

Ming-Kun Xie and Sheng-Jun Huang

X_1	X_2	…	X_n	C
a	b	…	b	a,b,c
b	b	…	a	a,c
a	a	…	b	d
b	a	…	b	b,c
a	b	…	a	a,d
b	a	…	a	a,b,d
a	a	…	b	b,c,d
a	b	…	a	c
a	a	…	b	b,c
b	a	…	a	b
a	a	…	a	a,b

PROBABILISTIC LABELS

LABEL DISTRIBUTIONS

X_1 , X_2 , … , X_n	c_1	c_2	c_3
a , b , … , b	0.3	0.3	0.4
b , b , … , a	0.4	0.2	0.4
a , a , … , b	0	1	0
a , b , … , a	0.7	0.2	0.1
b , a , … , a	0.5	0.5	0
a , a , … , b	0.3	0.1	0.6
a , b , … , a	0.4	0.2	0.4
a , b , … , b	0.7	0.2	0.4
b , a , … , b	0.9	0.1	0
b , b , … , a	0.6	0.2	0.2

Learning from data with uncertain labels by boosting credal classifiers

Benjamin Quost
HeuDiaSyC laboratory
deptartment of Computer Science
Compiègne University of Technology
Compiègne, France
quostben@hds.utc.fr

Thierry Denœux
HeuDiaSyC laboratory
deptartment of Computer Science
Compiègne University of Technology
Compiègne, France
tdenoeux@hds.utc.fr

IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 28, NO. 7, JULY 2016

Label Distribution Learning

Xin Geng, *Member, IEEE*

FULL-CLASS SET CLASSIFICATION

$X_1, X_2, \dots, X_n$	C
a , b , ... , b	b
b , b , ... , a	c
a , a , ... , b	c
b , a , ... , b	b
a , b , ... , a	a
b , a , ... , a	a
a , a , ... , b	c
a , b , ... , a	c
a , a , ... , b	b
b , a , ... , a	b
a , a , ... , a	a

Training

$X_1, X_2, \dots, X_n$	C
b , b , ... , a	Y_1
a , a , ... , b	
b , a , ... , b	
a , b , ... , a	Y_2
b , a , ... , a	
a , b , ... , a	
a , a , ... , b	Y_3
a , b , ... , b	
b , a , ... , b	
a , a , ... , b	Y_4
b , b , ... , a	
a , a , ... , a	

Test

$Y_i = (y_{i1}, y_{i2}, y_{i3}) = \text{PermutationOf}\{a, b, c\}$

Permutations						
y_{i1}	a	a	b	b	c	c
y_{i2}	b	c	a	c	a	b
y_{i3}	c	b	c	a	b	a

Int. J. Mach. Learn. & Cyber. (2010) 1:53–61
DOI 10.1007/s13042-010-0002-z

ORIGINAL ARTICLE

Full-class set classification using the Hungarian algorithm

Ludmila I. Kuncheva

- "Supervision degree" in prediction time !!

RESTRICTED SET CLASSIFICATION

Pattern Recognition 63 (2017) 158–170

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

ELSEVIER

Restricted set classification: Who is there?

Ludmila I. Kuncheva[a,*], Juan J. Rodríguez[b], Aaron S. Jackson[c]

[a] Bangor University, Dean Street, Bangor Gwynedd LL57 1UT, United Kingdom
[b] University of Burgos, Escuela Politécnica Superior, Avda. de Cantabria s/n, 09006 Burgos, Spain
[c] The University of Nottingham, NG8 1BB, United Kingdom

Classes in the chess-pieces recognition problem, and the limit number for each class in a standard chess game.

Class #:	1	2	3	4	5	6	7	8	9	10	11	12	13
	King	Queen	Rook	Bishop	Knight	Pawn	King	Queen	Rook	Bishop	Knight	Pawn	Empty
Numbers allowed:	1	1	2	2	2	8	1	1	2	2	2	8	62

This paper extends the above model to the more general case where X consists of m instances, and it is known that at most k_i instances may belong to class ω_i, $i = 1, \ldots, c$. Denoting $k = k_1 + \cdots + k_c$, we require that $m \leq k$. The who-is-who task is a special case where k_i=1, $i = 1, \ldots, c$, and m=c.

