Getting Lost in the Wealth of Classifier Ensembles?

Ludmila Kuncheva

School of Computer Science
Bangor University
mas00a@bangor.ac.uk

The Leverhulme Trust

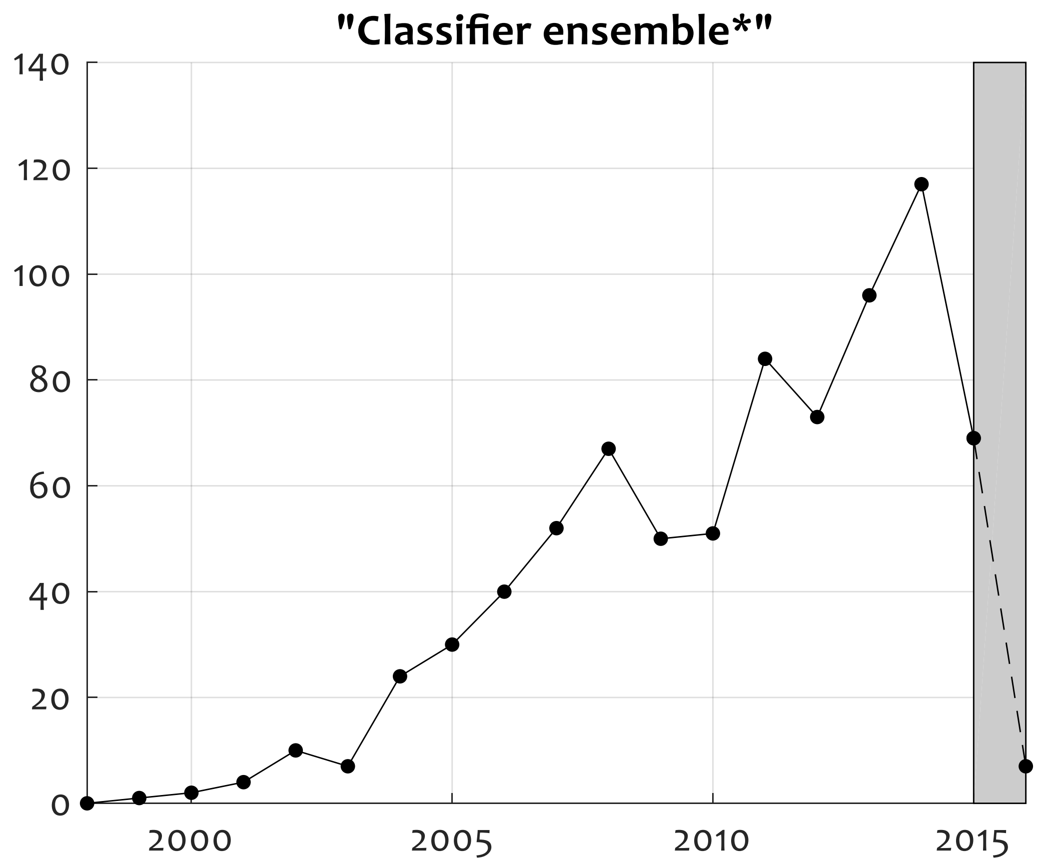
Supported by Project RPG-2015-188 Sponsored by the Leverhulme trust

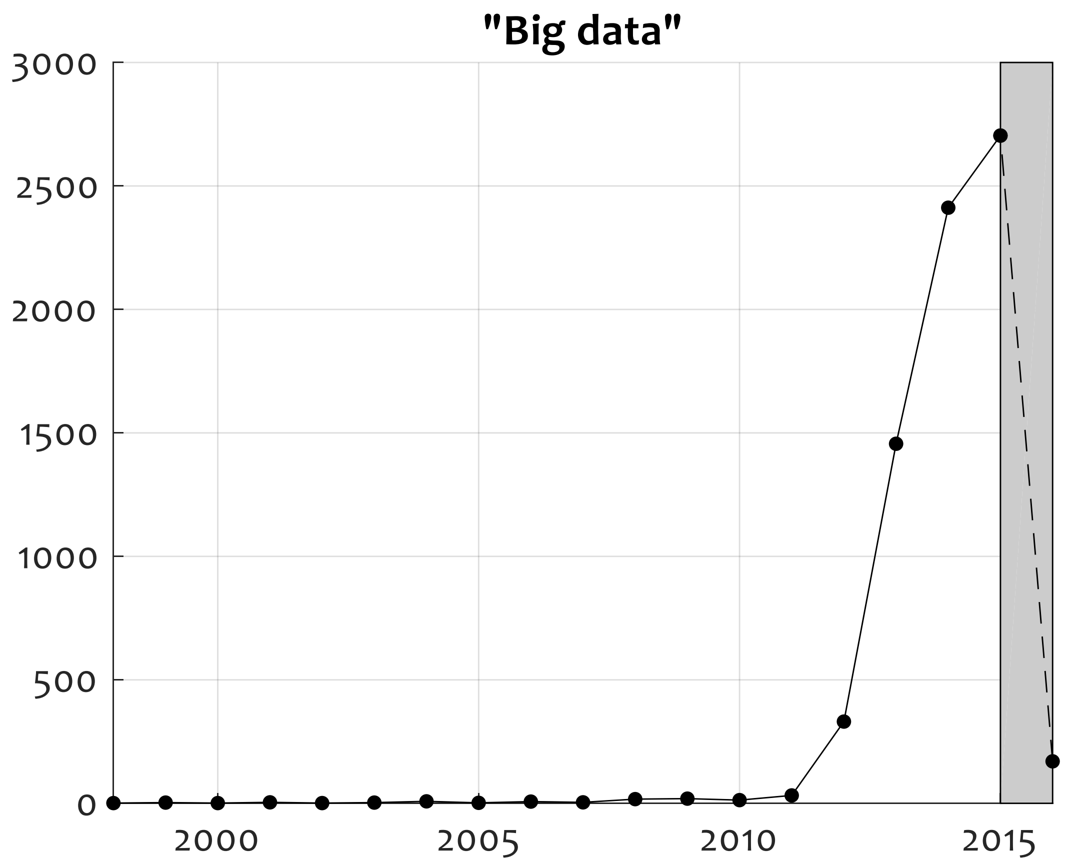
Doing research these days…

WEB OF SCIENCE™

Number of publications
08 February 2016

THOMSON REUTERS™





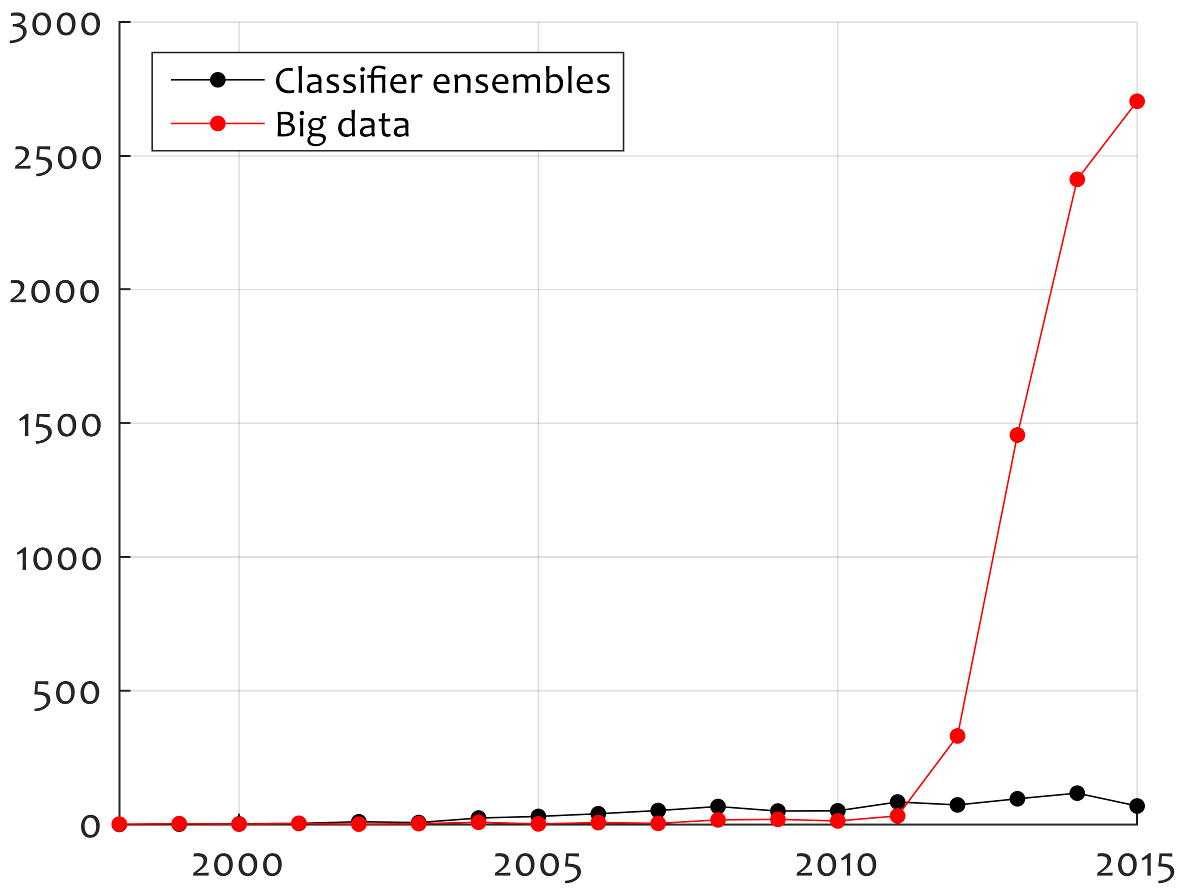
Web of Science TM | InCites TM | Journal Citation Reports® | Essential Science Indicators SM | EndNote TM | Sign In | Help | English

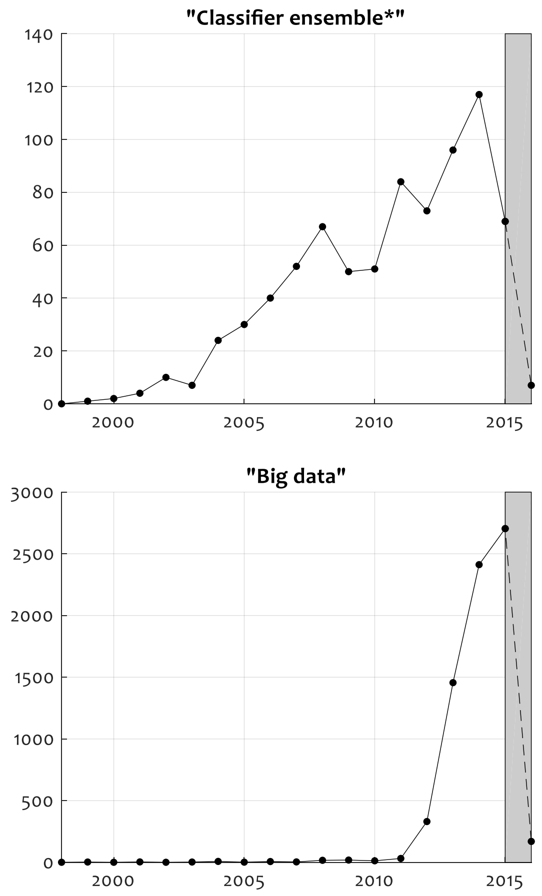
WEB OF SCIENCE™

Number of publications
08 February 2016

THOMSON REUTERS

Search | My Tools | Search History | Marked List





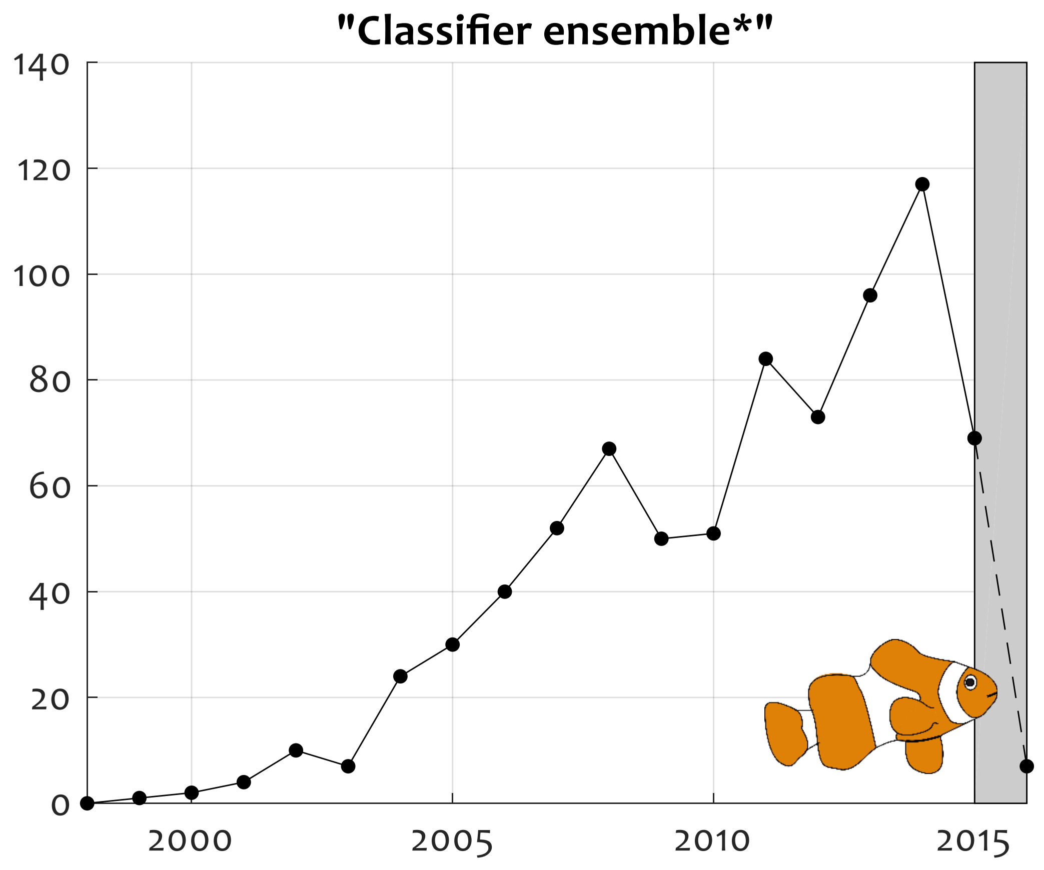
Web of Science ™ | InCites ™ | Journal Citation Reports ® | Essential Science Indicators SM | EndNote ™ | Sign In | Help | English

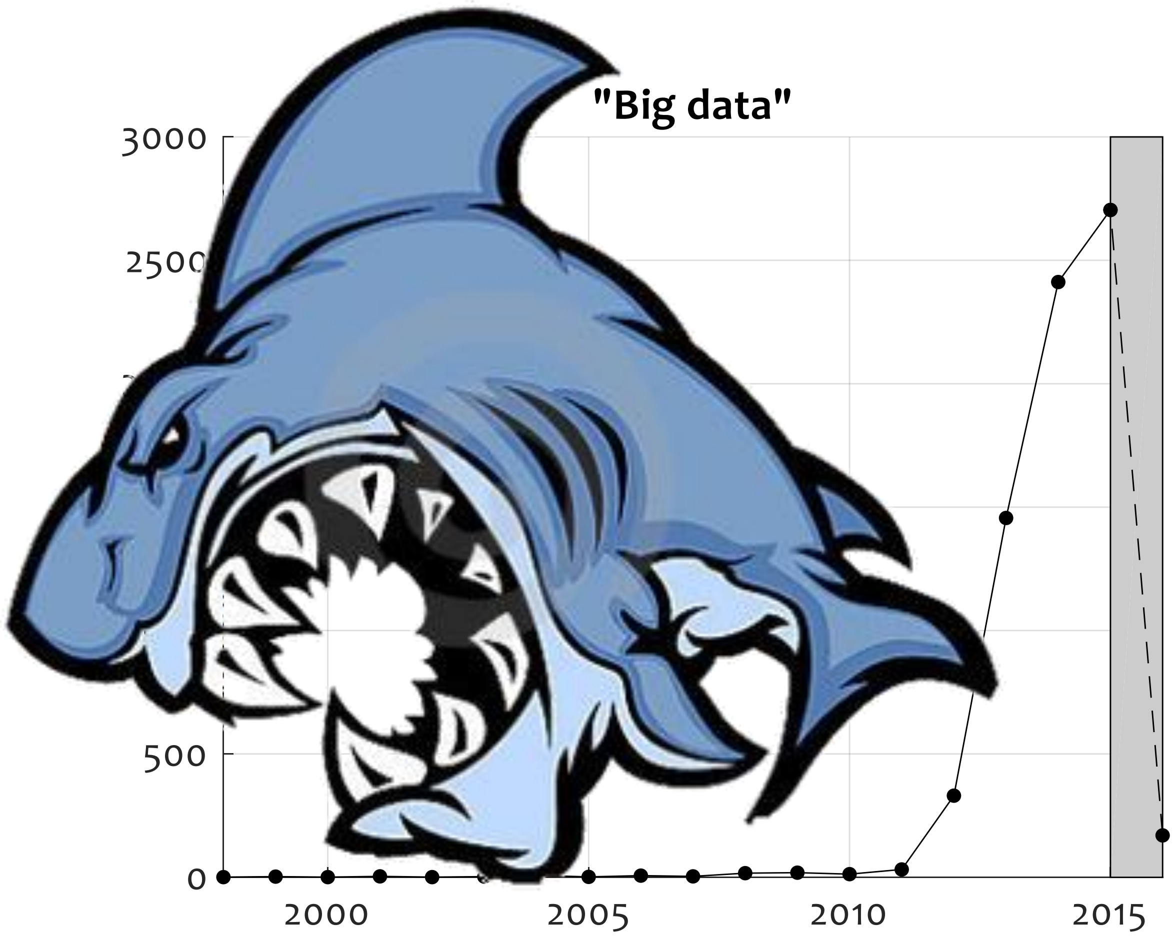
WEB OF SCIENCE™

THOMSON REUTERS

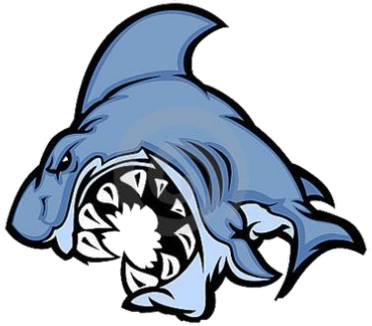
Number of publications
08 February 2016

Search | My Tools | Search History | Marked List





THE STACK NEWS | COMMUNITY | ABOUT US



CLOUD

Yahoo releases largest ever machine learning dataset to research

Martin Anderson Thu 14 Jan 2016 4.05pm

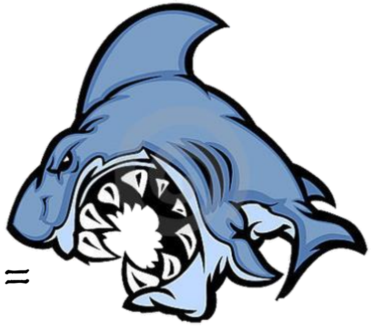
110 billion events

1.5 terabytes zipped

Ronald Fisher

The FAMOUS Iris Data

150 data points
4 features + labels



150*(4+1)*8 (bytes) =

6Kb

k-i-l-o-b-y-t-e-s , that is

WEB OF SCIENCE™

Search

09 February 2016

"iris data"

568 hits

8 inch floppy
250 Kb

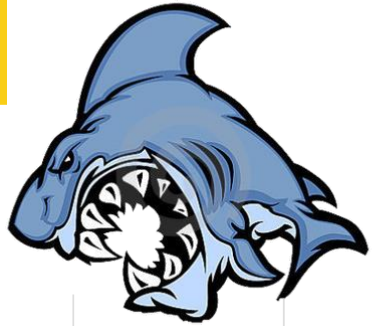
Dre-e-e-eam
dream dream dreaaam...


EC5274
SS/SD
СДЕЛАНО В БОЛГАРИИ
boedem
FLOPPY DISK
5,25" HD,DS
HD
WRITE ENABLE
WRITE PROTECT
INDEX

1.44 Mb
Wow!!!

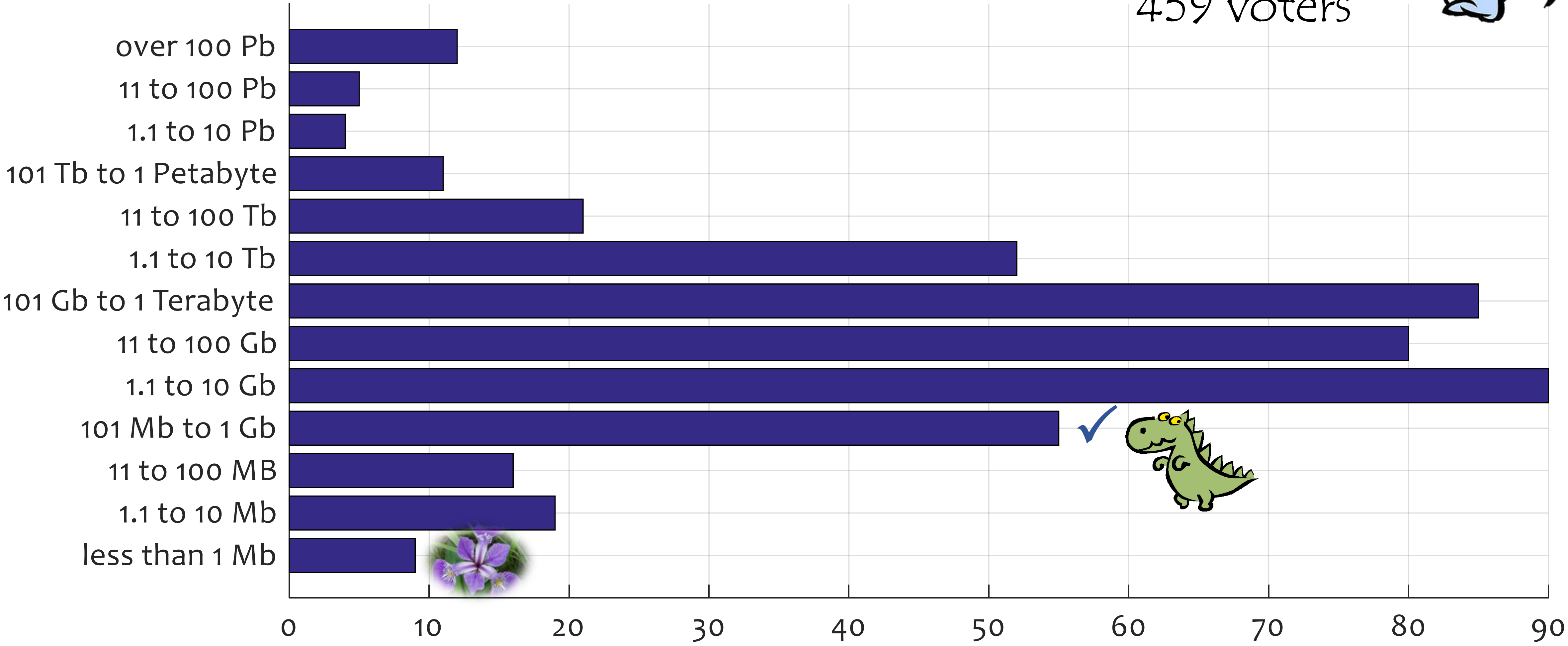
Let's say, generously, 1.5 Mb
Petabyte of storage is about 666,666,667 floppies

KDnuggets Home » Polls » largest dataset analyzed / data mined? Poll (Aug 2015)



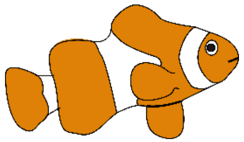
What was the largest dataset you analysed / data mined?