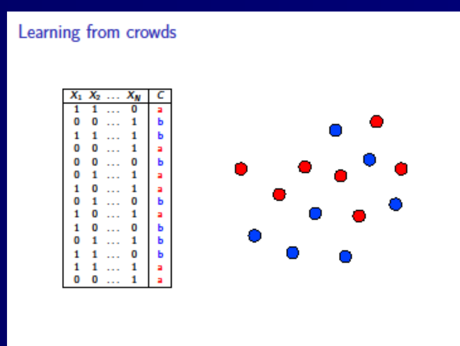
- When Predicting → Maximum number of samples per class is upper-bounded:
  - "Supervision degree" in prediction time !!
- Illustrative application → recognition of chess pieces

AND WHEN ANNOTATIONS ARE NOT FULLY RELIABLE?...

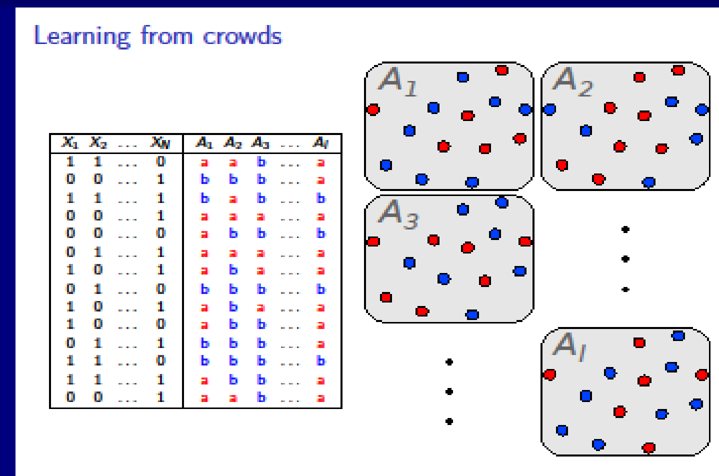
LEARNING FROM CROWDS

- real class for each object is not known: no "golden truth"

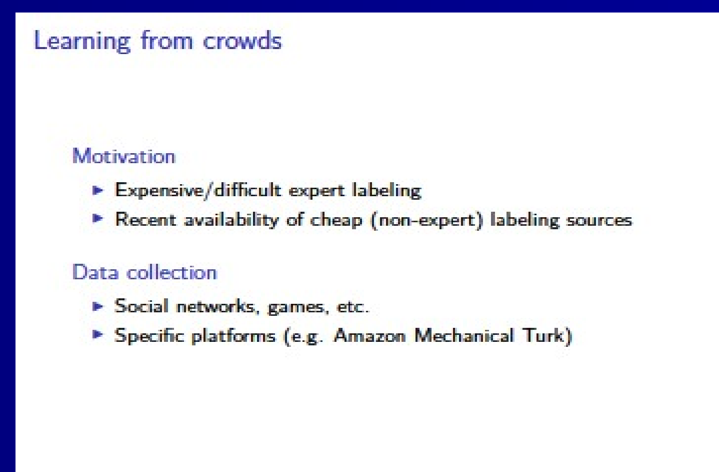
- humans (A_i) annotate their opinion about the label of each object → experts? novices?



X_1	X_2	....	X_N	C
1	1	....	0	a
0	0	....	1	b
1	1	....	1	b
0	0	....	1	a
0	0	....	0	b
0	1	....	1	a
1	0	....	1	a
0	1	....	0	b
1	0	....	1	a
1	0	....	0	b
0	1	....	1	b
1	1	....	0	b
1	1	....	1	a
0	0	....	1	a



X_1	X_2	....	X_N	A_1	A_2	A_3	....	A_I
1	1	....	0	a	a	b	....	a
0	0	....	1	b	b	b	....	a
1	1	....	1	b	a	b	....	b
0	0	....	1	a	a	a	....	a
0	0	....	0	a	b	b	....	b
0	1	....	1	a	a	a	....	a
1	0	....	1	a	b	a	....	a
0	1	....	0	b	b	b	....	b
1	0	....	1	a	b	a	....	a
1	0	....	0	a	b	b	....	a
0	1	....	1	b	b	b	....	a
1	1	....	0	b	b	b	....	b
1	1	....	1	a	b	b	....	a
0	0	....	1	a	a	b	....	a

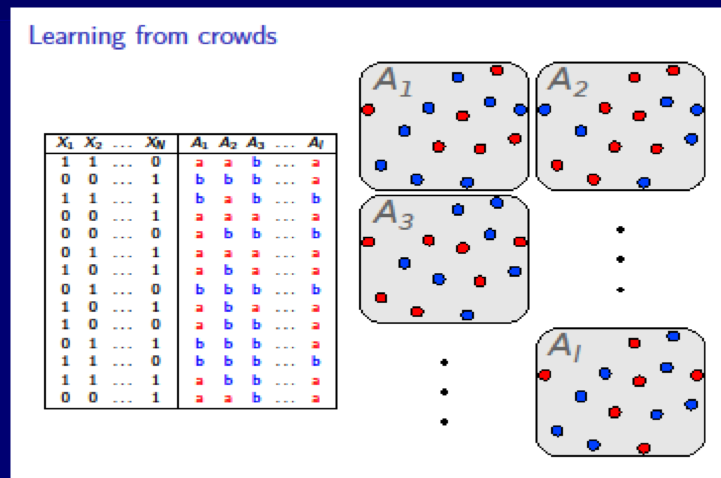


- Crowd annotation platforms

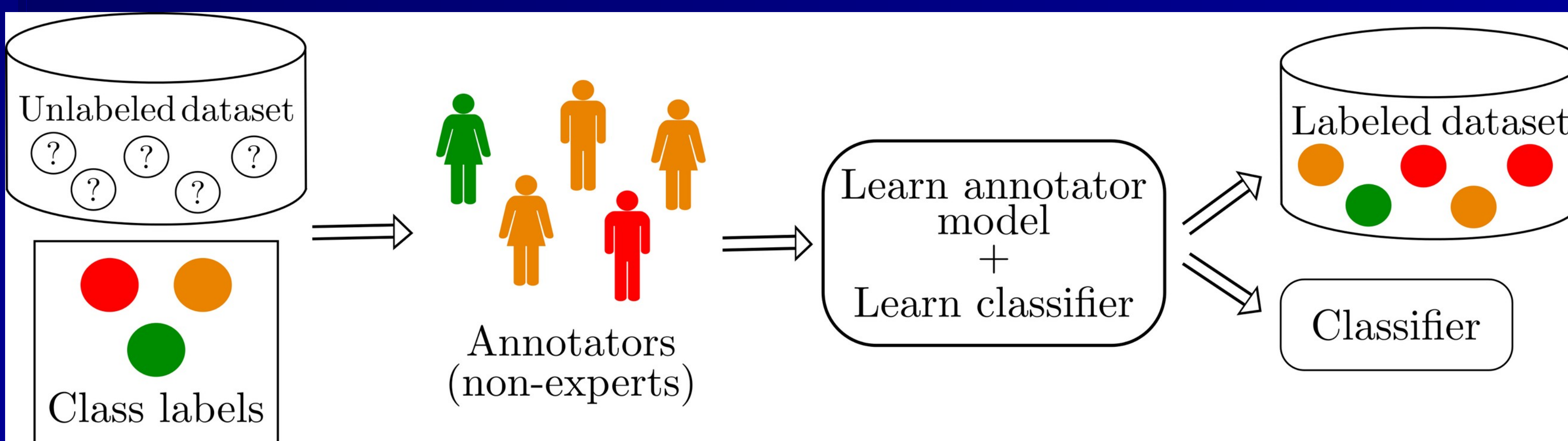
- TolokaAI, AI Crowd, Amazon Mechanial Turk...

AND WHEN ANNOTATIONS ARE NOT FULLY RELIABLE?...

LEARNING FROM CROWDS



X_1	X_2	...	X_N	A_1	A_2	A_3	...	A_l
1	1	...	0	a	a	b	...	a
0	0	...	1	b	b	b	...	a
1	1	...	1	b	a	b	...	b
0	0	...	1	a	a	a	...	a
0	0	...	0	a	b	b	...	b
0	1	...	1	a	a	a	...	a
1	0	...	1	a	b	a	...	a
0	1	...	0	b	b	b	...	b
1	0	...	1	a	b	a	...	a
1	0	...	0	a	b	b	...	a
0	1	...	1	b	b	b	...	a
1	1	...	0	b	b	b	...	b
1	1	...	1	a	b	b	...	a
0	0	...	1	a	a	b	...	a



AND WHEN ANNOTATIONS ARE NOT FULLY RELIABLE?…

LEARNING FROM CROWDS

Journal of Machine Learning Research 11 (2010) 1297-1322 Submitted 9/09; Revised 2/10; Published 4/10