459 voters

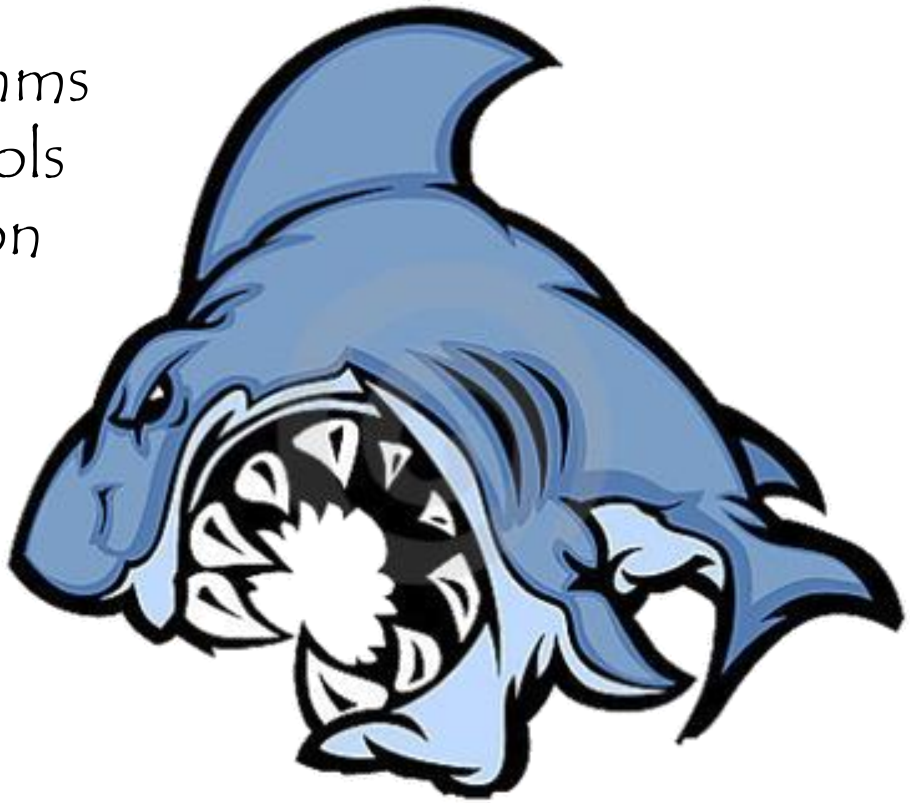


http://www.kdnuggets.com/polls/2015/largest-dataset-analyzed-data-mined.html

Pattern recognition algorithms
Training and testing protocols
Clever density approximation
Ingenious models



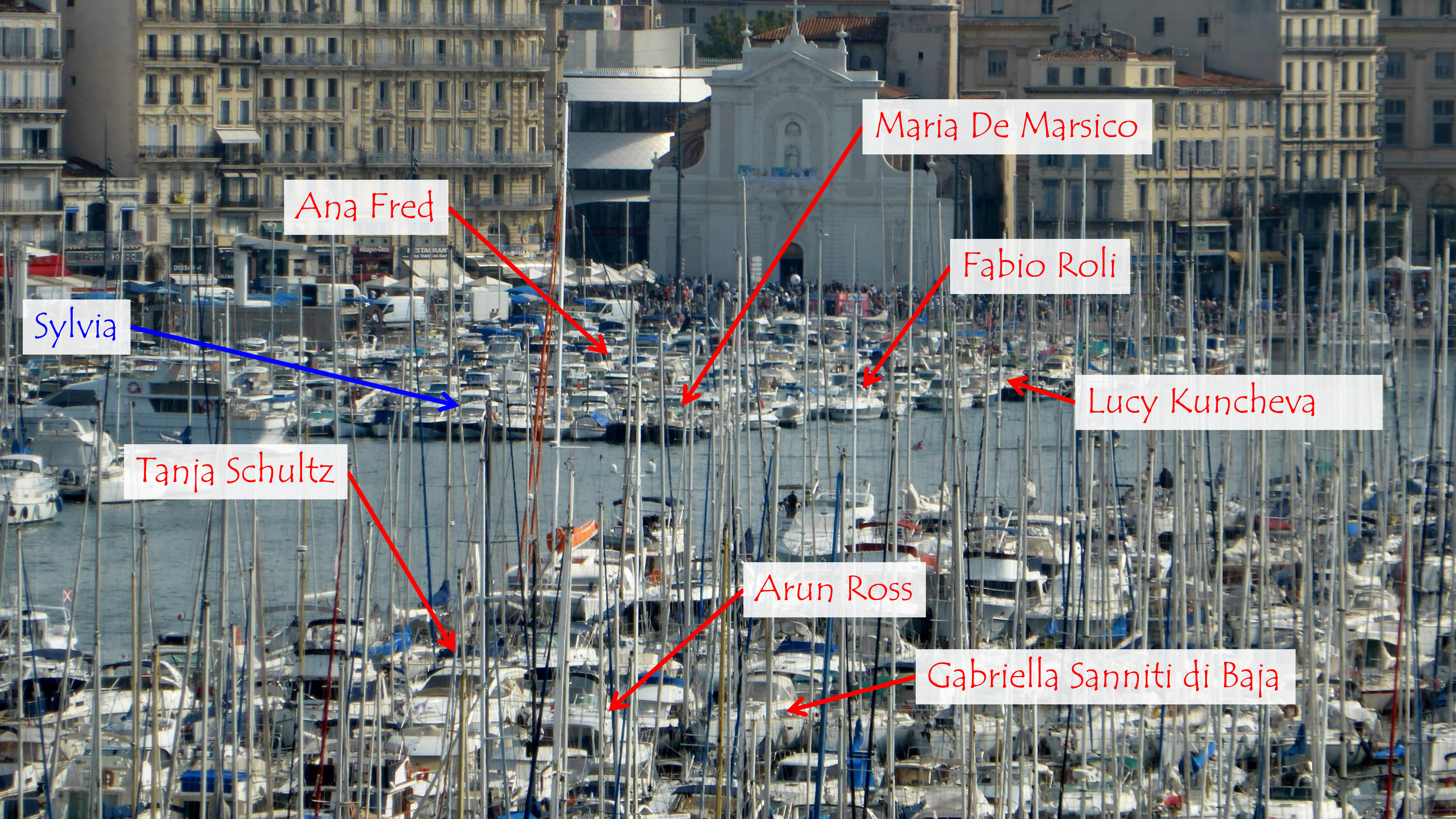
bye-bye...



HELLO

Data management and organisation
Efficient storage
Distributed computing
Fast computing
Optimisation / search
Computer clusters

A bit depressing...


Maria De Marsico
Ana Fred
Fabio Roli
Sylvia
Lucy Kuncheva
Tanja Schultz
Arun Ross
Gabriella Sanniti di Baja


Enjoying the ride but kind of... insignificant
YAHOO!
amazon
Google

All the Giants

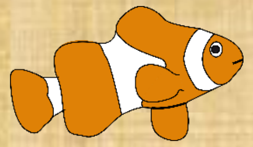
You...


By ESO - http://www.eso.org/public/images/eso1253a/, CC BY 4.0, https://commons.wikimedia.org/w/index.php?curid=23340651
Supercomputer

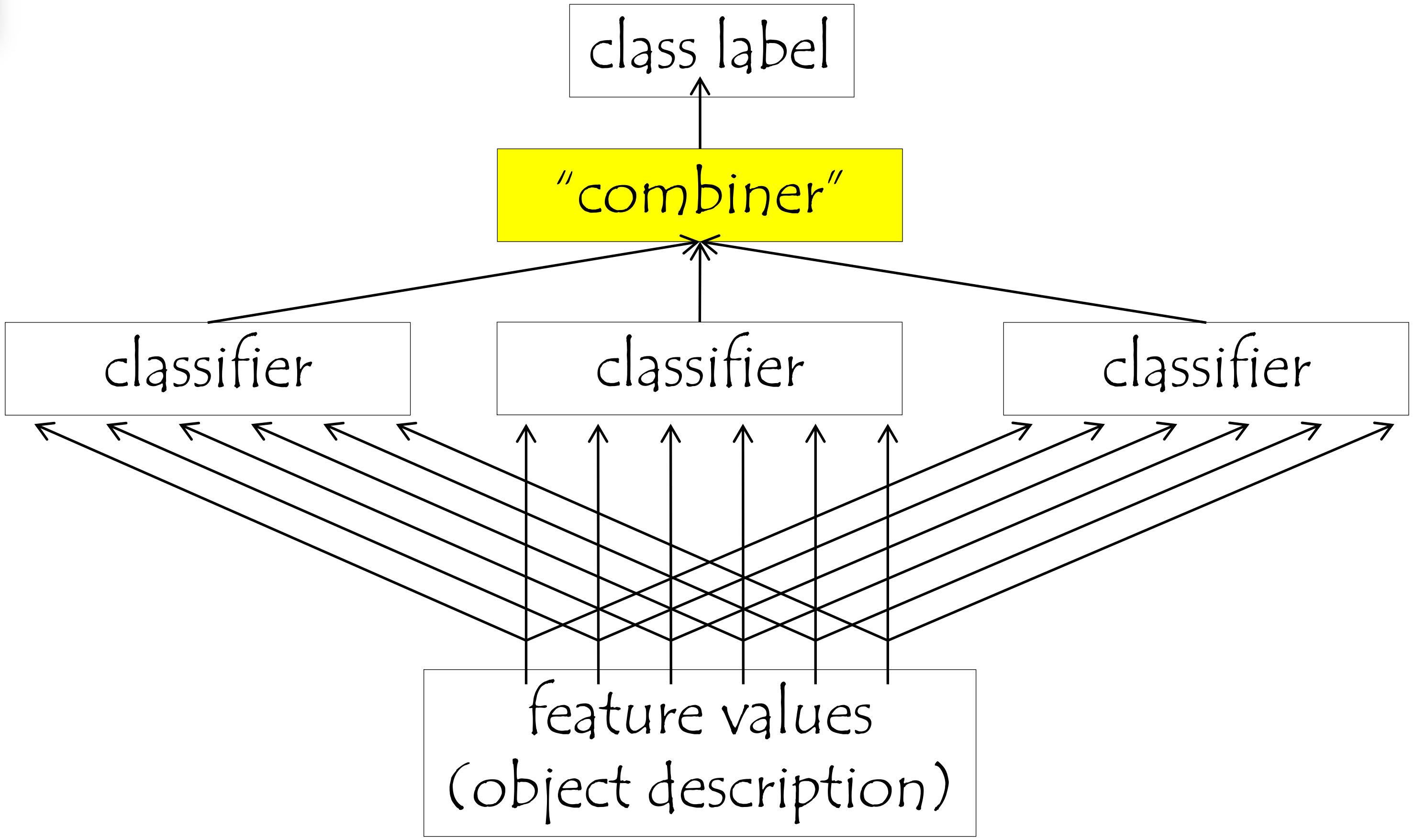
What chance do you have with your little...

You...

Classifier Ensembles?



Classifier ensembles



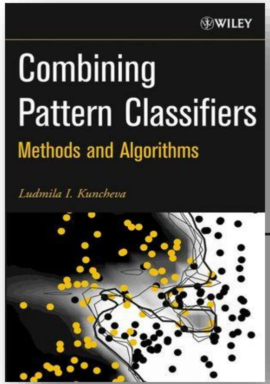
Classifier ensembles

VERY well liked classification approach. And this is why:

1. Typically better than the individual ensemble members and other individual classifiers.

2. And the above is enough ☺

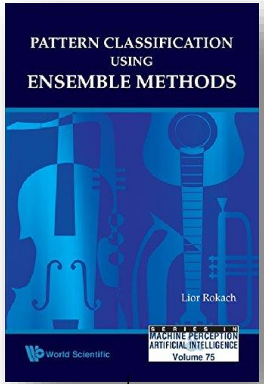
Classifier ensembles



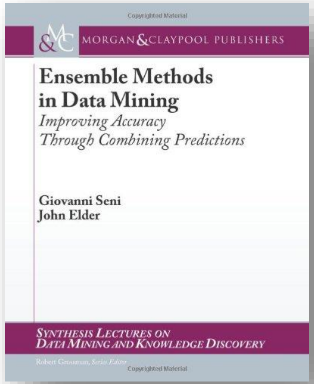
2004

Ludmila Kuncheva
Combining Pattern Classifiers.
Methods and Algorithms
Wiley, 2004

Lior Rokach
Pattern Classification Using Ensemble Methods
World Scientific
2010

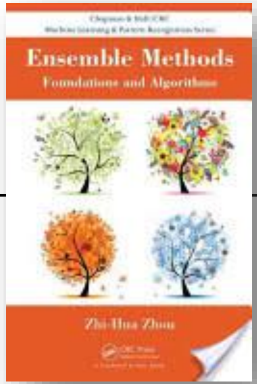


2010



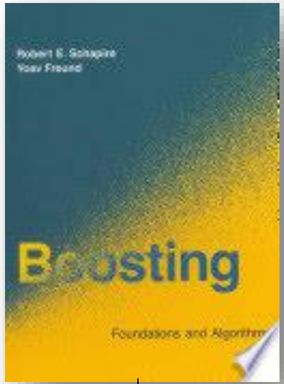
Giovanni Seni and John F. Elder
Ensemble Methods in Data Mining
Morgan & Claypool Publishers, 2010

Zhi-Hua Zhou
Ensemble Methods:
Foundations and Algorithms
Chapman & Hall/Crc , 2012

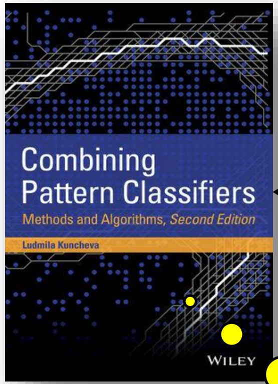


2012

Robert E. Schapire and Yoav Freund
Boosting:
Foundations and Algorithms
MIT Press, 2014



2014



Second edition

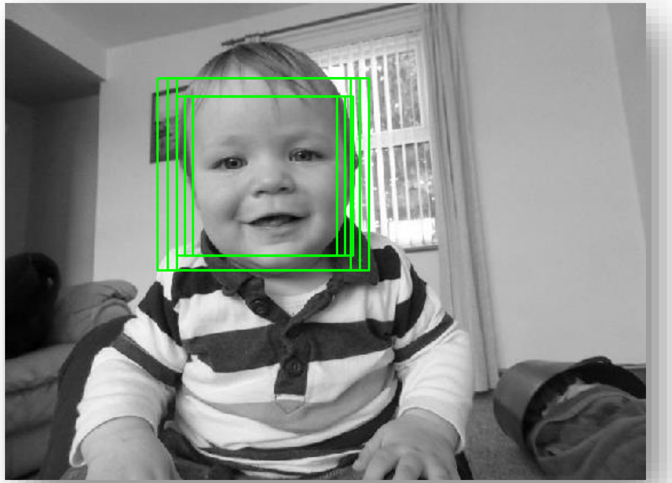
Classifier ensembles

2009 – Winner 1,000,000 UD$ – an ensemble method

US Patents

Satellite classifier ensemble
US 7769701 B2

Methods for feature selection using classifier ensemble based genetic algorithms
US 8762303 B



cited 12,403 times by 19/02/2016 (Google Scholar)

Rapid Object Detection using a Boosted Cascade of Simple Features

Paul Viola
viola@merl.com
Mitsubishi Electric Research Labs
201 Broadway, 8th FL
Cambridge, MA 02139

Michael Jones
mjones@crl.dec.com
Compaq CRL
One Cambridge Center
Cambridge, MA 02142