Learning From Crowds

Vikas C. Raykar VIKAS.RAYKAR@SIEMENS.COM
Shipeng Yu SHIPENG.YU@SIEMENS.COM
CAD and Knowledge Solutions (IKM CKS)
Siemens Healthcare
Malvern, PA 19355 USA

Linda H. Zhao LZHAO@WHARTON.UPENN.EDU
Department of Statistics
University of Pennsylvania
Philadelphia, PA 19104 USA

Gerardo Hermosillo Valadez GERARDO.HERMOSILLOVALADEZ@SIEMENS.COM
Charles Florin CHARLES.FLORIN@SIEMENS.COM
Luca Bogoni LUCA.BOGONI@SIEMENS.COM
CAD and Knowledge Solutions (IKM CKS)
Siemens Healthcare
Malvern, PA 19355 USA

Linda Moy LINDA.MOY@NYUMC.ORG
Department of Radiology
New York University School of Medicine
New York, NY 10016 USA

SUPERVISION MODELS

Table 2
Collection of supervision models.

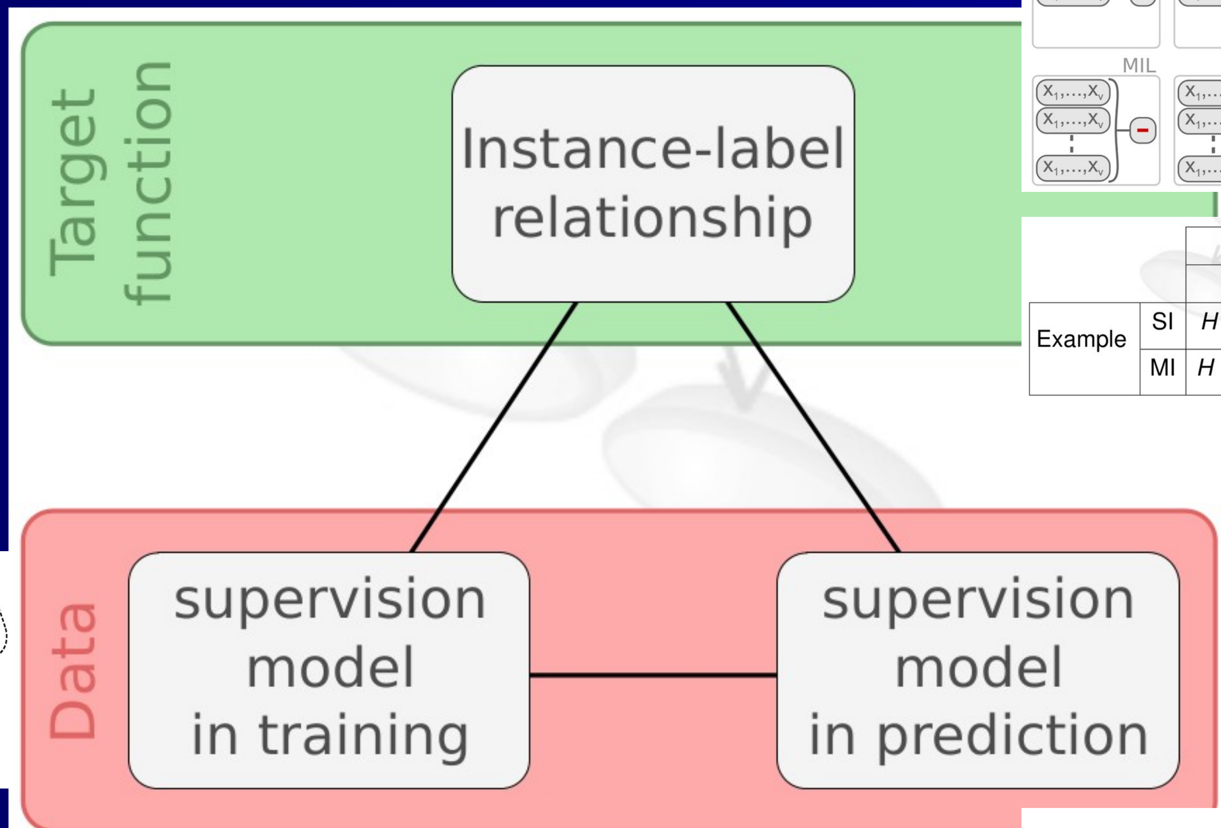
Model	References	Description
Full-supervision	[9,24,34,43]	For each example, complete class information is provided.
Unsupervision	[24]	No class information is provided with the examples.
Semi-supervision	[5]	Part of the examples are provided fully supervised. The rest are unsupervised.
Positive-unlabeled	[4,10,21,32]	Part of the examples are provided fully supervised, all of them with the same categorization. The rest are unsupervised.
Candidate labels	[7,13,16]	For each example, a set of class labels is provided. In this set, the class label(s) that compose the real categorization of the example are included.
Probabilistic labels	[18]	For each example, the probability of belonging to each class label is provided. This probability distribution is expected to assign high probability to the real label(s).
Incomplete	[3,33,42]	For each example, a subset of the labels that compose its real categorization is provided (SIML or MIML, Table 1).
Noisy labels	[2,44]	For each example, complete class information is provided, although its correctness is not guaranteed.
Crowd	[30,40]	For each example, many different non-expert annotators provide their (noisy) categorization.
Mutual label constraints	[19,20,31]	For each **group** of examples, an explicit relationship between their class labels is provided (e.g., all the examples have the same categorization).
Candidate labeling vectors	[22]	For each **group** of examples, a set of labeling vectors (including the real one) is provided. A labeling vector provides a class label for each examples of a group.
Label proportions	[15,25,28]	For each **group** of examples, the proportion of examples belonging to each class label is provided.