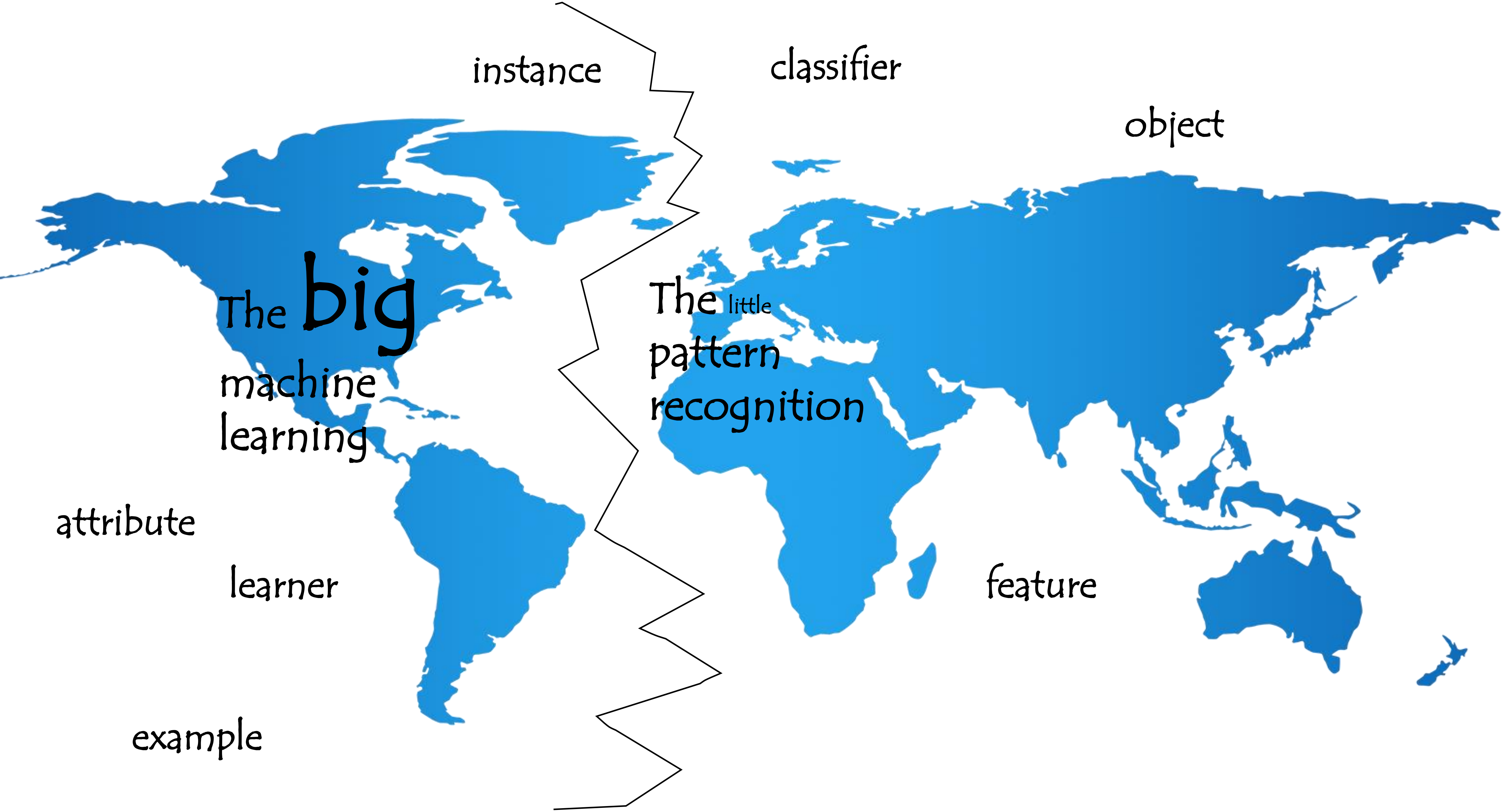
AdaBoost

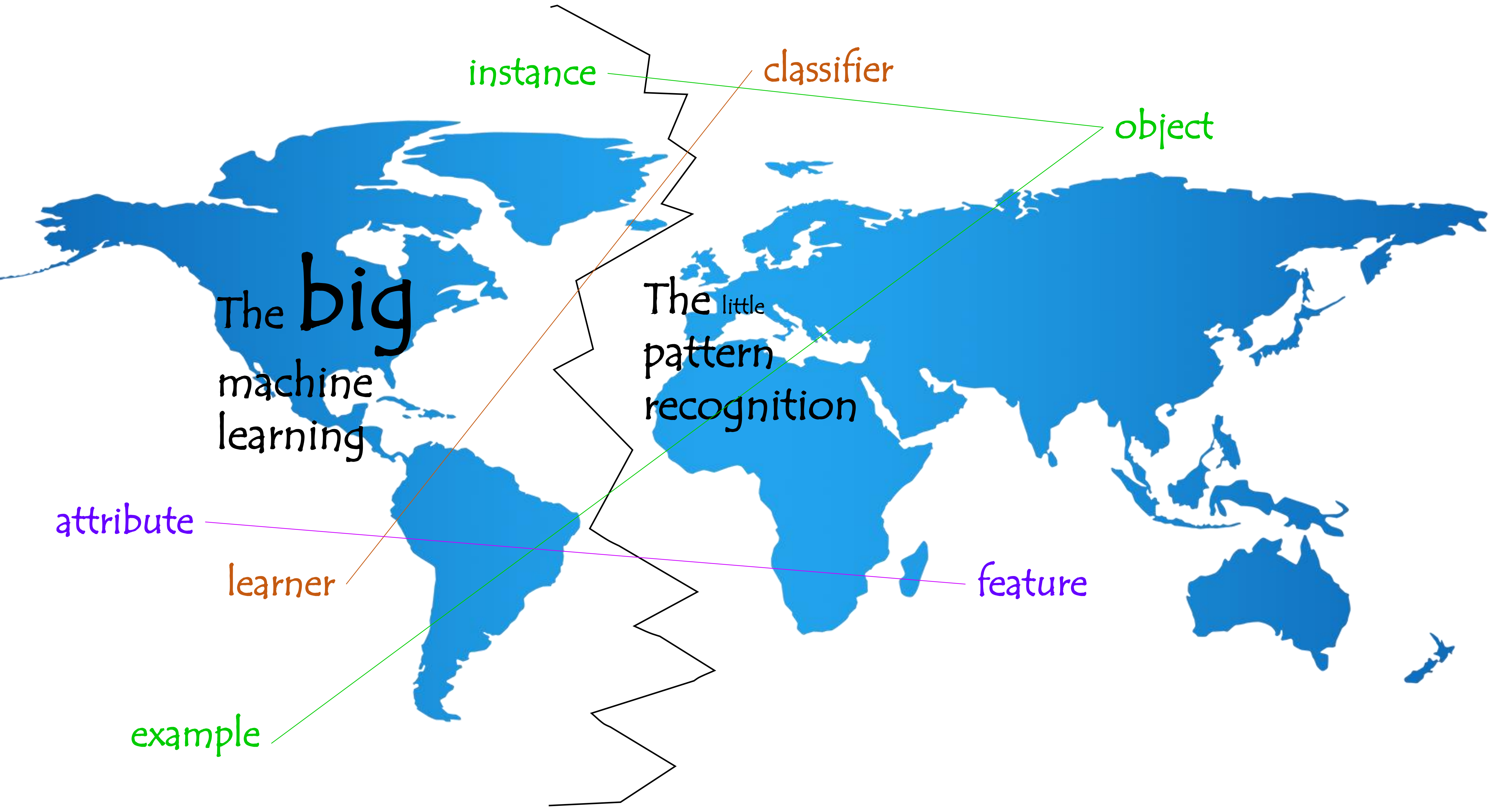
combination of multiple classifiers [Lam95,Woods97,Xu92,Kittler98]
classifier fusion [Cho95,Gader96,Grabisch92,Keller94,Bloch96]
mixture of experts [Jacobs91,Jacobs95,Jordan95,Nowlan91]
committees of neural networks [Bishop95,Drucker94]
consensus aggregation [Benediktsson92,Ng92,Benediktsson97]
voting pool of classifiers [Battiti94]
dynamic classifier selection [Woods97]
composite classifier systems [Dasarathy78] oldest
classifier ensembles [Drucker94,Filippi94,Sharkey99]
bagging, boosting, arcing, wagging [Sharkey99]
modular systems [Sharkey99]
collective recognition [Rastrigin81,Barabash83] oldest
stacked generalization [Wolpert92]
divide-and-conquer classifiers [Chiang94]
pandemonium system of reflective agents [Smieja96] fanciest
change-glasses approach to classifier selection [KunchevaPRL93]
etc.

combination of multiple classifiers [Lam95,Woods97,Xu92,Kittler98]
classifier fusion [Cho95,Gader96,Grabisch92,Keller94,Bloch96]
mixture of experts [Jacobs91,Jacobs95,Jordan95,Nowlan91]
committees of neural networks [Bishop95,Drucker94]
consensus aggregation [Benediktsson92,Ng92,Benediktsson97]
voting pool of classifiers [Battiti94]
dynamic classifier selection [Woods97]
composite classifier systems [Dasarathy78]
classifier ensembles [Drucker94,Filippi94,Sharkey99]
bagging, boosting, arcing, wagging [Sharkey99]
modular systems [Sharkey99]
collective recognition [Rastrigin81,Barabash83]
stacked generalization [Wolpert92]
divide-and-conquer classifiers [Chiang94]
pandemonium system of reflective agents [Smieja96]
change-glasses approach to classifier selection [Kunche
etc.

Out of fashion

Subsumed


instance
classifier
object
The big
machine
learning
The little
pattern
recognition
attribute
learner
feature
example


instance
classifier
object
The big machine learning
The little pattern recognition
attribute
learner
feature
example

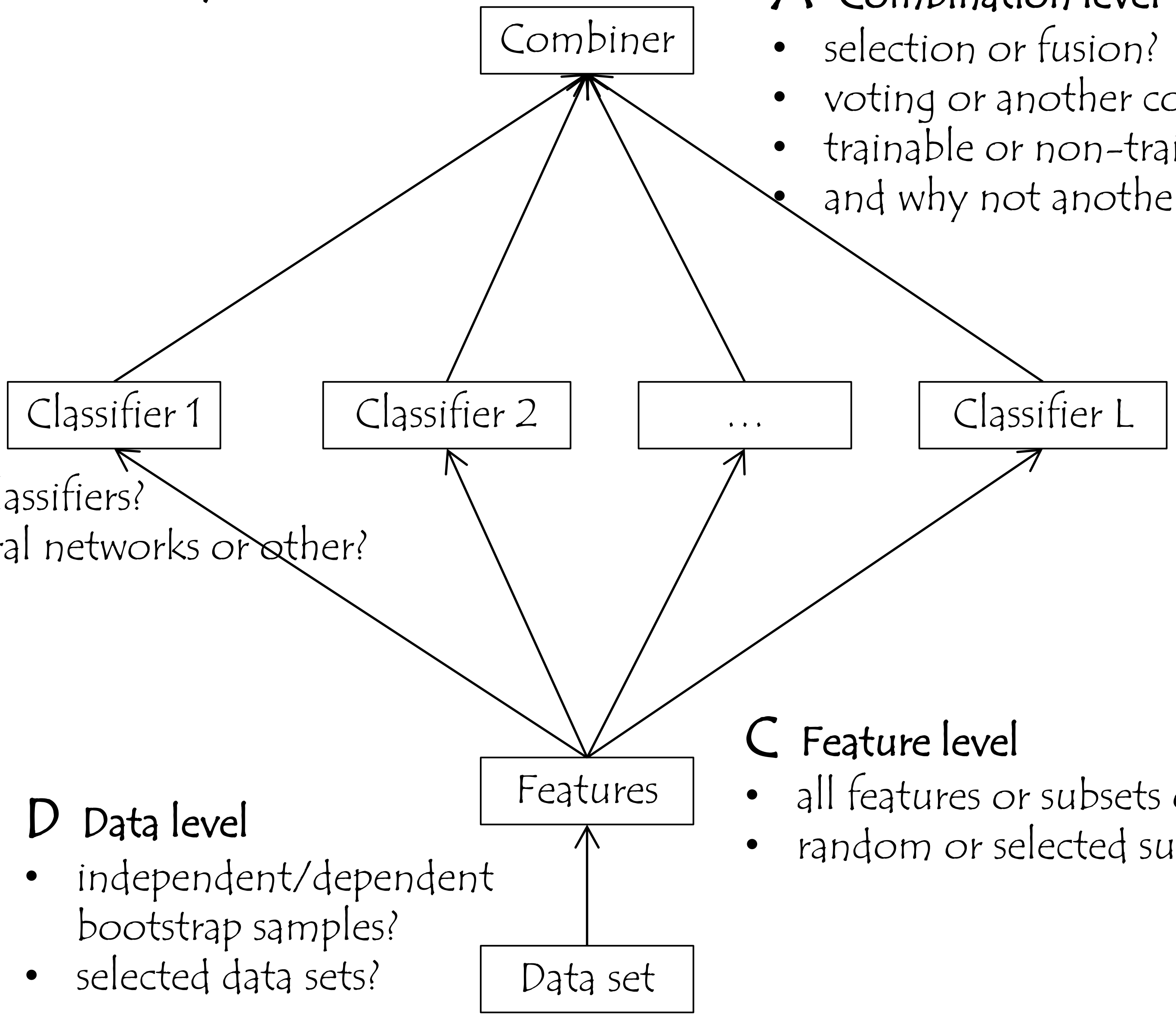

instance
classifier
object
The big machine learning
The little pattern recognition
attribute
learner
feature
example
And don't even start me on what they call things :)

International Workshops on Multiple Classifier Systems 2000 – 2014 - continuing

How do we design/build/train classifier ensembles?

Building ensembles

Levels of questions



A Combination level

- selection or fusion?
- voting or another combination method?
- trainable or non-trainable combiner?
- and why not another classifier?

B Classifier level

- same or different classifiers?
- decision trees, neural networks or other?
- how many?

C Feature level

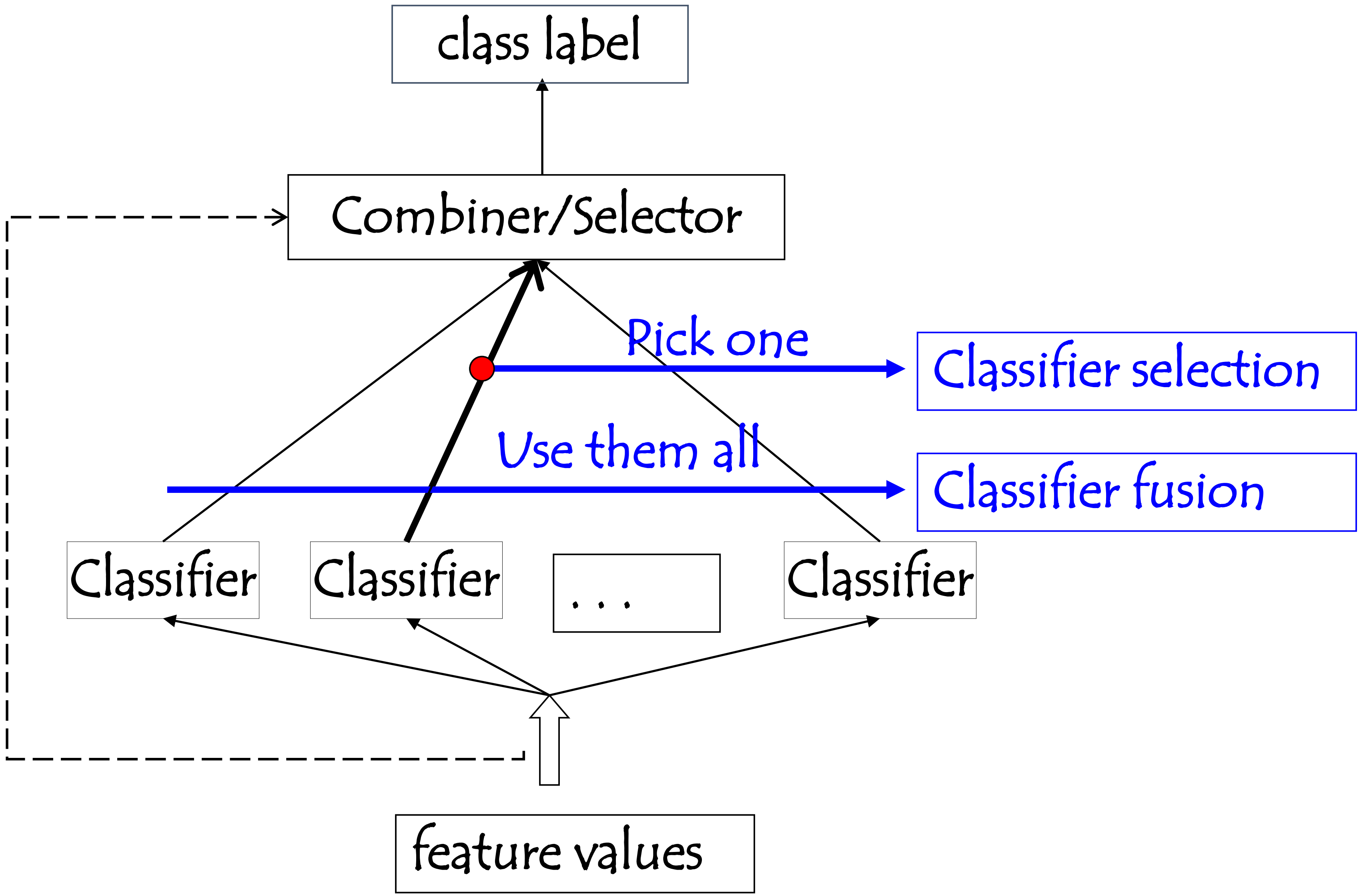
- all features or subsets of features?
- random or selected subsets?

D Data level

- independent/dependent bootstrap samples?
- selected data sets?

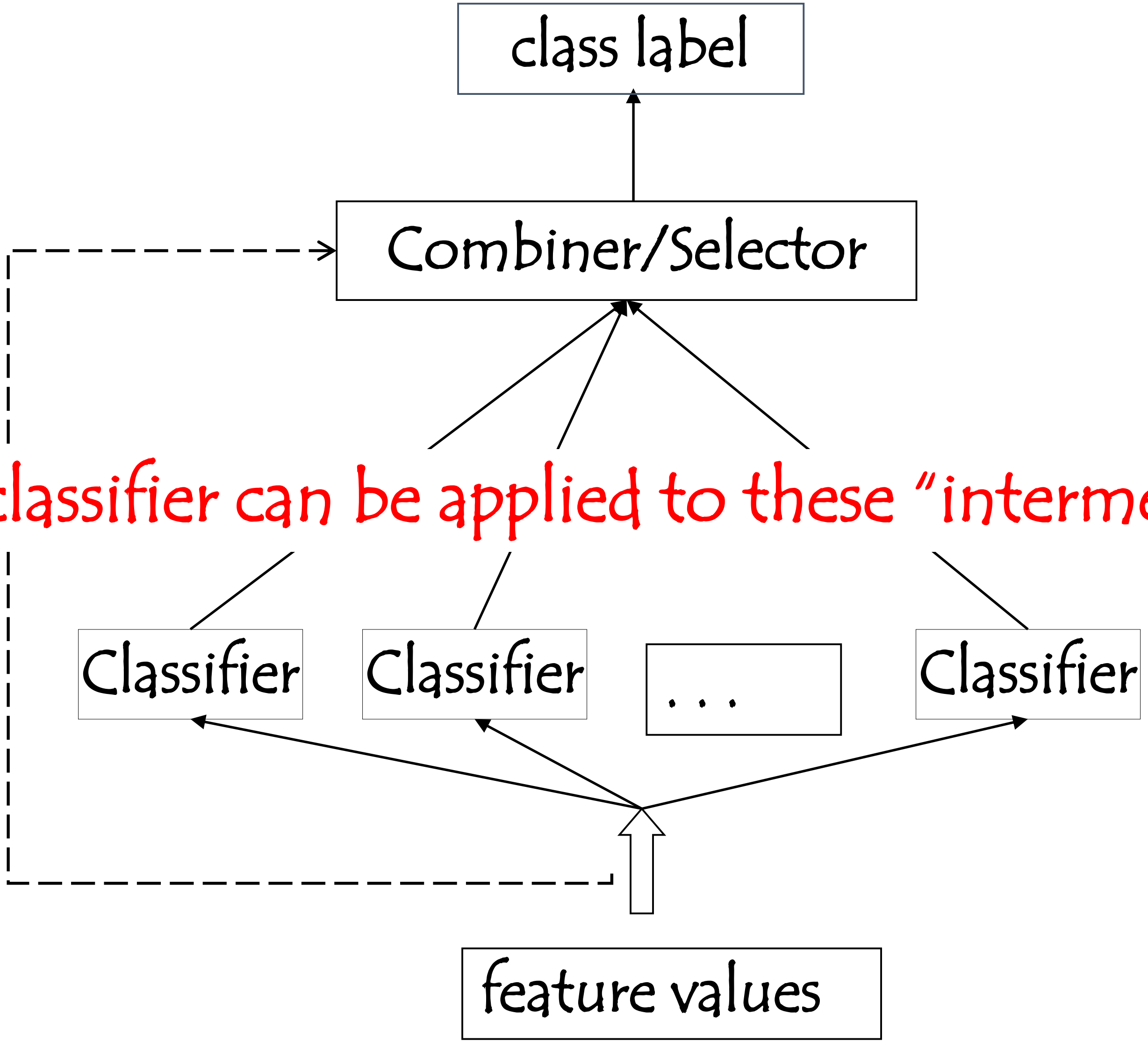
Building ensembles

The two strategies: fusion versus selection



Building ensembles

The two strategies: fusion versus selection



In fact, any classifier can be applied to these "intermediate features"

Classifier ensembles: the "CLASSICS"

- Bagging

  sample sample sample sample

  Independent bootstrap samples

- Boosting

  sample sample sample sample

  Dependent bootstrap samples

- Random Subspace

  sample sample sample sample

  Independent feature subsamples

- Random Forest

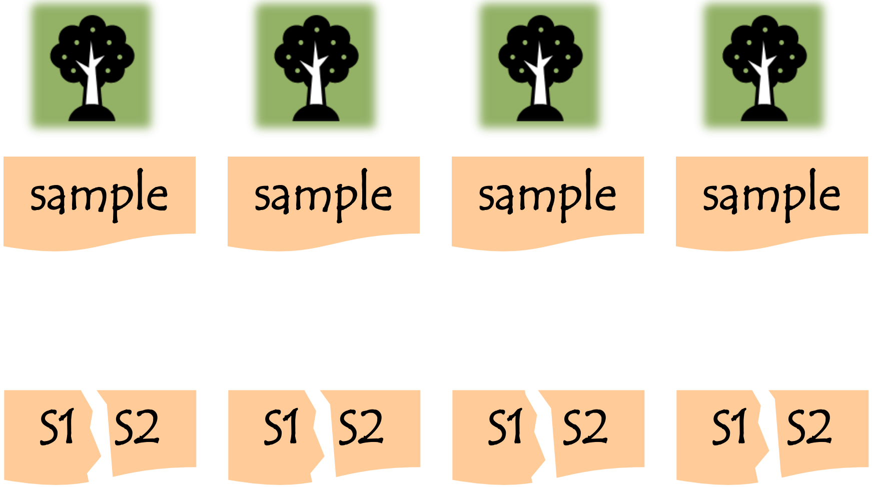
  Random trees

  sample sample sample sample

  Independent bootstrap samples

Classifier ensembles: the not-so-classics

- Rotation Forest
- Linear Oracle



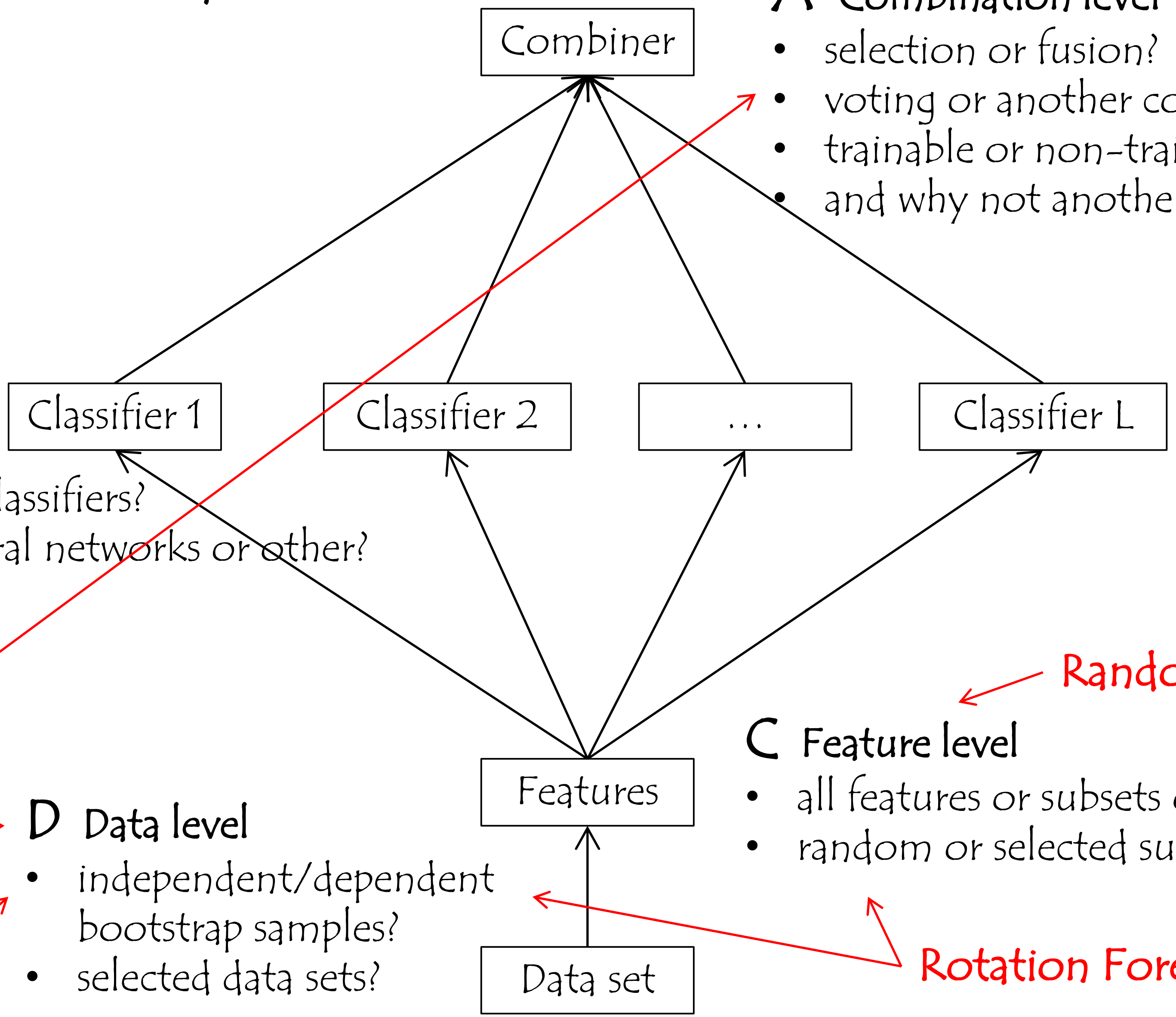
Standard or random trees

Independent bootstrap samples

Independently split samples

Building ensembles

Levels of questions



A Combination level

- selection or fusion?
- voting or another combination method?
- trainable or non-trainable combiner?
- and why not another classifier?

B Classifier level

- same or different classifiers?
- decision trees, neural networks or other?
- how many?

C Feature level

- all features or subsets of features?
- random or selected subsets?

D Data level

- independent/dependent bootstrap samples?
- selected data sets?

Random Forest

Boosting

Bagging

Linear Oracle

Random subspace

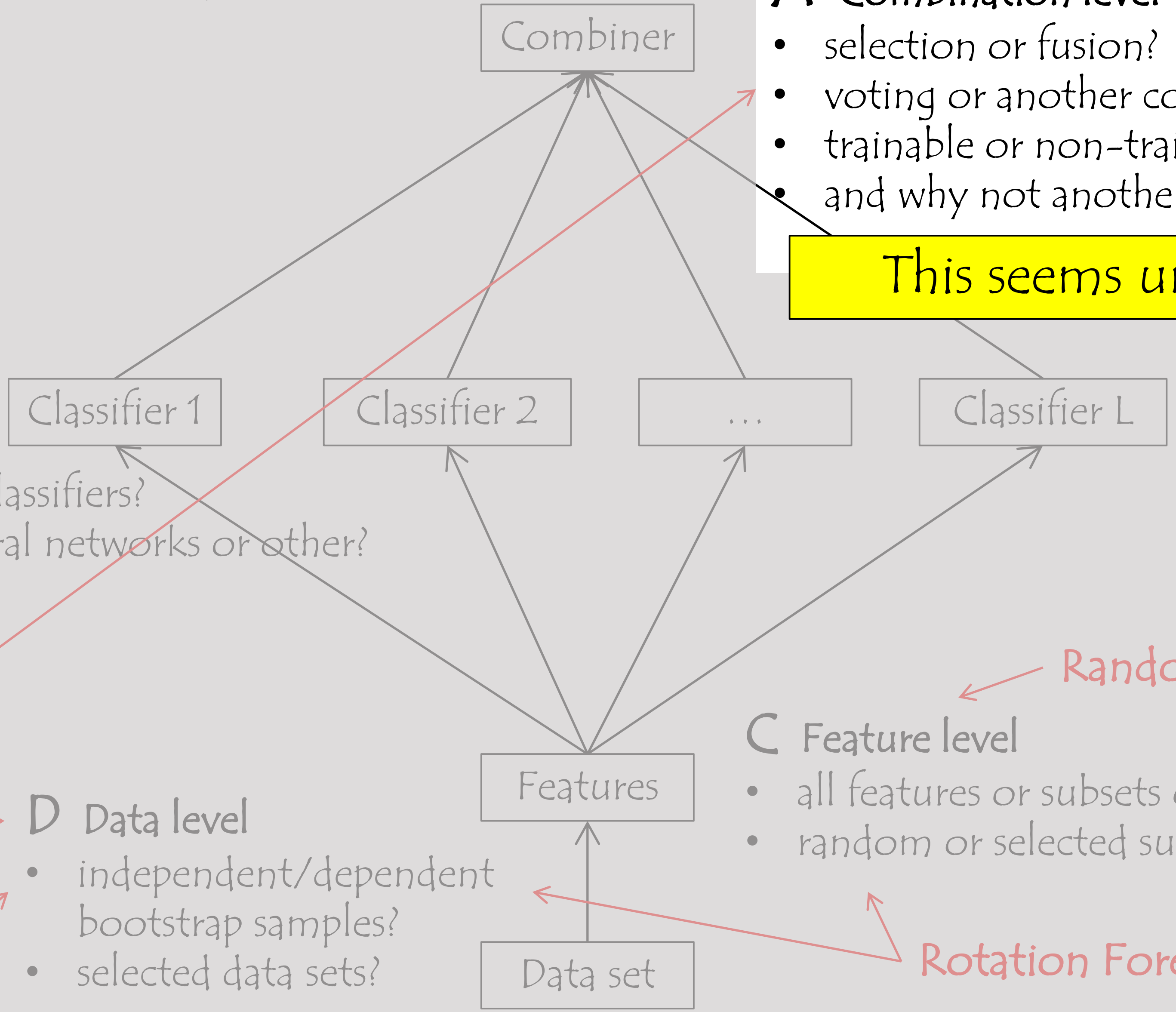
Rotation Forest

"Anchor" points

1. Combiner

Building ensembles

Levels of questions



A Combination level

- selection or fusion?
- voting or another combination method?
- trainable or non-trainable combiner?
- and why not another classifier?

This seems under-researched...

B Classifier level

- same or different classifiers?
- decision trees, neural networks or other?
- how many?

C Feature level

- all features or subsets of features?
- random or selected subsets?

D Data level

- independent/dependent bootstrap samples?
- selected data sets?

Random Forest

Boosting

Bagging

Linear Oracle

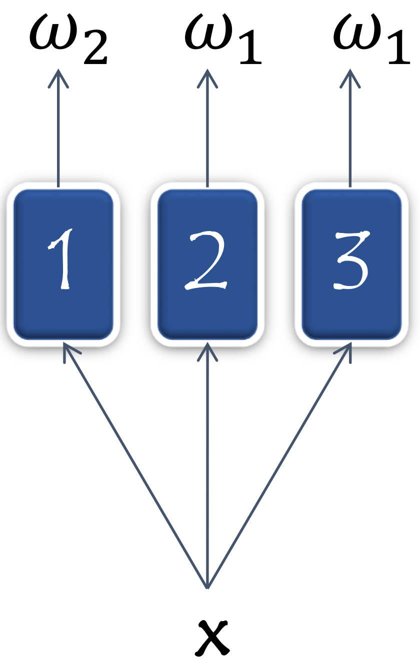
Random subspace

Rotation Forest

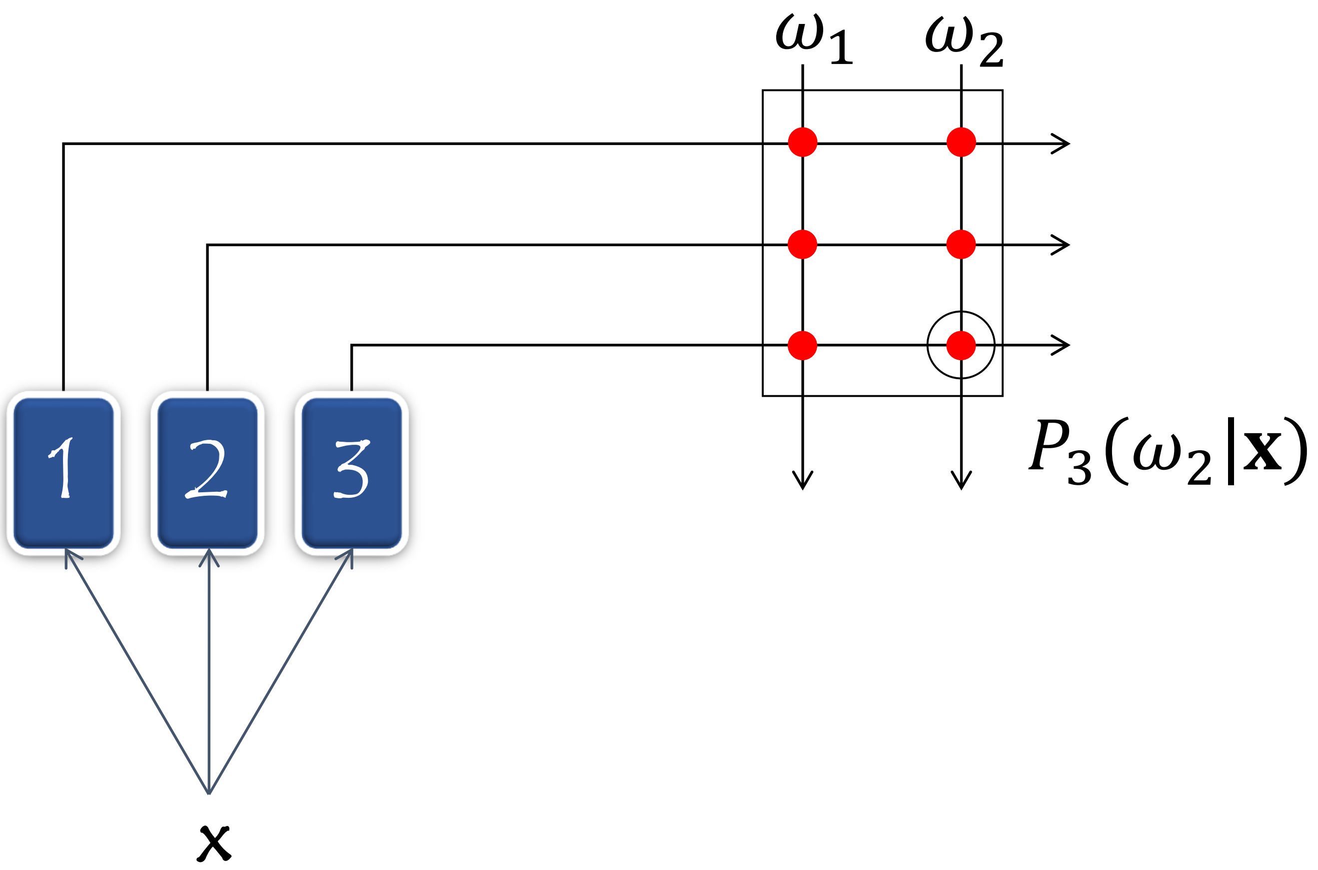
Combiner

Continuous-valued outputs

Label outputs

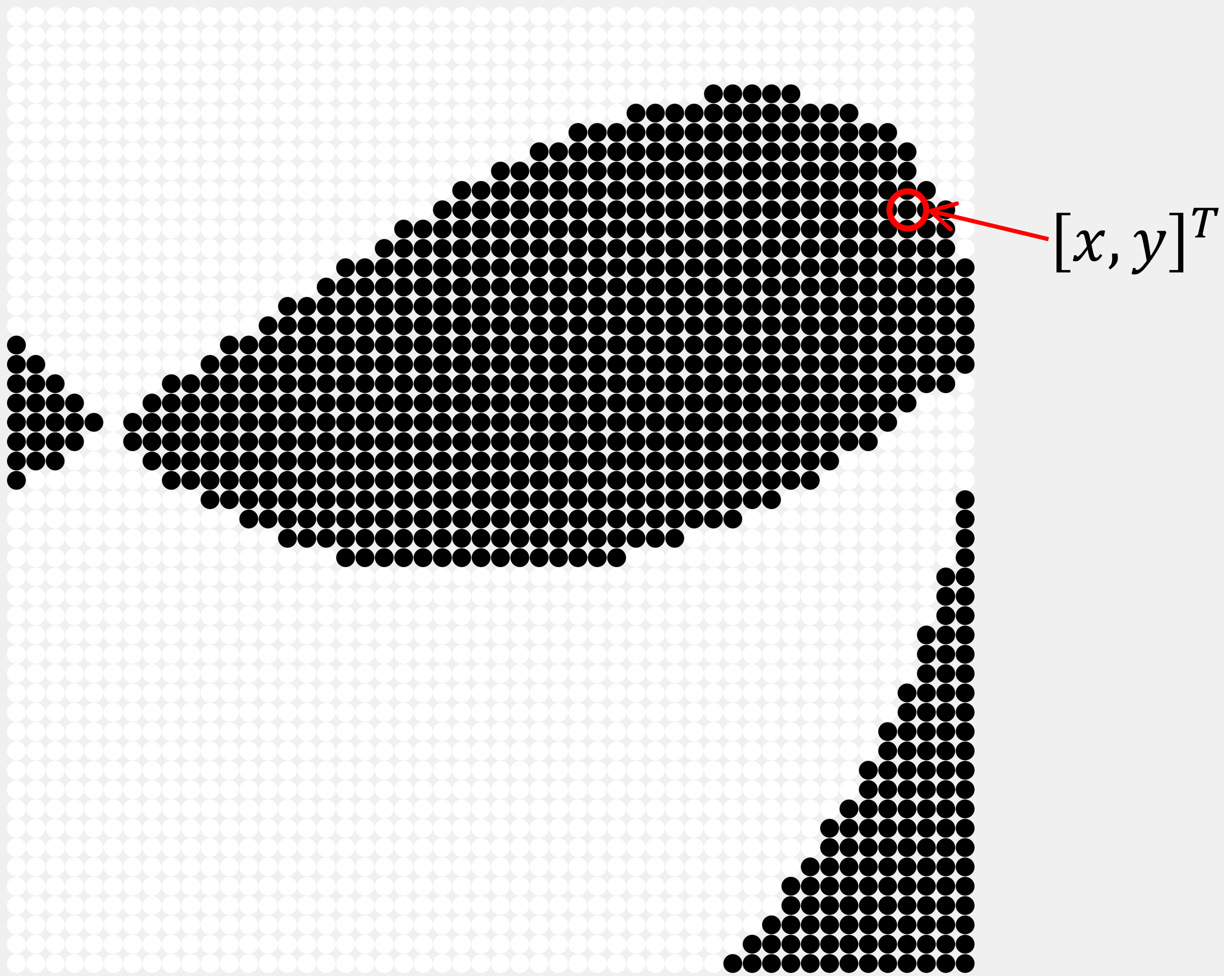


Decision profile



Data set:

Let's call this data "The Tropical Fish" or just the **fish data**.

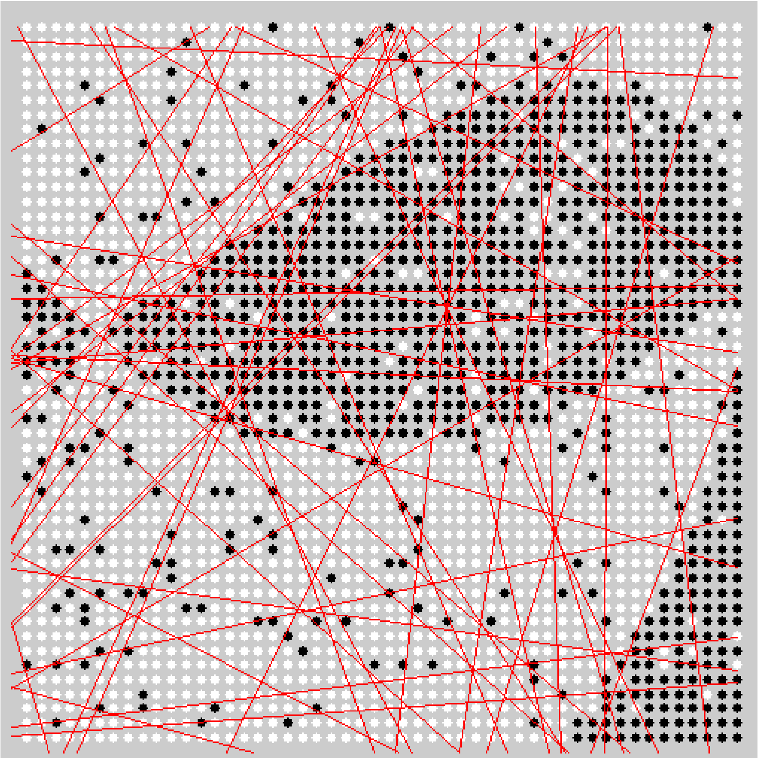


50-by-50 =
2500 objects in 2-d

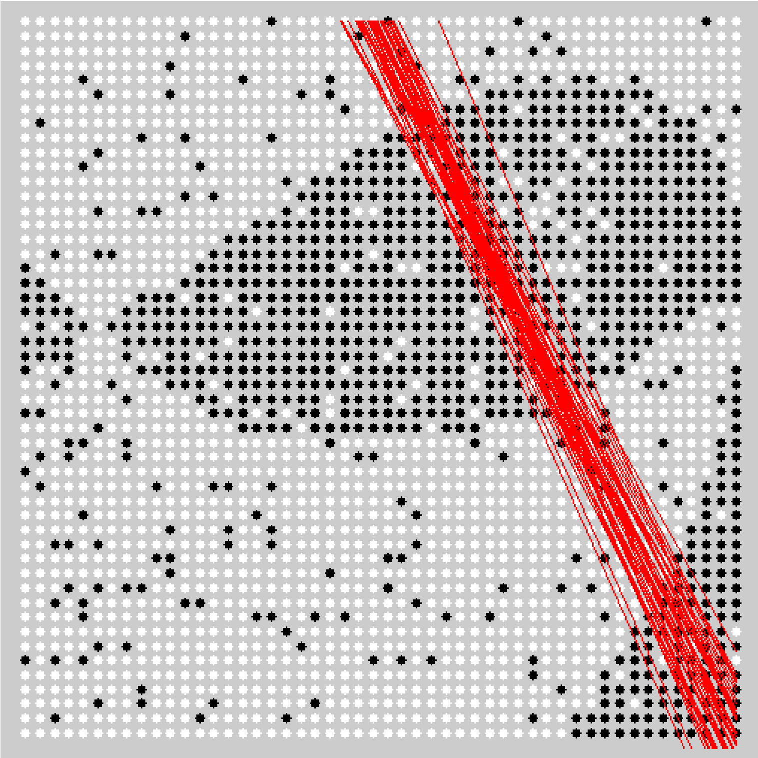
Bayes error rate = 0%

Example: 2 ensembles

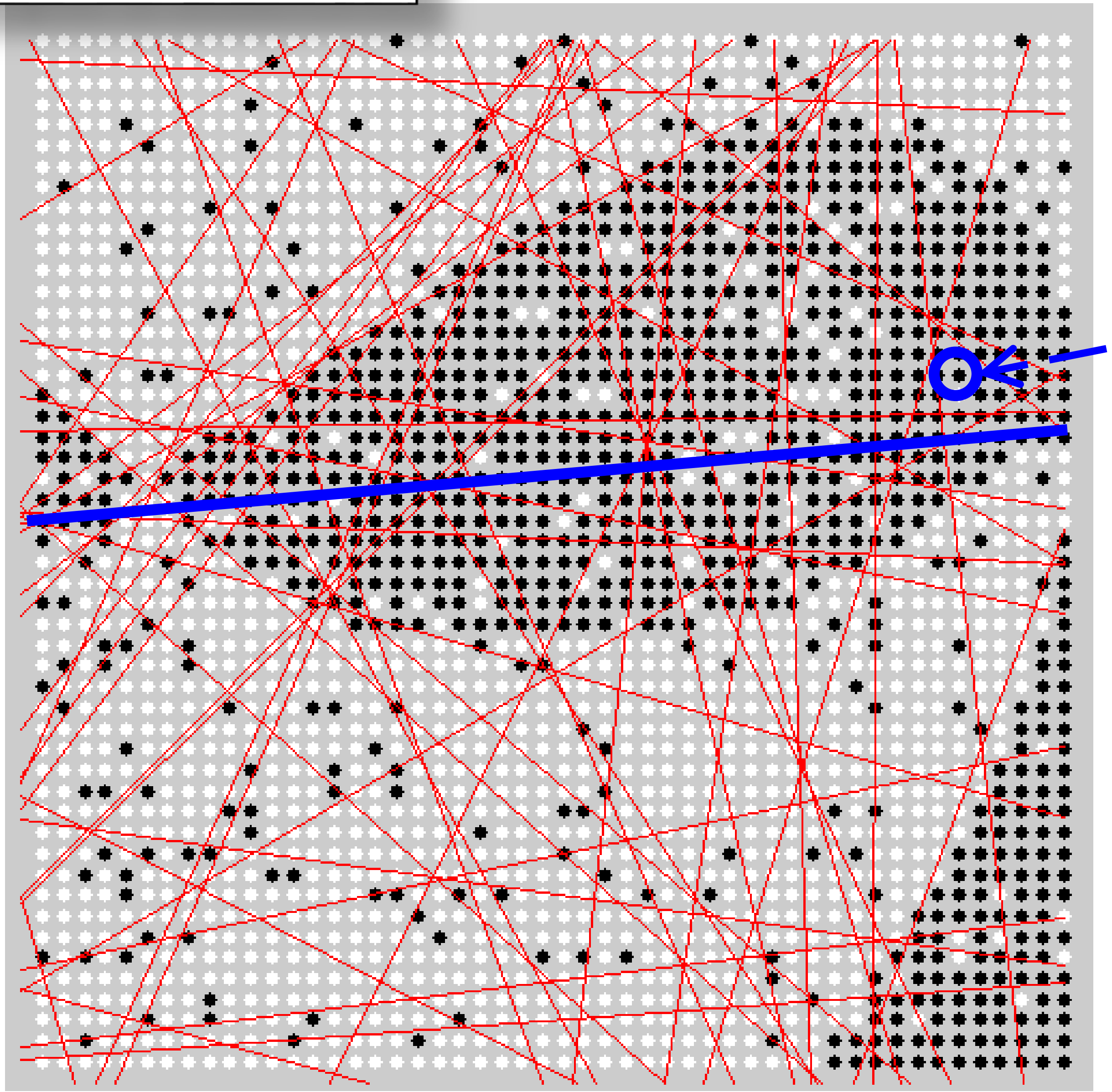
Throw 50 “straws” and label the sides so that the accuracy is greater than 0.5



Train 50 linear classifiers on bootstrap samples



Example: 2 ensembles



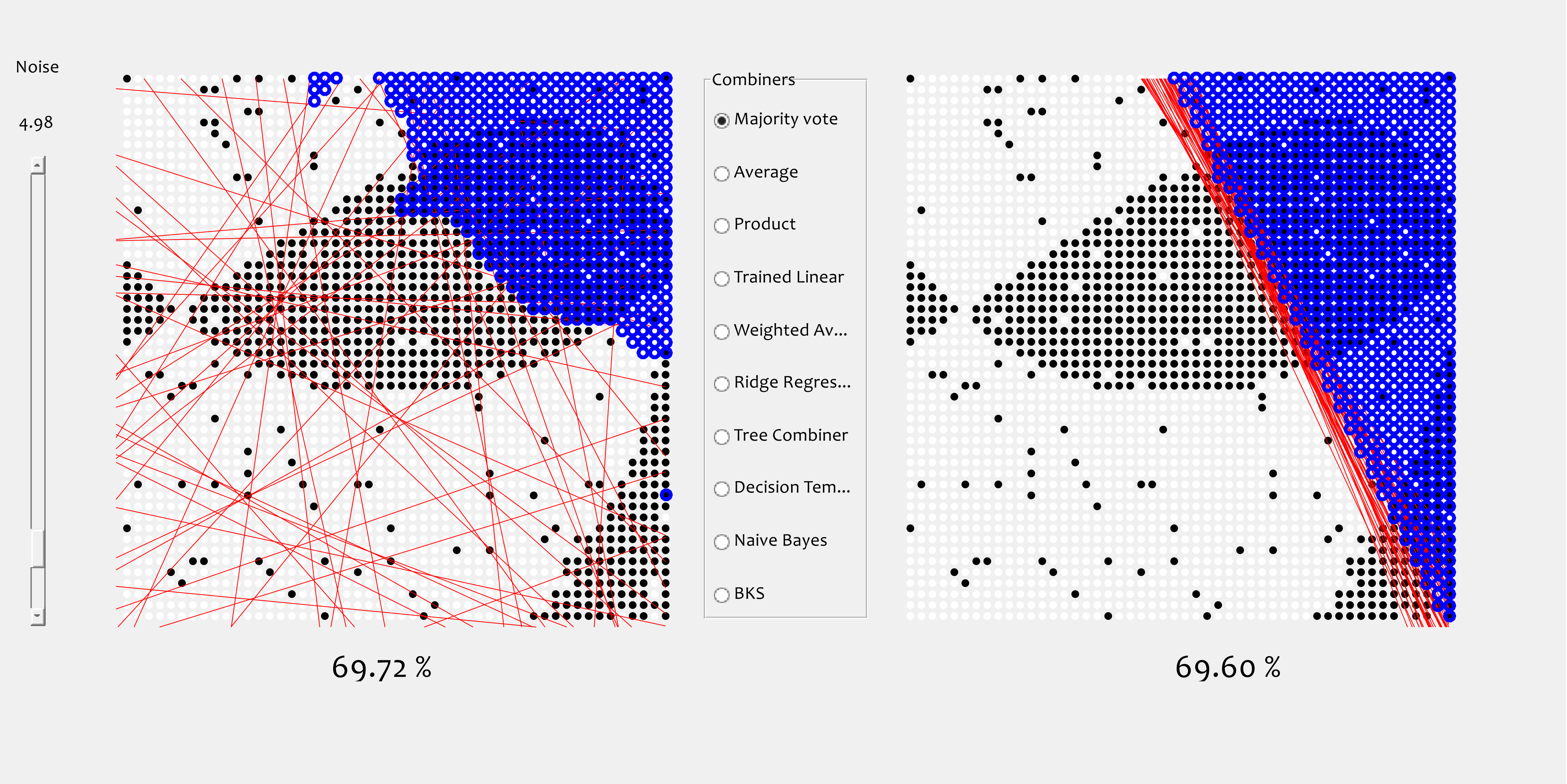
Each classifier returns an estimate for class "Fish"

$P(F \mid [x,y]^T)$

And, of course, we have
$P(\bar{F}|[x,y]^T) = 1 - P(F|[x,y]^T)$
but we will not need this.