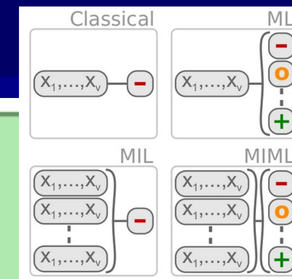
LEARNING SCENARIOS

Table 3
Brief description of classification problems and characterization according to the three axes of the taxonomy.

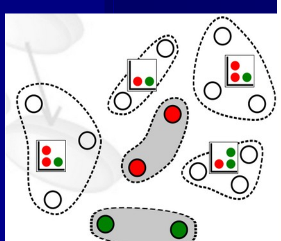
Problem	Description	Application (e.g.)	IL rel	SUPERVISION MODEL Learning	Prediction
Standard problem [24]	Learning with full categorized examples	Handwritten digit recogn.	SISL	Full-supervision	Unsupervision
Semi-supervised [5]	Learning with categorized and uncategorized examples	Text classification	SISL	Semi-supervision	Unsupervision
Positive-unlabeled [4]	Learning with examples of a category and other uncategorized examples	Spam detection, Gene prediction	SISL	Positive-unlabeled	Unsupervision
Mislabeled data [2] Ambiguous labels [44]	Learning with maybe wrong-categorized examples	Subjective labeler	SISL	Noisy Labels	Unsupervision
Partial labels [7]	Learning and prediction with uncategorized examples that have a set of possible categorizations	Classifying photographs with captions	SISL	Candidate labels	Unsupervision / Candidate labels
Multiple labels [18]	Learning with uncategorized examples that, with some probability, belong to a certain categorization	Bioinformatics	SISL	Probabilistic labels	Unsupervision
Partial equivalence relations [19]	Learning with groups of examples of the same/different categorization	Computer vision	SISL	Mutual label constraints	Unsupervision
Full-class set [20]	Prediction for a group of examples, all of them with a different categorization	Automatic attendance recording	SISL	Full-supervision	Mutual label constraints
Label proportions [15] Aggregate outputs [25]	Learning with groups of examples only knowing how many of them belong to each categorization	Embryo Selection, Polls prediction	SISL	Label proportions	Unsupervision
Candidate labeling sets [22]	Learning with groups of examples and sets of possible categorizing vectors	Classifying photographs with captions	SISL	Candidate labeling vectors	Unsupervision
Learning from crowds [30,40]	Learning with examples categorized with many candidate noisy categorizations	Image annotation	SISL	Crowd	Unsupervision
Multi-label [34]	Learning with examples that belong to several categorizations at the same time	Film genre prediction	SIML	Full-supervision	Unsupervision
Semi-supervised multi-label [6]	Learning with examples categorized with multiple labels or uncategorized	Text categorization	SIML	Semi-supervision	Unsupervision
ML with weak label [33] ML incomplete class [3]	Learning with examples categorized with a subset of the real multiple labels	Image annotation	SIML	Incomplete	Unsupervision
Set classification [26]	Prediction for a group of examples, all of them with the same categorization	Face recognition with multiple photos	SIML	Full-supervision	Mutual label constraints
MIL [9]	Learning with multiple-instance examples that are positive if at least one of their instances is	Molecule activation prediction	MISL	Full-supervision	Unsupervision
G-MIL [39]	Learning with examples represented by several instances with generalized function for positives	Key-and-lock prediction problem	MISL	Full-supervision	Unsupervision
MISSL [29]	Learning with categorized and uncategorized multiple-instances examples	Content-based image retrieval	MISL	Semi-supervision	Unsupervision
MIML [43]	Learning with examples represented with several instances that belong to several categorizations	Classifying texts, images or videos	MIML	Full-supervision	Unsupervision
SSMIML [41]	Learning with multiple-instance examples categorized with multiple labels or uncategorized	Video annotation	MIML	Semi-supervision	Unsupervision
MIML with weak labels [42]	Learning with multiple-instance examples categorized with a subset of the real multiple labels	Image annotation	MIML	Incomplete	Unsupervision

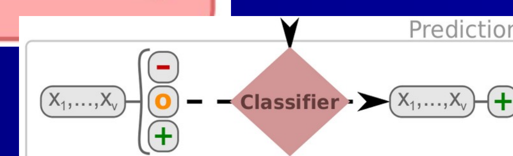
TAXONOMY OF WEAKLY SUPERVISED SCENARIOS





		Categorization	
		SL	ML
Example	SI	$H : \mathcal{X} \to \mathcal{C}$	$H : \mathcal{X} \to 2^{\mathcal{C}}$
	MI	$H : 2^{\mathcal{X}} \to \mathcal{C}$	$H : 2^{\mathcal{X}} \to 2^{\mathcal{C}}$





ALGORITHMS FOR WEAKLY SUPERVISED LEARNING

Document Classification Using Expectation Maximization with Semi Supervised Learning

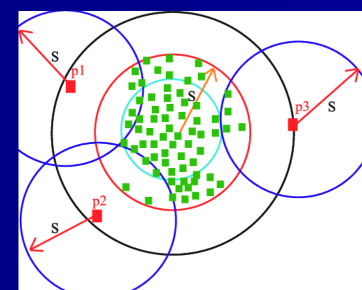
Bhawna Nigam[1], Poorvi Ahirwal[2], Sonal Salve[3], Swati Vamney[4]

1. Start from MLE $\theta = \{w, \mu, \Sigma\}_{1:2}$ on (X_l, Y_l),
   - w_c=proportion of class c
   - μ_c=sample mean of class c
   - Σ_c=sample cov of class c

   repeat:
2. The E-step: compute the expected label $p(y|x,\theta) = \frac{p(x,y|\theta)}{\sum_{y'} p(x,y'|\theta)}$ for all $x \in X_u$
   - label $p(y = 1|x,\theta)$-fraction of x with class 1
   - label $p(y = 2|x,\theta)$-fraction of x with class 2
3. The M-step: update MLE θ with (now labeled) X_u

LOF: Identifying Density-Based Local Outliers

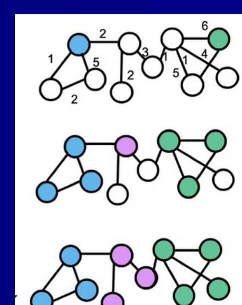
Markus M. Breunig†, Hans-Peter Kriegel†, Raymond T. Ng‡, Jörg Sander†



Learning from Labeled and Unlabeled Data with Label Propagation

Xiaojin Zhu*
* School of Computer Science
Carnegie-Mellon University
zhuxj@cs.cmu.edu

Zoubin Ghahramani*†
†Gatsby Computational Neuroscience Unit
University College London
zoubin@gatsby.ucl.ac.uk



LIBRARIES FOR WEAKLY SUPERVISED LEARNING

RSSL: Semi-supervised Learning in R

Jesse H. Krijthe[1,2]

[1] Pattern Recognition Laboratory, Delft University of Technology
[2] Department of Molecular Epidemiology, Leiden University Medical Center
jkrijthe@gmail.com

Expectation Maximization
Moment Constrained
Self-learning

`sklearn.neighbors`.LocalOutlierFactor

AdaSampling

An R implementation of the AdaSampling algorithm for positive unlabeled and label noise learning

pickLabel: Pick the optimal label from candidate labels
In Luwei-Ying/validateIt: Validating Topic Coherence and Topic Labels

`sklearn.svm`.OneClassSVM