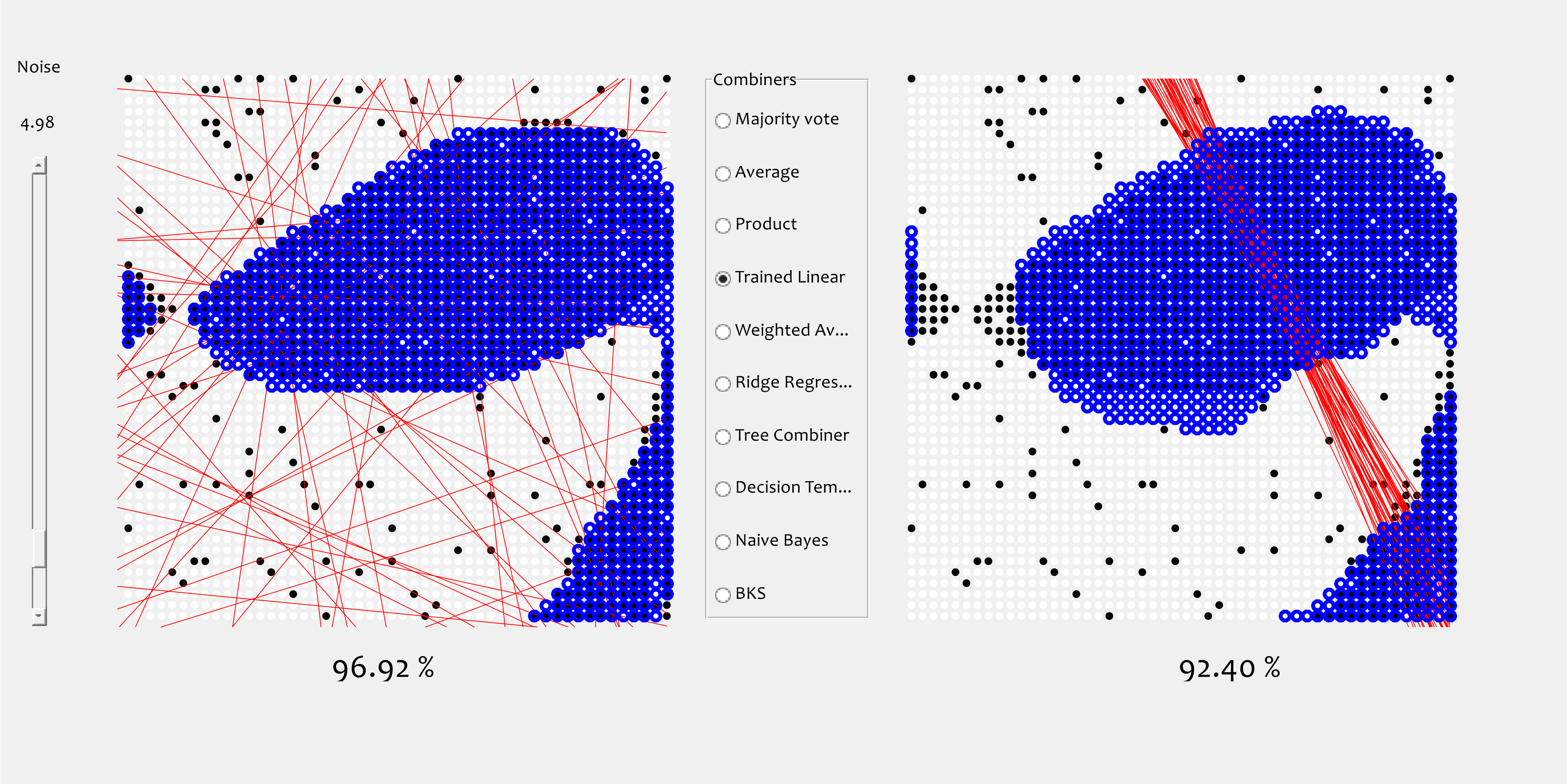
Example: 2 ensembles

5% label noise –
Majority vote



Example: 2 ensembles

5% label noise – trained linear combiner



Example: 2 ensembles

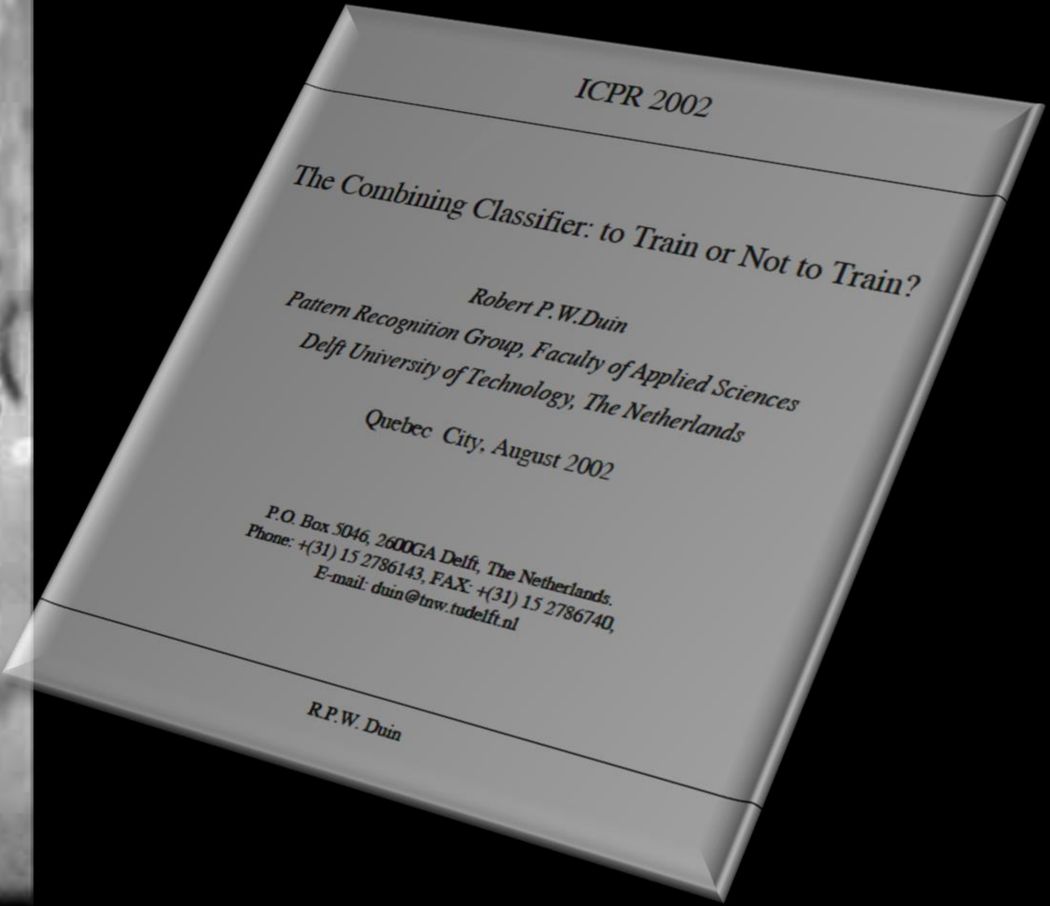
What does the example show?

- The combiner matters (a lot)
- The trained combiner works better

However, nothing is as simple as it looks...

The Combining Classifier: to Train or Not to Train?

http://samcnitt.tumblr.com/



ICPR 2002

The Combining Classifier: to Train or Not to Train?

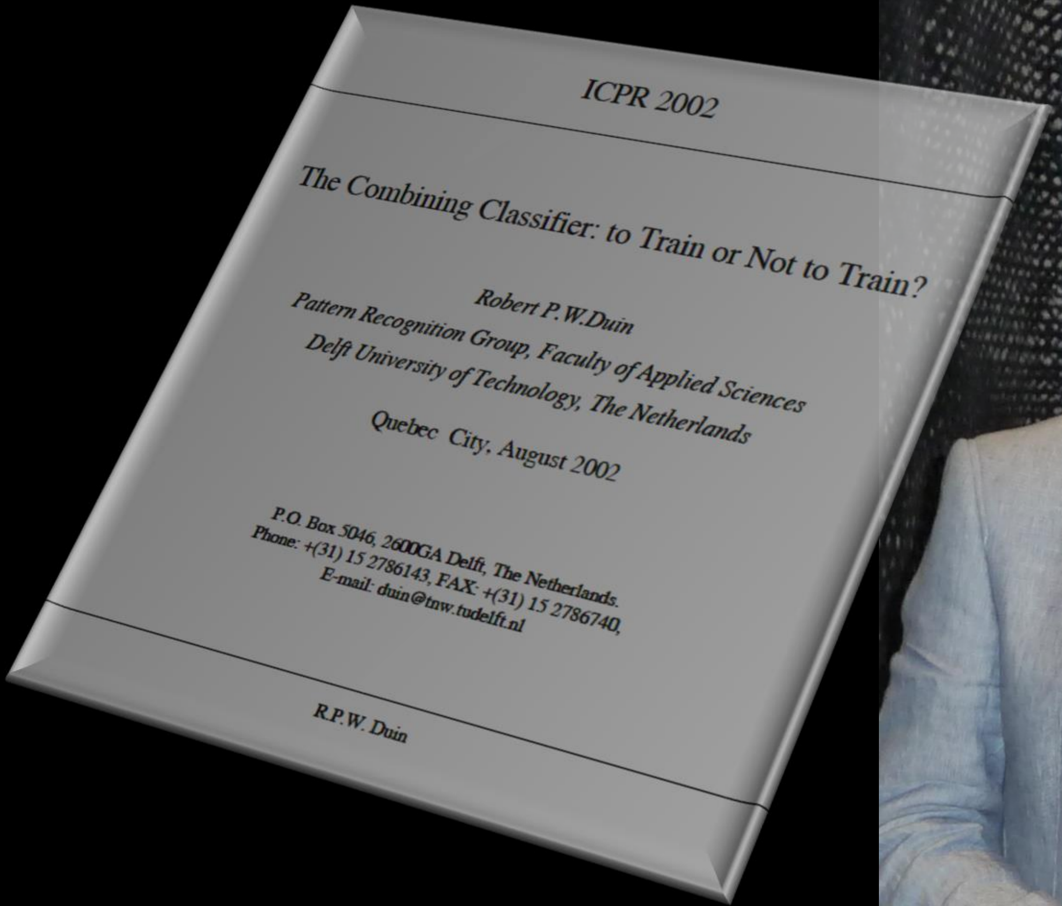
Robert P.W.Duin
Pattern Recognition Group, Faculty of Applied Sciences
Delft University of Technology, The Netherlands

Quebec City, August 2002

P.O. Box 5046, 2600GA Delft, The Netherlands.
Phone: +(31) 15 2786143, FAX: +(31) 15 2786740,
E-mail: duin@tnw.tudelft.nl

R.P.W. Duin

The Combining Classifier: to Train or Not to Train?


ICPR 2002

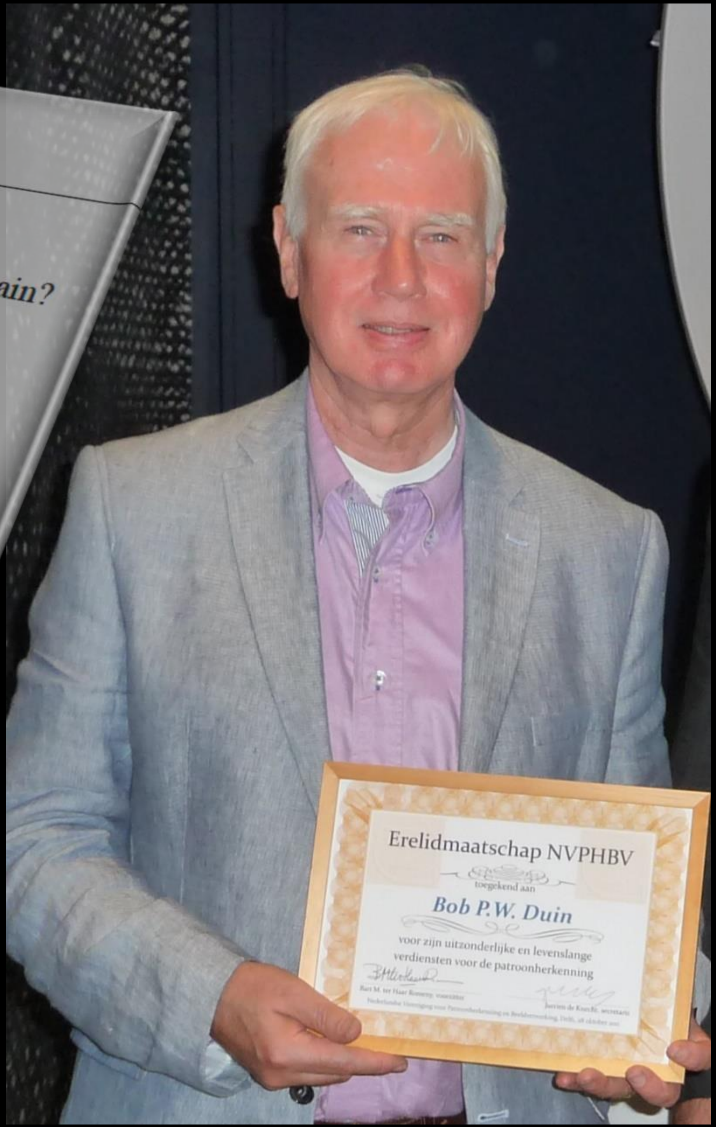
The Combining Classifier: to Train or Not to Train?

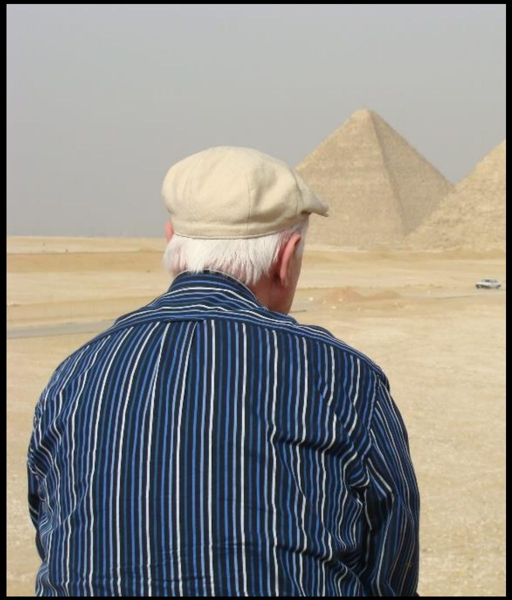
Robert P.W.Duin
Pattern Recognition Group, Faculty of Applied Sciences
Delft University of Technology, The Netherlands

Quebec City, August 2002

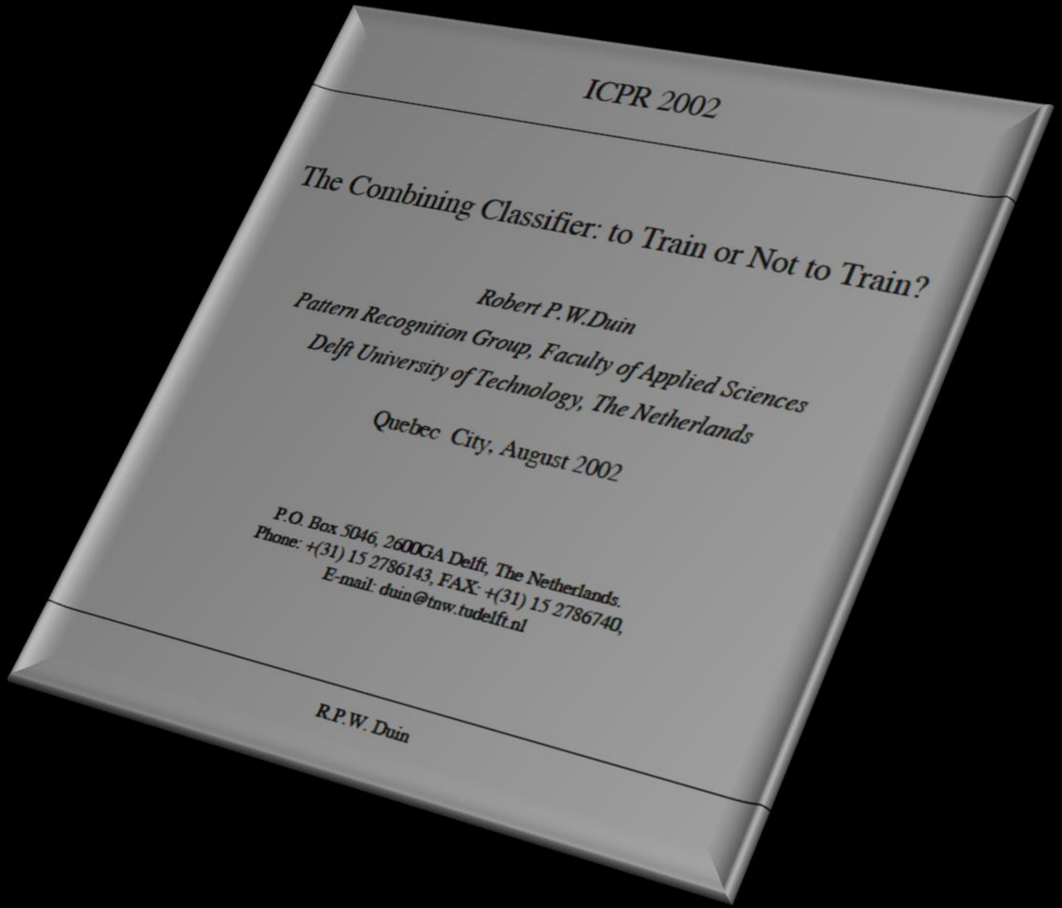
P.O. Box 5046, 2600GA Delft, The Netherlands.
Phone: +(31) 15 2786143, FAX: +(31) 15 2786740,
E-mail: duin@tnw.tudelft.nl

R.P.W. Duin





The Combining Classifier: to Train or Not to Train?



ICPR 2002

The Combining Classifier: to Train or Not to Train?

Robert P.W.Duin

Pattern Recognition Group, Faculty of Applied Sciences

Delft University of Technology, The Netherlands

Quebec City, August 2002

P.O. Box 5046, 2600GA Delft, The Netherlands.

Phone: +(31) 15 2786143, FAX: +(31) 15 2786740,

E-mail: duin@tnw.tudelft.nl

R.P.W. Duin

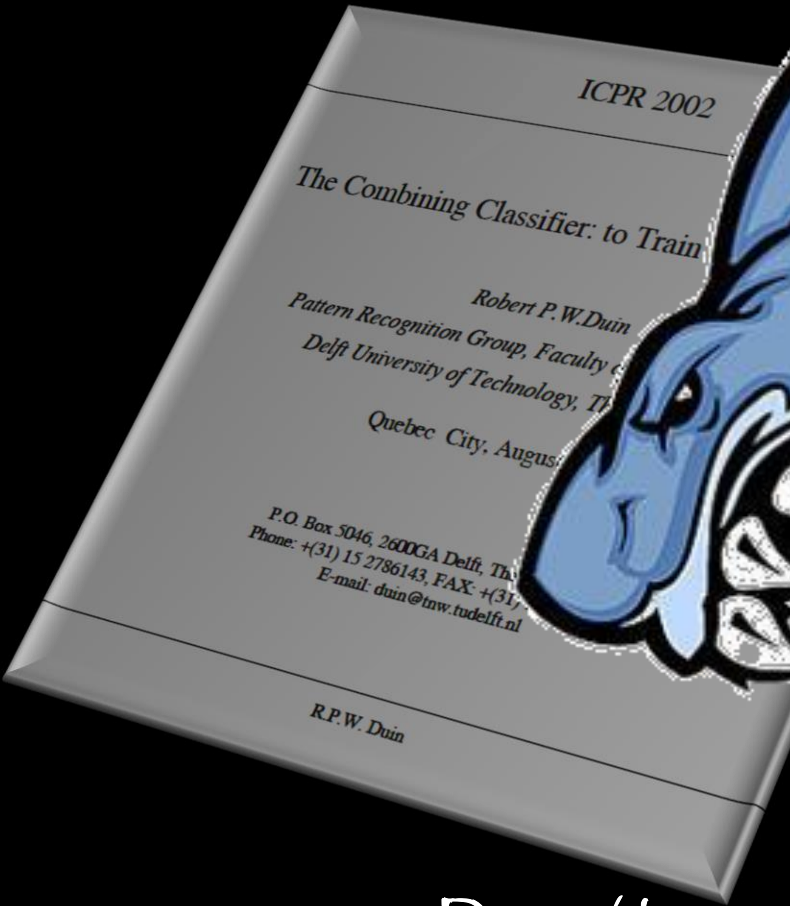
Train the COMBINER if you have "enough" data!

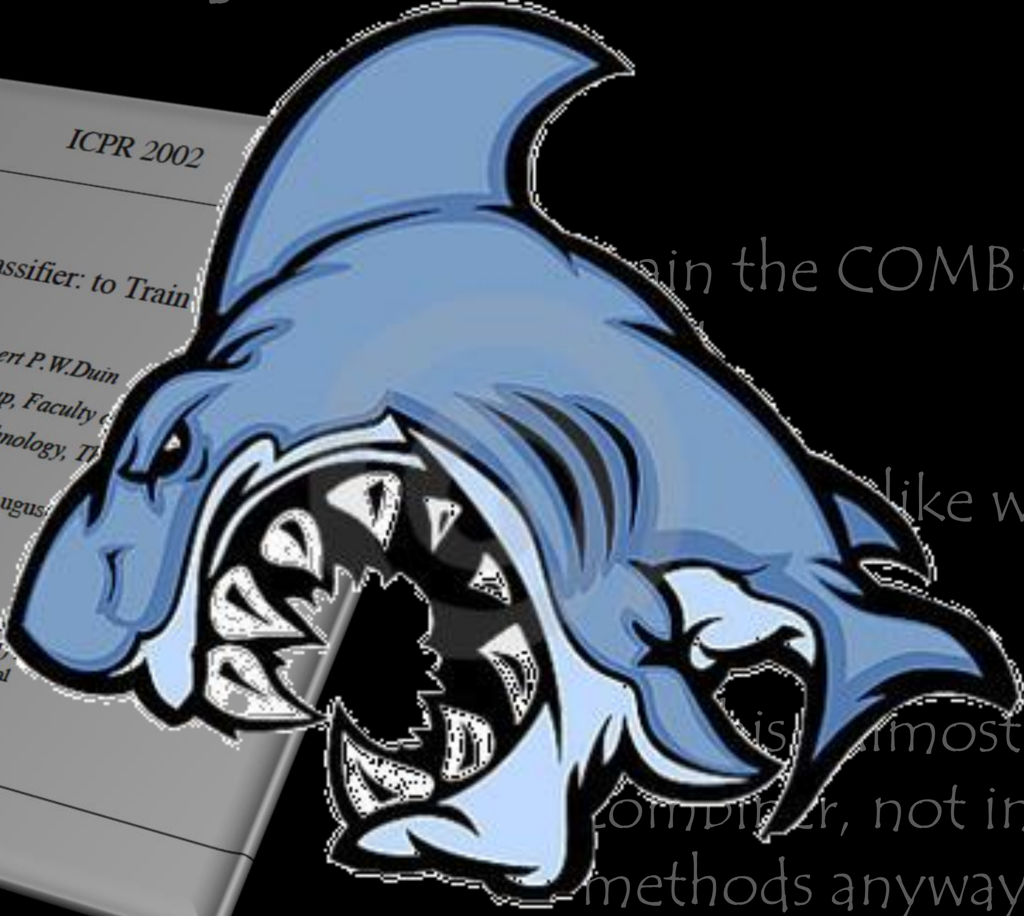
Otherwise, like with any classifier, we may over-fit the data.

Get this: Almost NOBODY trains the combiner, not in the CLASSIC ensemble methods anyway.

Ha-ha-ha, what is "enough" data?

The Combining Classifier: to Train or Not to Train?


ICPR 2002
The Combining Classifier: to Train
Robert P.W.Duin
Pattern Recognition Group, Faculty
Delft University of Technology, Th
Quebec City, Augus
P.O. Box 5046, 2600GA Delft, Th
Phone: +(31) 15 2786143, FAX: +(31)
E-mail: duin@tnw.tudelft.nl
R.P.W. Duin



ain the COMBINER if you have “enough”

like with any classifier, we may over-

is almost NOBODY trains the

ombiner, not in the CLASSIC ensemble

methods anyway.

Don’t you worry… BIG DATA is coming

Ha-ha-ha, what is “**enough**” data?

Train the combiner and live happily ever after!

"Anchor" points

2. Diversity

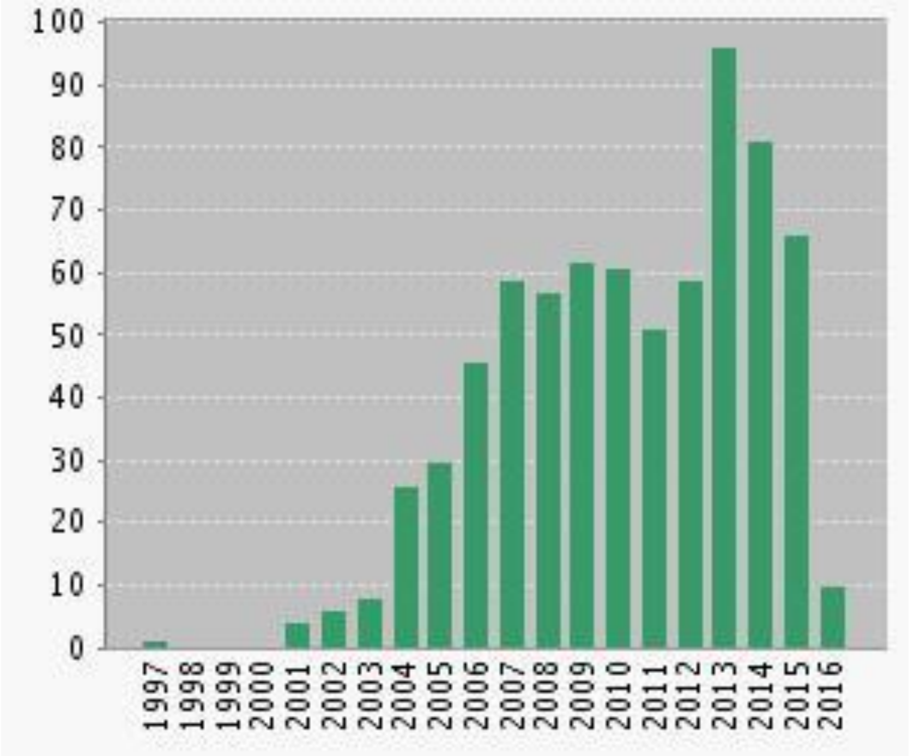
All ensemble methods we have seen so far strive to keep the individual accuracy high while increasing diversity.

- How can we measure diversity?
- WHAT can we do with the diversity value?
  - Compare ensembles
  - Explain why a certain ensemble heuristic works and others don't
  - Construct ensemble by overproducing and selecting classifiers with high accuracy and high diversity

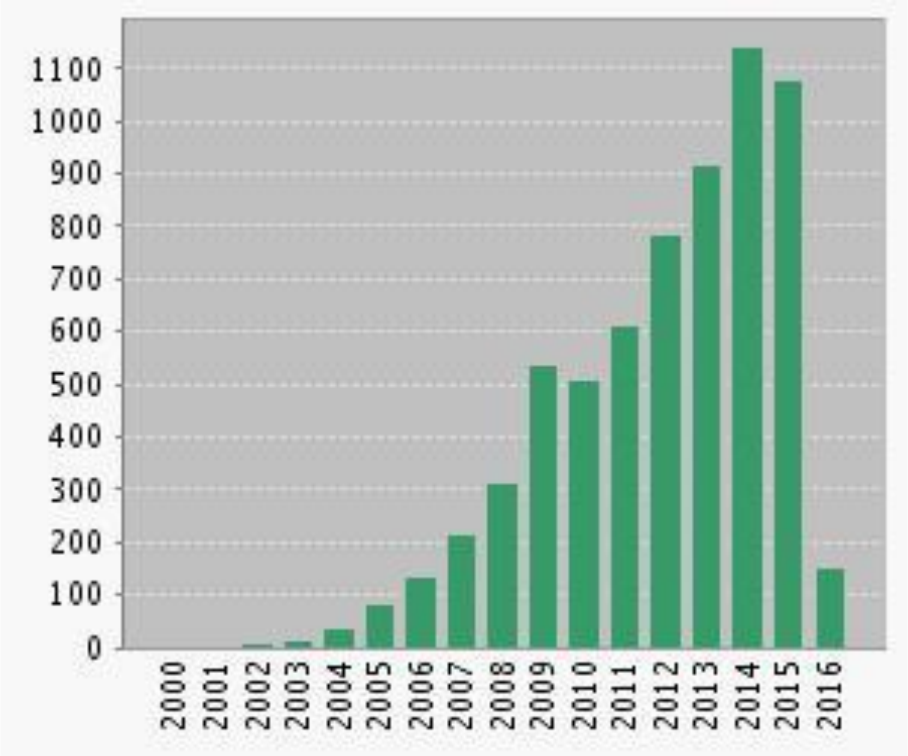
Are we still talking about diversity?

classifier + ensemble + diversity: 713 papers, 6543 citations

Published each year (713)



Cited each year (6543)



Search on 22 Feb 2016

Measure diversity for a PAIR of classifiers

independent outputs ≠ independent errors
hence, use ORACLE outputs

		Classifier 2	
		correct	wrong
Classifier 1	correct	a	b
	wrong	c	d

Number of instances labelled correctly by classifier 1 and mislabelled by classifier 2

Diversity

Measure diversity for a PAIR of classifiers

Objects	C1	C2
1	1	1
2	0	0
3	1	1
4	1	1
5	1	0
6	1	1
7	0	1
8	1	0

Classifier 2

Classifier 1	correct	wrong
correct	4	2
wrong	1	1

Diversity

Measure diversity for a PAIR of classifiers

Objects	C1	C2
1	1	1
2	0	0
3	1	1
4	1	1
5	1	1
6	1	1
7	0	0
8	0	0

Not diverse

Classifier 1 \ Classifier 2	correct	wrong
correct	5	0
wrong	0	3

Diversity

Measure diversity for a PAIR of classifiers

Objects	C1	C2
1	0	1
2	1	0
3	0	1
4	0	1
5	0	1
6	0	1
7	1	0
8	1	0

Totally diverse

Classifier 1 \ Classifier 2	correct	wrong
correct	0	3
wrong	5	0

Diversity

Measure diversity for a PAIR of classifiers

Diversity resides here

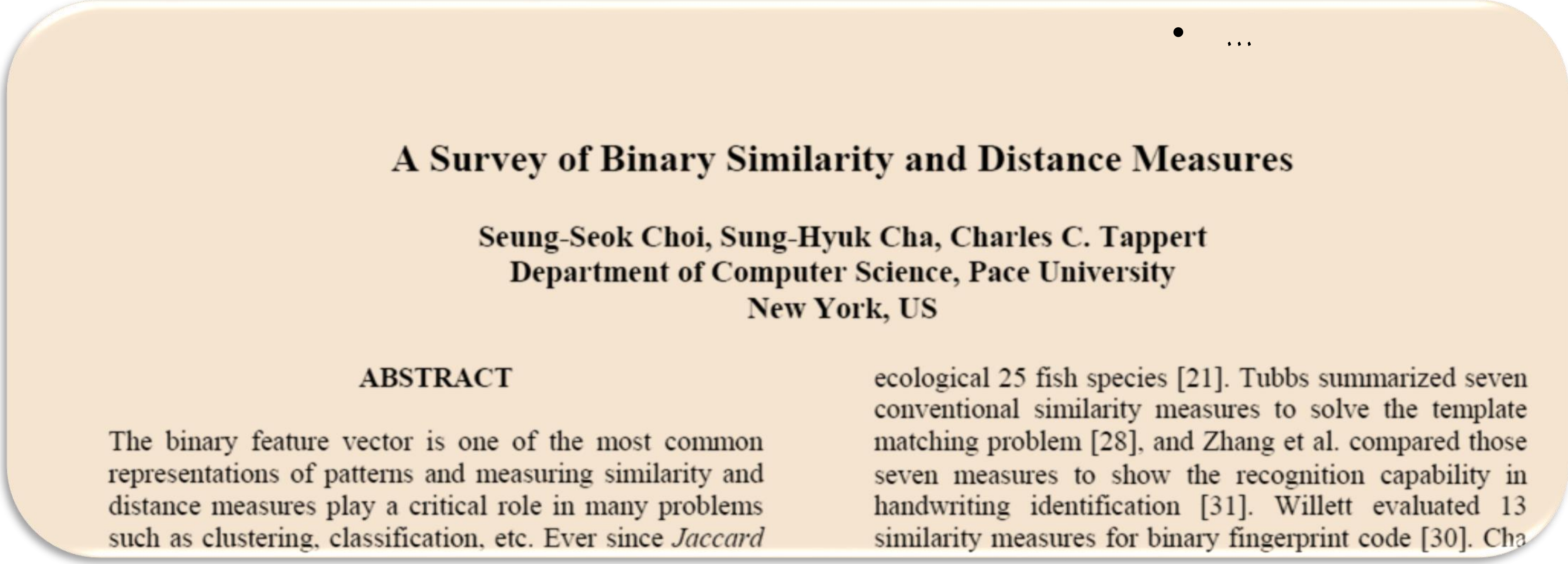
		Classifier 2	
		correct	wrong
Classifier 1	correct	a	b
	wrong	c	$\boxed{d}$

Joint error

Diversity

		Classifier 2	
		correct	wrong
Classifier 1	correct	a	b
	wrong	c	d

- Q
- kappa
- correlation (rho)
- disagreement
- double fault
- …



A Survey of Binary Similarity and Distance Measures

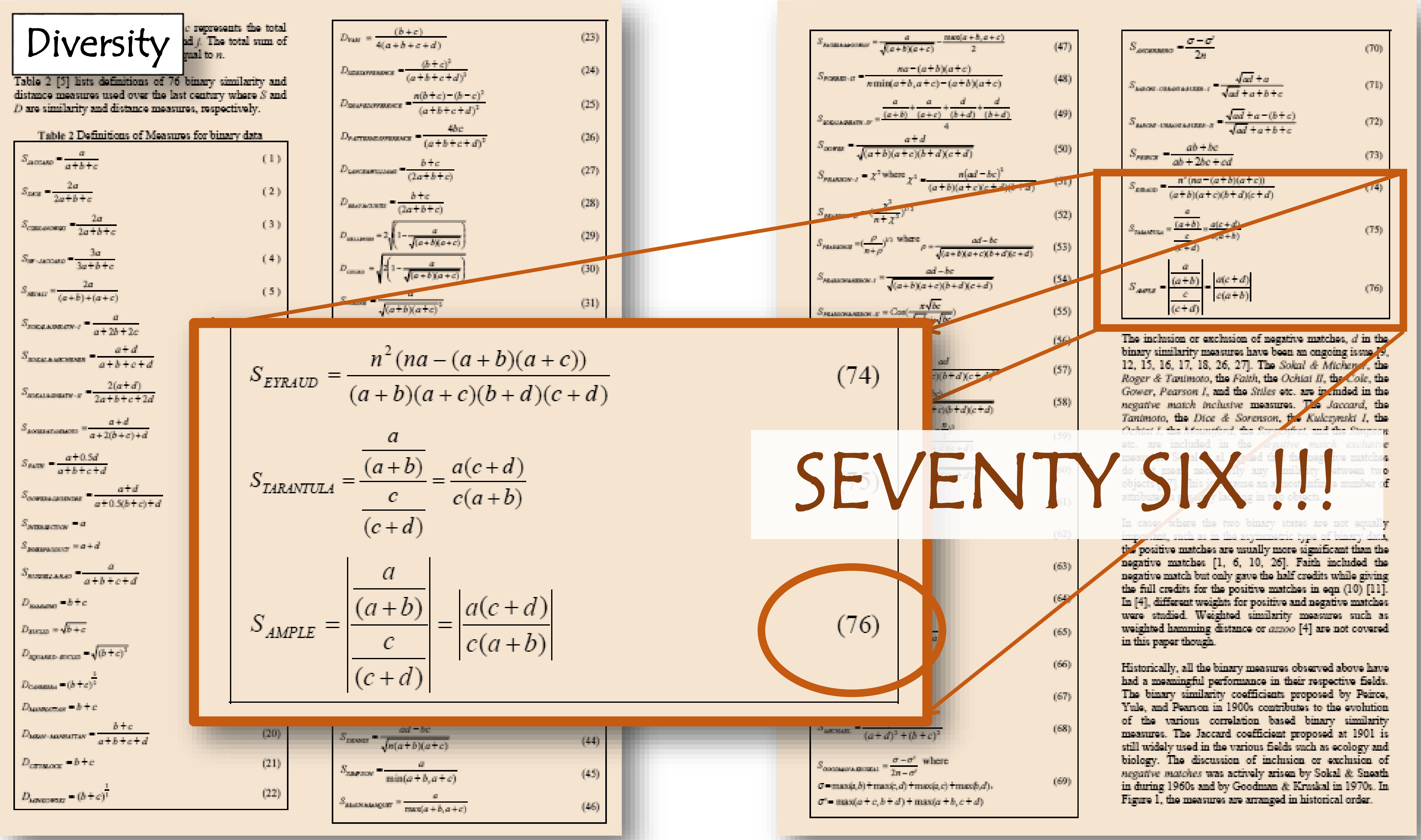
Seung-Seok Choi, Sung-Hyuk Cha, Charles C. Tappert
Department of Computer Science, Pace University
New York, US

ABSTRACT

The binary feature vector is one of the most common representations of patterns and measuring similarity and distance measures play a critical role in many problems such as clustering, classification, etc. Ever since *Jaccard*

ecological 25 fish species [21]. Tubbs summarized seven conventional similarity measures to solve the template matching problem [28], and Zhang et al. compared those seven measures to show the recognition capability in handwriting identification [31]. Willett evaluated 13 similarity measures for binary fingerprint code [30]. Cha

Diversity



$$S_{EYRAUD} = \frac{n^2(na-(a+b)(a+c))}{(a+b)(a+c)(b+d)(c+d)} \quad (74)$$

$$S_{TARANTULA} = \frac{\dfrac{a}{(a+b)}}{\dfrac{c}{(c+d)}} = \frac{a(c+d)}{c(a+b)}$$

$$S_{AMPLE} = \left|\frac{\dfrac{a}{(a+b)}}{\dfrac{c}{(c+d)}}\right| = \left|\frac{a(c+d)}{c(a+b)}\right| \quad (76)$$

SEVENTY SIX !!!

Do we need more "NEW" pairwise diversity measures?

Looks like we don't...

And the same holds for non-pairwise measures...
Far too many already.

Take just ONE measure – kappa – $\varkappa$ – not because it is "the best" but because one is enough.

Kappa-error diagrams

- proposed by Margineantu and Dietterich in 1997
- visualise individual accuracy and diversity in a 2-dimensional plot
- have been used to decide which ensemble members can be pruned without much harm to the overall performance

Kappa-error diagrams

		C2	
		correct	wrong
C1	correct	a	b
	wrong	c	d

error $$e_1 = \frac{c+d}{a+b+c+d} \quad ; \quad e_2 = \frac{b+d}{a+b+c+d}$$

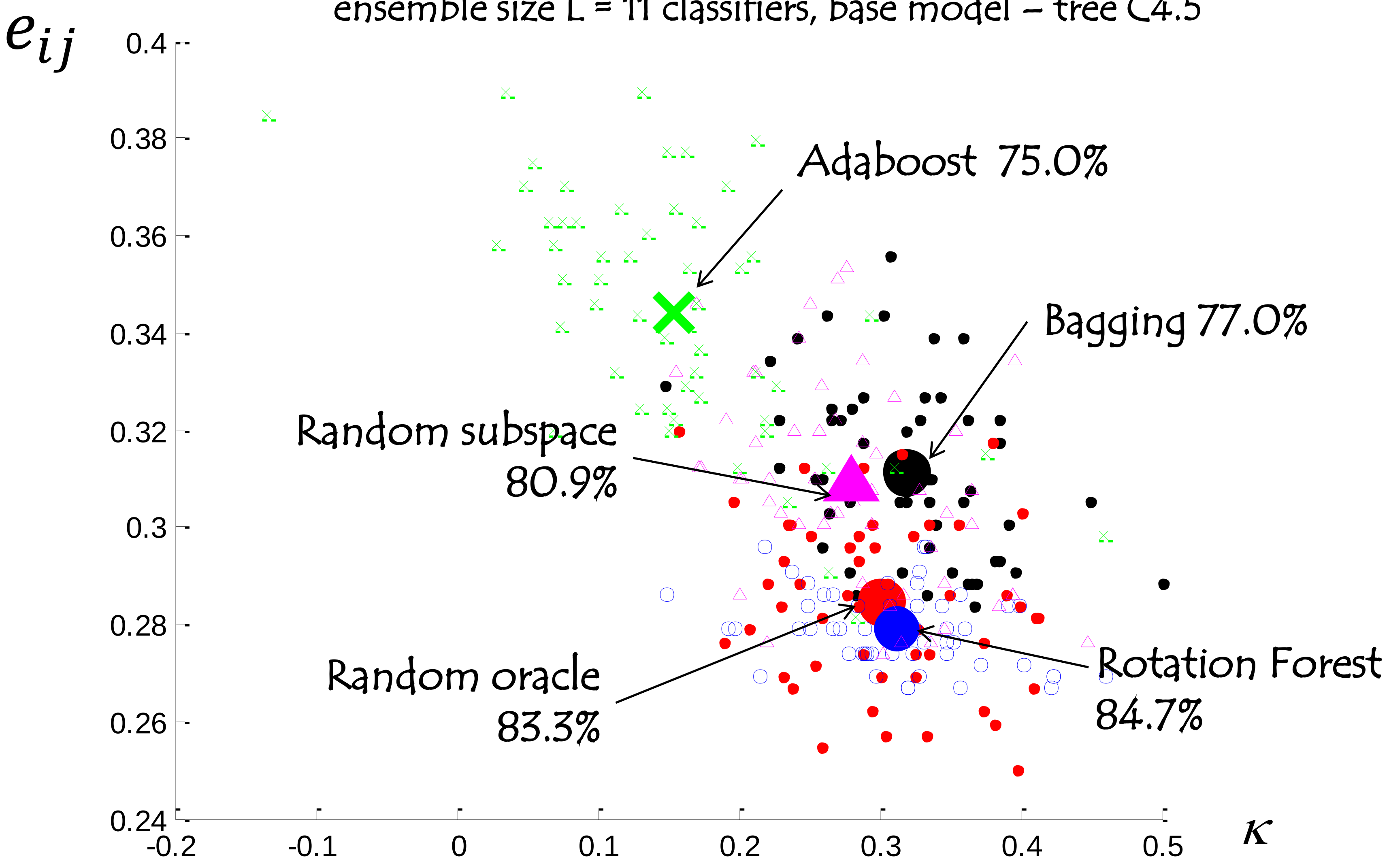
$$e = \frac{b+c+2d}{2(a+b+c+d)}$$

kappa = (observed – chance)/(1-chance)

$$\kappa = \frac{2\ (ad-bc)}{(a+b)(b+d)+(a+c)(c+d)}$$

Example

sonar data (UCI): 260 instances, 60 features, 2 classes,
ensemble size L = 11 classifiers, base model – tree C4.5



Kappa-error diagrams

error $$e = \frac{b+c+2d}{2(a+b+c+d)}$$

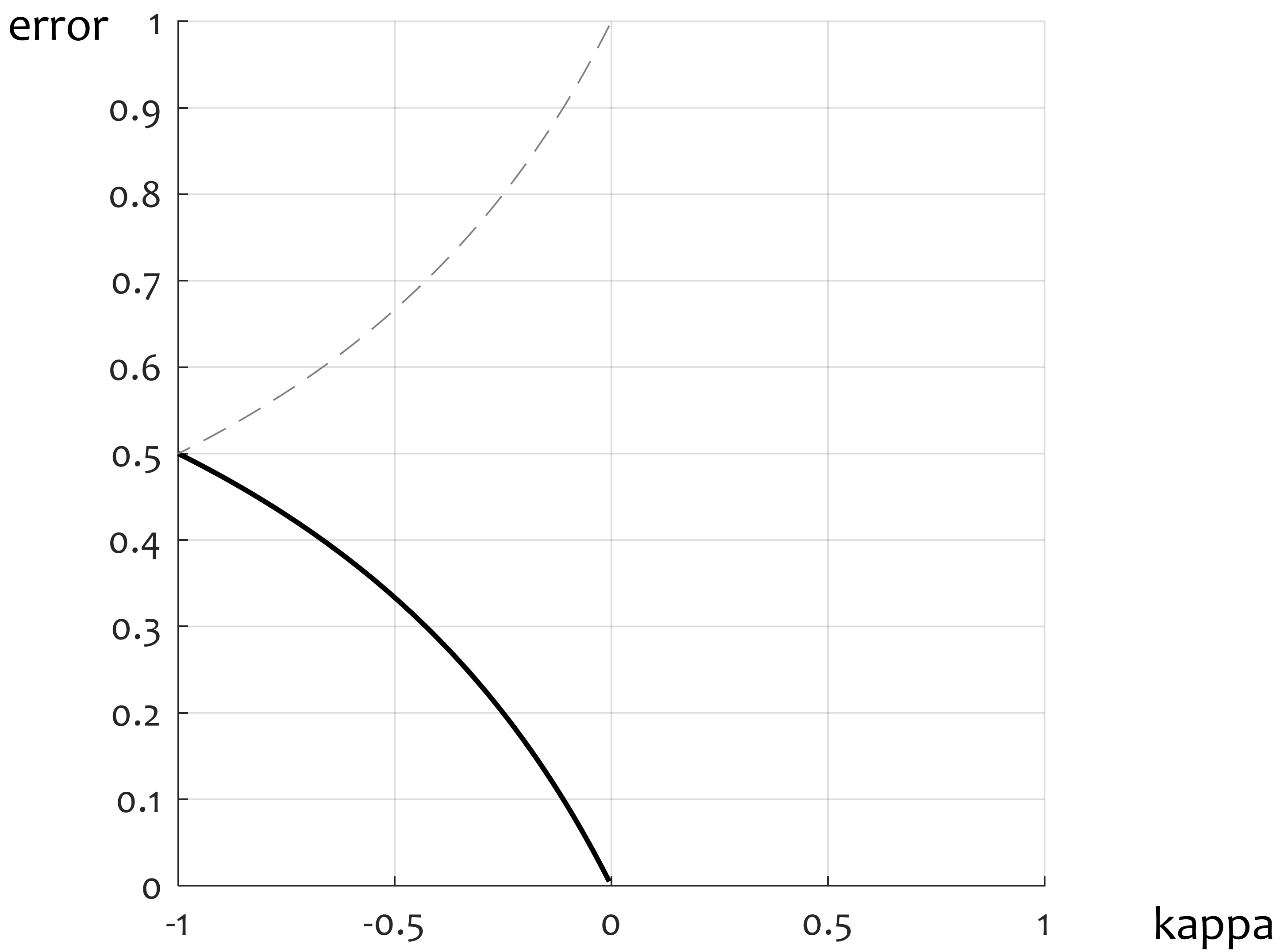
kappa $$\kappa = \frac{2\ (ad-bc)}{(a+b)(b+d)+(a+c)(c+d)}$$

bound (tight)

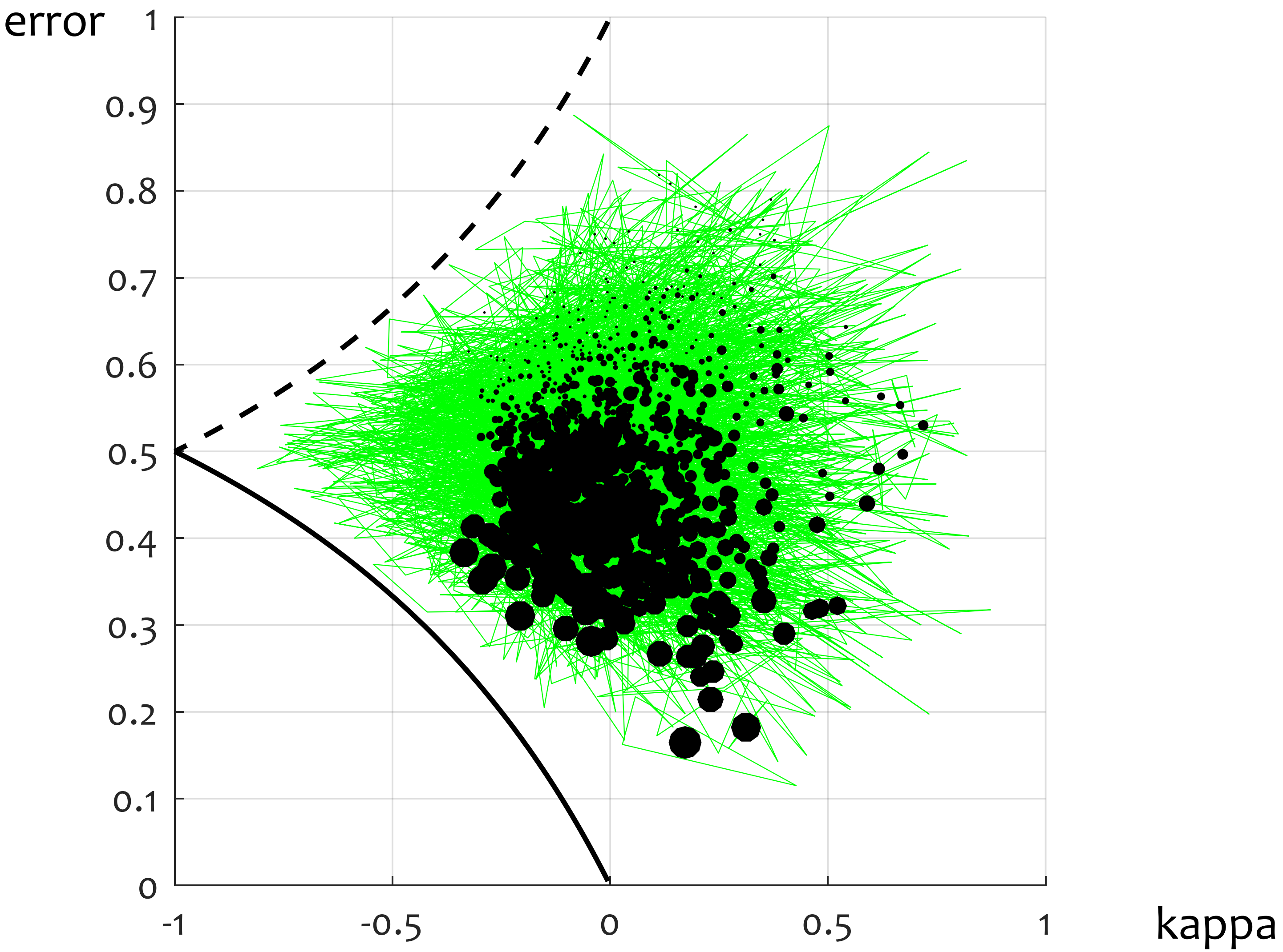
$$\kappa_{\min} = \begin{cases} 1-\dfrac{1}{1-e}, & 0 < e \le 0.5 \\ \dfrac{1}{1-e}, & 0.5 < e \le 1 \end{cases}$$

Kuncheva L.I., A bound on kappa-error diagrams for analysis of classifier ensembles, IEEE Transactions on Knowledge and Data Engineering, 2013, 25 (3), 494-501.

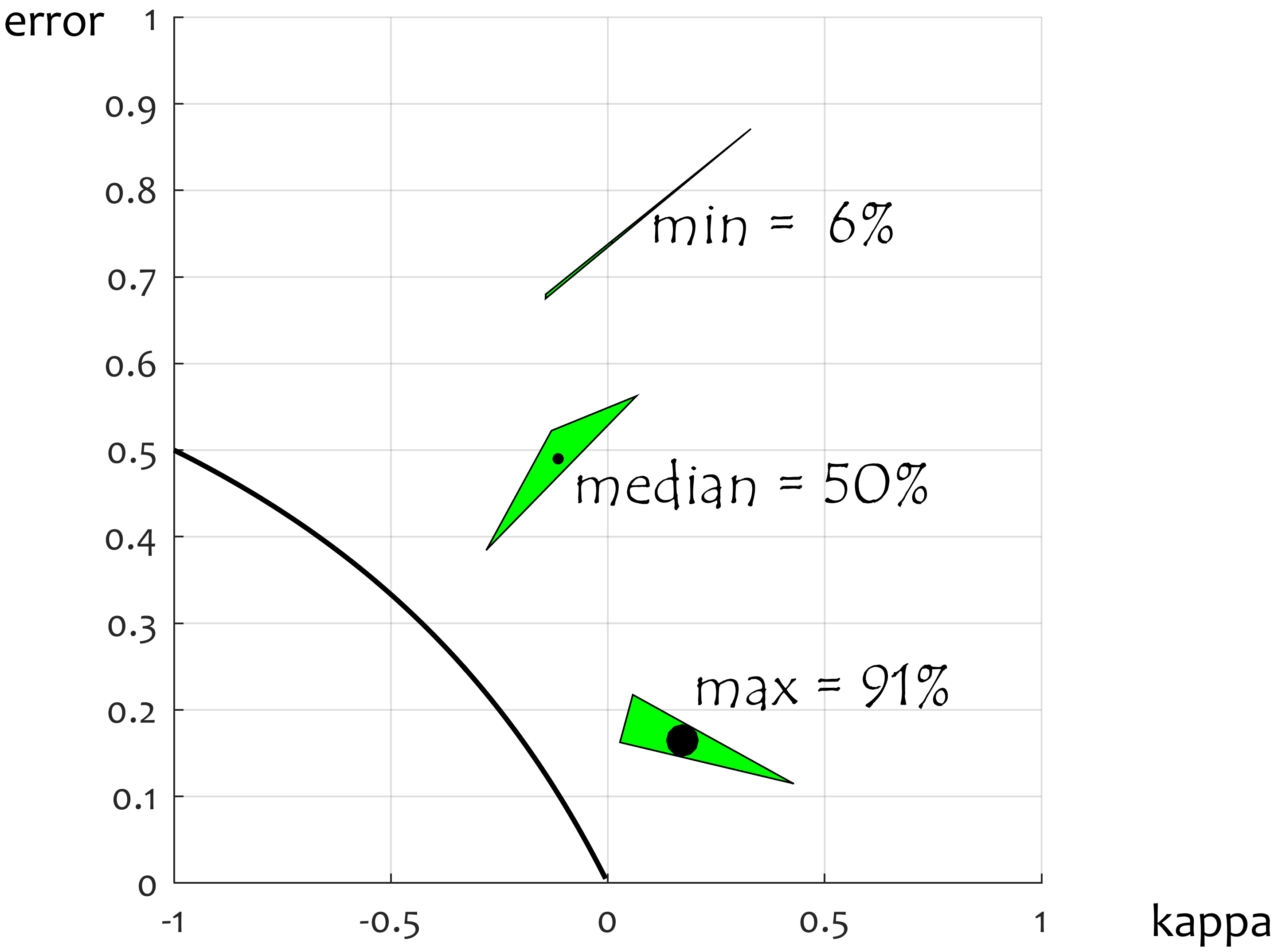
Kappa-error diagrams – bounds



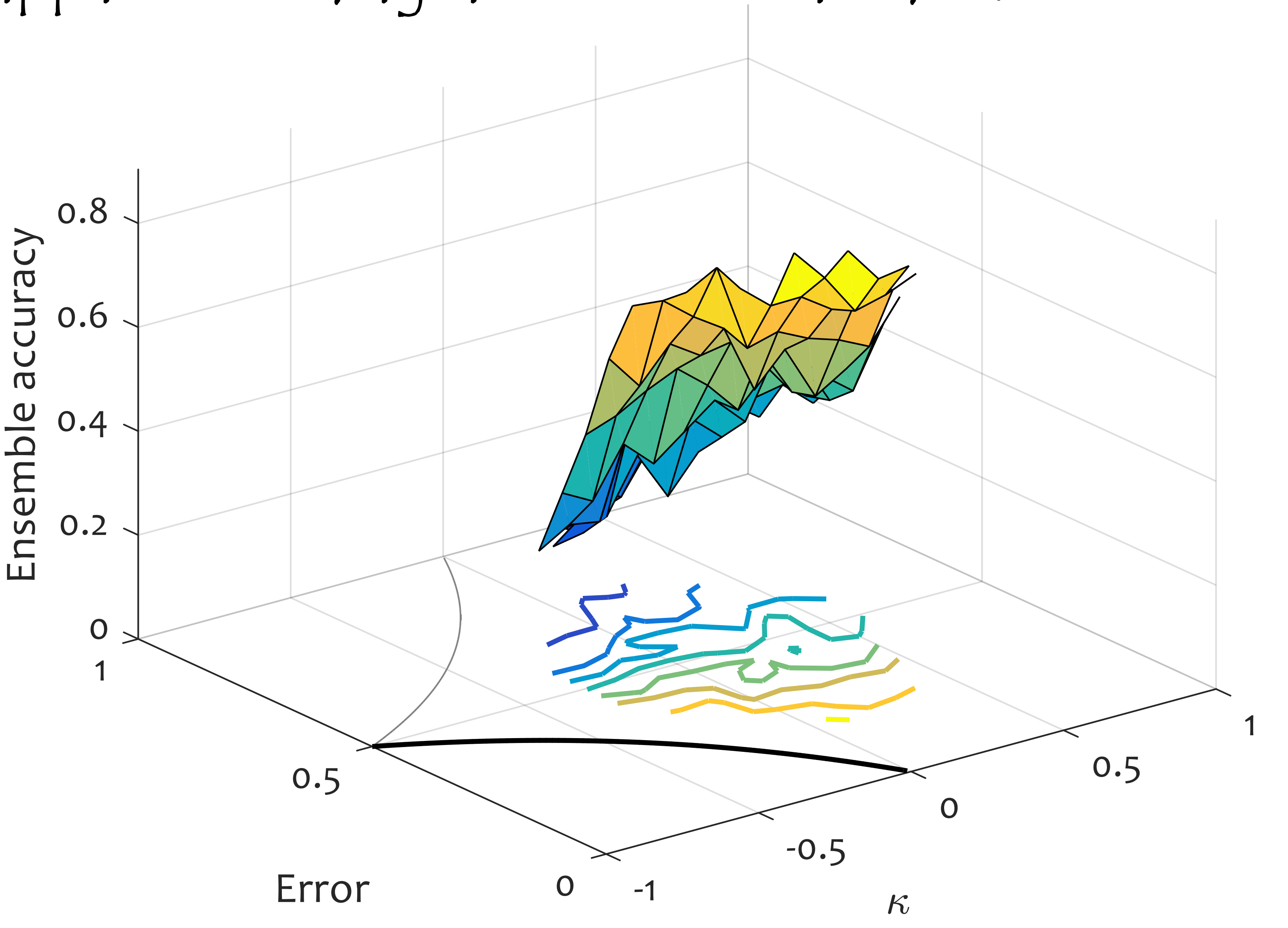
Kappa-error diagrams – simulated ensembles L = 3



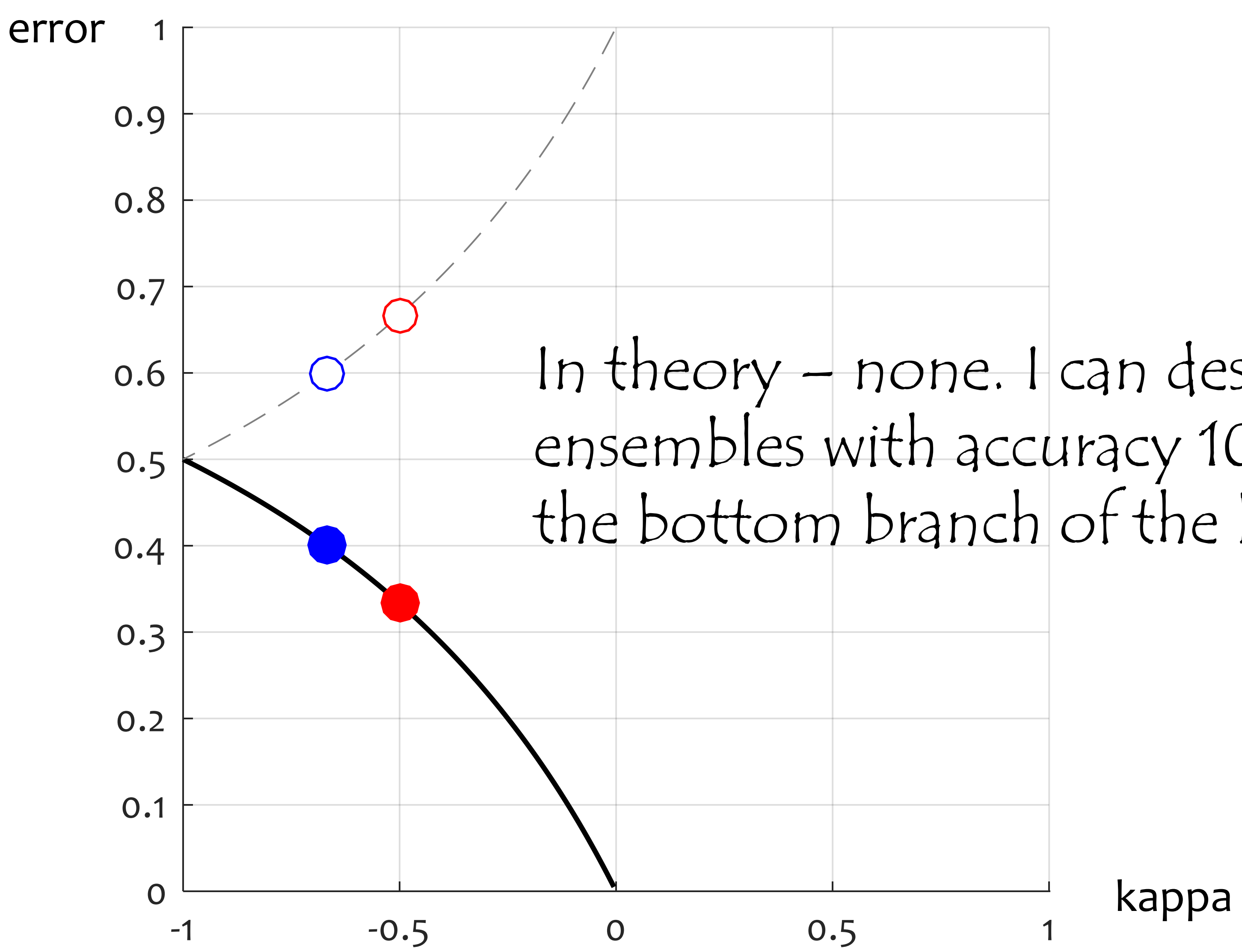
Kappa-error diagrams – simulated ensembles L = 3



Kappa-error diagrams – simulated ensembles L = 3

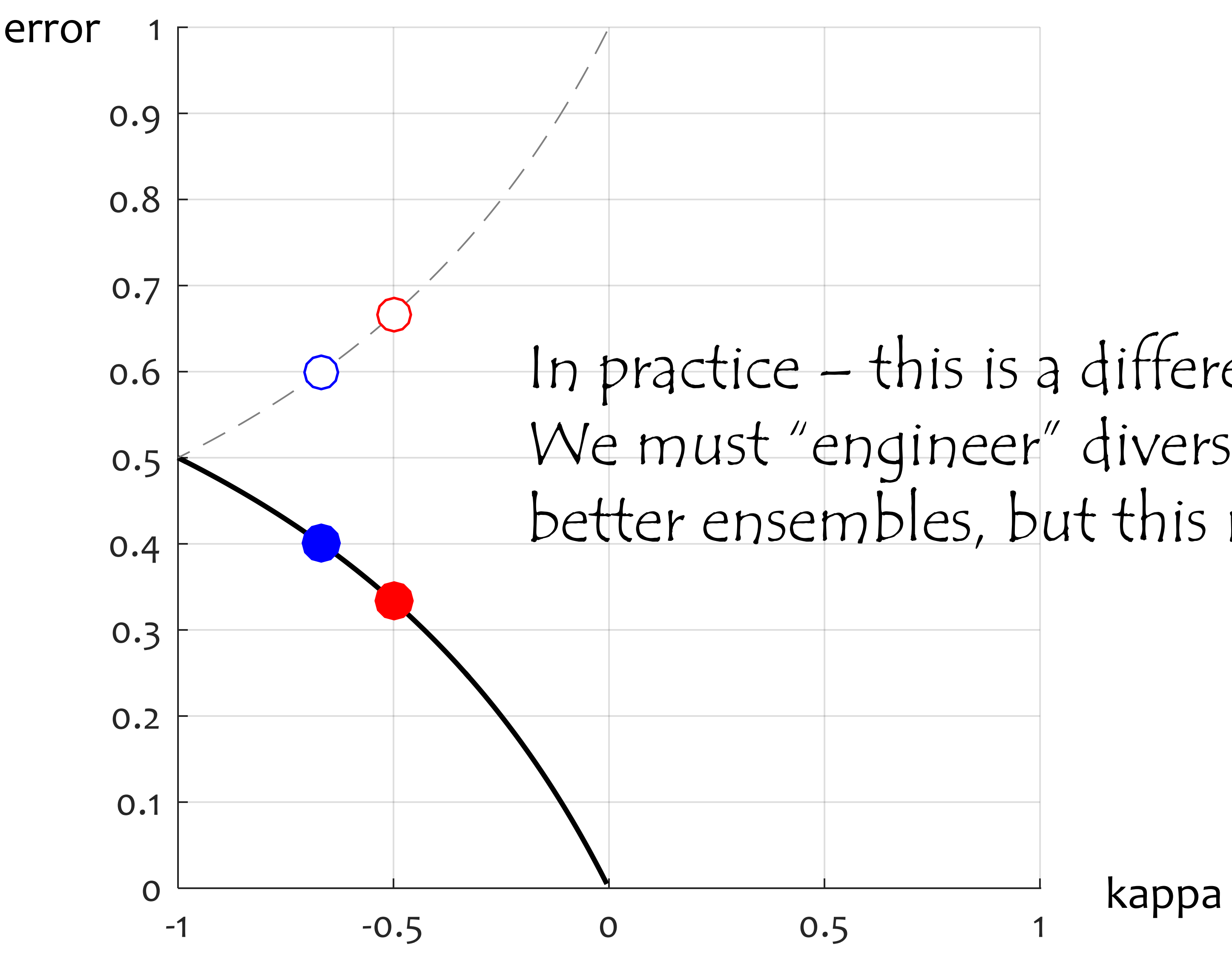


Kappa-error diagrams – How much SPACE do we have to the bound?



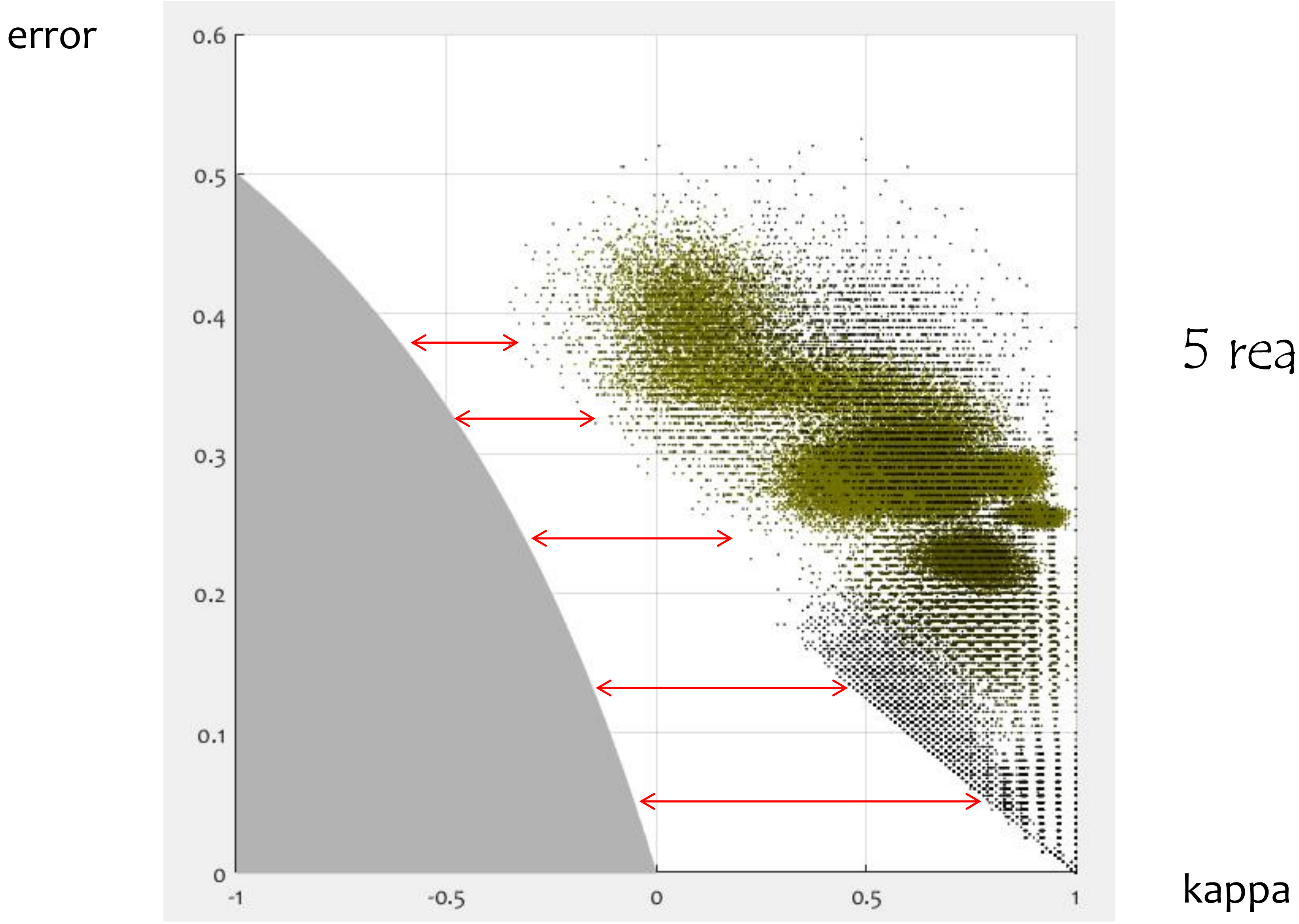
In theory – none. I can design ensembles with accuracy 100%, all on the bottom branch of the bound

Kappa-error diagrams – How much SPACE do we have to the bound?



In practice – this is a different story. We must "engineer" diversity to get better ensembles, but this is not easy...

Kappa-error diagrams – How much SPACE do we have to the bound?



5 real data sets

Is there space for new classifier ensembles?

Looks like yes...

But we need revolutionary ideas about embedding diversity into the ensemble

Why is diversity so baffling?

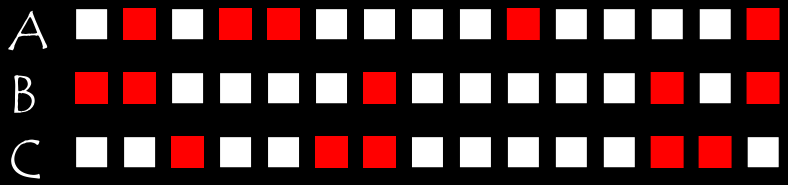
The problem is that diversity is NOT monotonically related to the ensemble accuracy.

In other words, diverse ensembles may be good or may be bad…

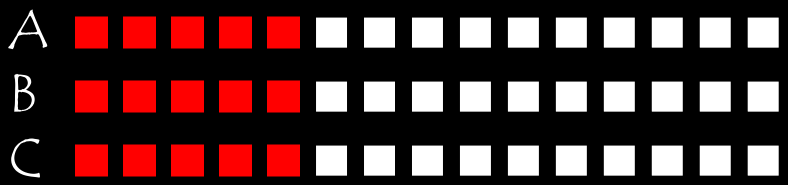
Good and Bad diversity

3 classifiers: A, B, C
15 objects, ■ wrong vote, ■ correct vote
individual accuracy = 10/15 = 0.667
P = ensemble accuracy

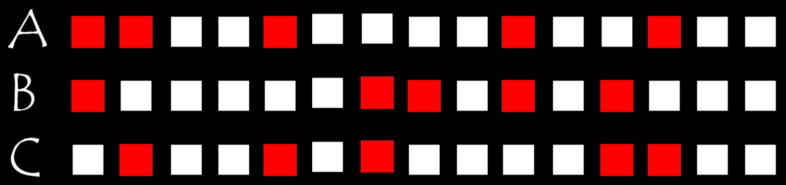
MAJORITY VOTE



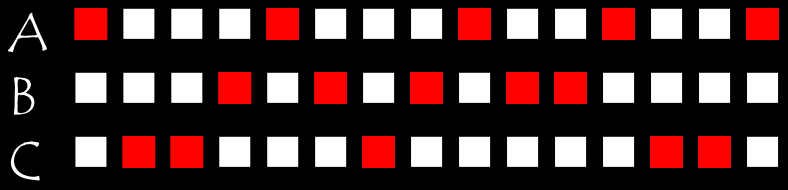
independent classifiers
P = 11/15 = 0.733



identical classifiers
P = 10/15 = 0.667



dependent classifiers 1
P = 7/15 = 0.467

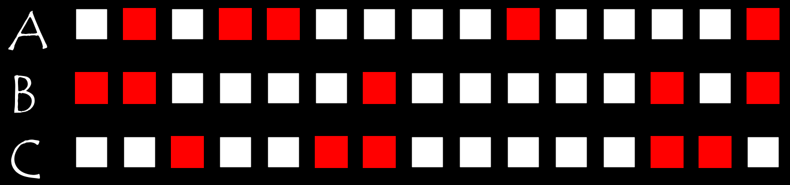


dependent classifiers 2
P = 15/15 = 1.000

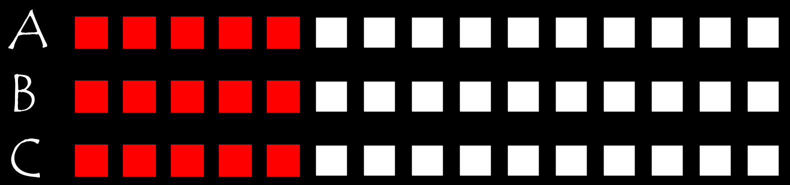
Good and Bad diversity

3 classifiers: A, B, C
15 objects, ■ wrong vote, ■ correct vote
individual accuracy = 10/15 = 0.667
P = ensemble accuracy

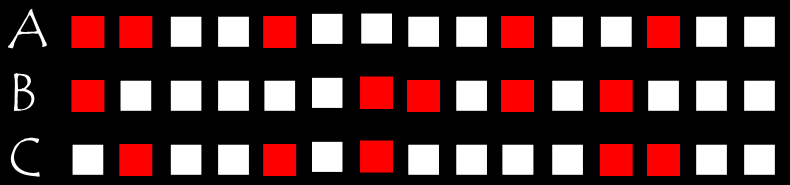
MAJORITY VOTE



independent classifiers
P = 11/15 = 0.733

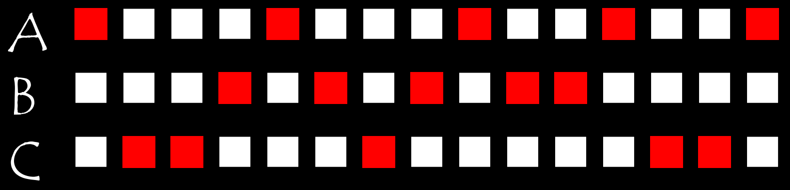


identical classifiers
P = 10/15 = 0.667



dependent classifiers 1
P = 7/15 = 0.467

Bad diversity



dependent classifiers 2
P = 15/15 = 1.000

Good diversity

Good and Bad diversity

l_i	number of classifiers with correct output for z_i
$L - l_i$	number of classifiers with wrong output for z_i
$\bar{p}$	mean individual accuracy
N	number of data points

Decomposition of the Majority Vote Error

$$E_{maj} = (1-\bar{p}) - \frac{1}{NL}\sum_{\text{maj}\,✔}(L-l_i) + \frac{1}{NL}\sum_{\text{maj}\,✘} l_i$$

Individual error

Subtract GOOD diversity

Add BAD diversity

Brown G., L.I. Kuncheva, "Good" and "bad" diversity in majority vote ensembles, Proc. Multiple Classifier Systems (MCS'10), Cairo, Egypt, LNCS 5997, 2010, 124-133.

Good and Bad diversity

✓ ✓ ✓ ✓ ✗ ✗ ✗ This object will contribute $L - l_i$ = (7 – 4) = 3 to good diversity

✓ ✓ ✓ ✗ ✗ ✗ ✗ This object will contribute l_i = 3 to bad diversity

Note that diversity **quantity is 3** in both cases

Ensemble Margin

The <u>voting margin</u> for object z_i is the proportion of <correct minus wrong votes>

$$m_i = \frac{l_i - (L - l_i)}{L}$$

✓ ✓ ✓ ✓ ✗ ✗ ✗

$$m_i = \frac{4 - (7-4)}{7} = \frac{1}{7}$$

POSITIVE

✓ ✓ ✓ ✗ ✗ ✗ ✗

$$m_i = \frac{3 - (7-3)}{7} = -\frac{1}{7}$$

NEGATIVE

Ensemble Margin

Average margin

$$\bar{m} = \frac{1}{N}\sum_{i=1}^{N} m_i = \frac{1}{N}\sum_{i=1}^{N} \frac{l_i - (L - l_i)}{L}$$

Large $\bar{m}$ corresponds to BETTER ensembles…

However, nearly all diversity measures are functions of

Average absolute margin

$$\overline{|m|} = \frac{1}{N}\sum_{i=1}^{N} |m_i|$$

or

Average square margin

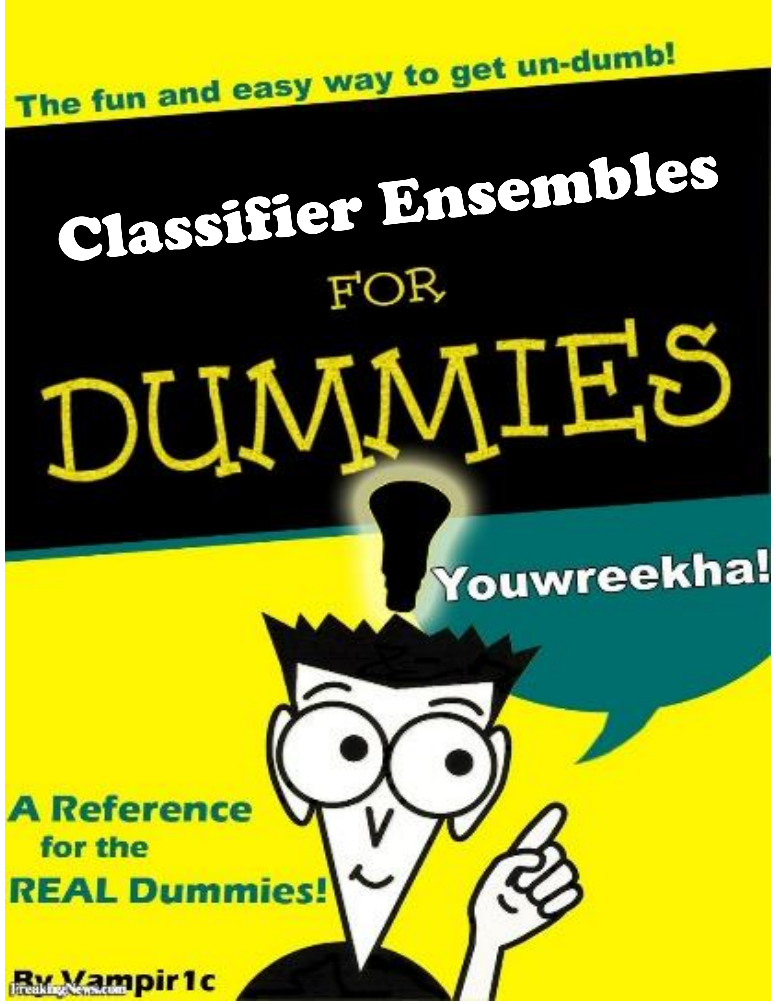
$$\overline{m^2} = \frac{1}{N}\sum_{i=1}^{N} m_i^2$$

The bottom line is:
Diversity is <u>not</u> MONOTONICALLY related to ensemble accuracy

So, stop looking for what is not there…

Where next in classifier ensembles?

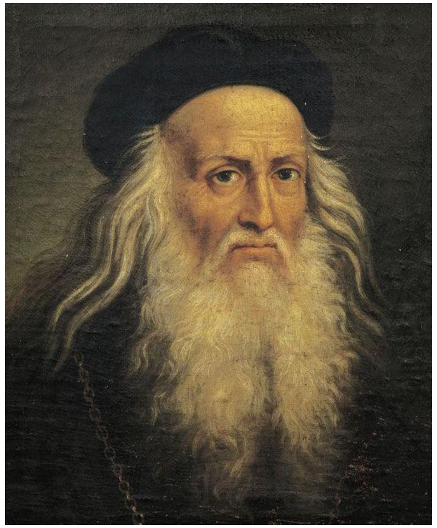
20 years from now, what will stay in the textbooks on classifier ensembles?



We will branch out like every other walk of science

Leonardo da Vinci

1452 - 1519

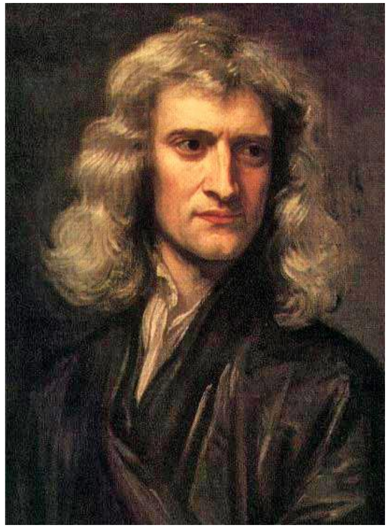


A polymath.

Invention, painting, sculpting, architecture, science, music, and mathematics

Isaac Newton

1643 - 1727



English physicist and mathematician

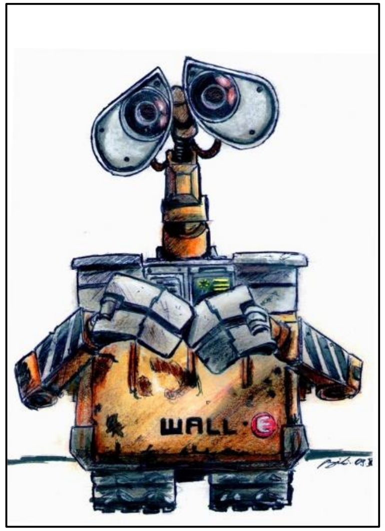
Leo Breiman

1928 - 2005

Statistician

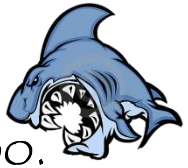
RZ21 4X@#3

2216 - 2354



Classifier ensemblist (Concept drift, imbalanced classes)

Ah, and Big Datist too.



A polymath is a person whose expertise spans a significant number of different subject areas.

Instead of conclusions :)

For the winner

by my favourite illustrator Marcello Barenghi

Well, I'll give you a less crinkled one :)

A guessing game:
CITATIONS
on Web of Science
23/02/2016

"Deep learning"
1,490

"Big data"
8,151

Sort 2-9

from highest to lowest

1. "Classifier Ensemble*" 788
2. (1) + "concept drift"
3. (1) + "rotation forest"
4. (1) + adaboost
5. (1) + (imbalanced or unbalanced)
6. (1) + diversity
7. (1) + combiner
8. (1) + "big data"
9. (1) + "deep learning"

Citation counts

1. "Classifier Ensemble*"	788
2. (1) + "concept drift"	18
3. (1) + "rotation forest"	27
4. (1) + adaboost	78
5. (1) + (imbalanced or unbalanced)	37
6. (1) + diversity	205
7. (1) + combiner	11
8. (1) + "big data"	1
9. (1) + "deep learning"	0

"Deep learning"
1,490

"Big data"
8,151

Solution

"Classifier Ensemble*"
788

"Deep learning"
1,490

"Big data"
8,151

6. (1) + diversity	205
4. (1) + adaboost	78
5. (1) + (imbalanced or unbalanced)	37
3. (1) + "rotation forest"	27
2. (1) + "concept drift"	18
7. (1) + combiner	11
8. (1) + "big data"	1
9. (1) + "deep learning"	0

And thank you for listening to me!

The Leverhulme Trust

Supported by Project RPG-2015-188 Sponsored by the Leverhulme trust