Learning Error-Correcting Representations For Multi-class Problems

Miguel Ángel Bautista Martín

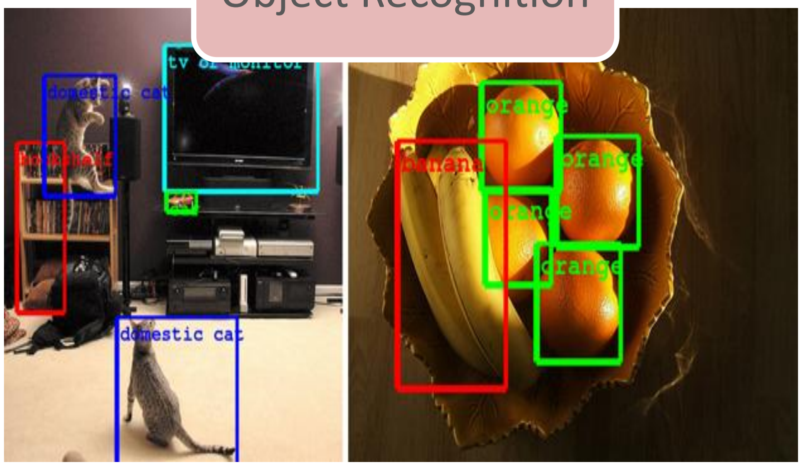
Advisors
Dr. Sergio Escalera
Dr. Oriol Pujol

- The **quintessential goal of Artificial Intelligence** is to build machines that are capable of intelligent behavior, by **perceiving**, interacting and learning from their environment.
- Perceptual related tasks share at its core a decision making process.
- Given some sensorial stimulus and previous experience, choose a single option amongst a defined set of possible decisions.
- Most perception tasks can be interpreted as a **classification-categorization** problems.

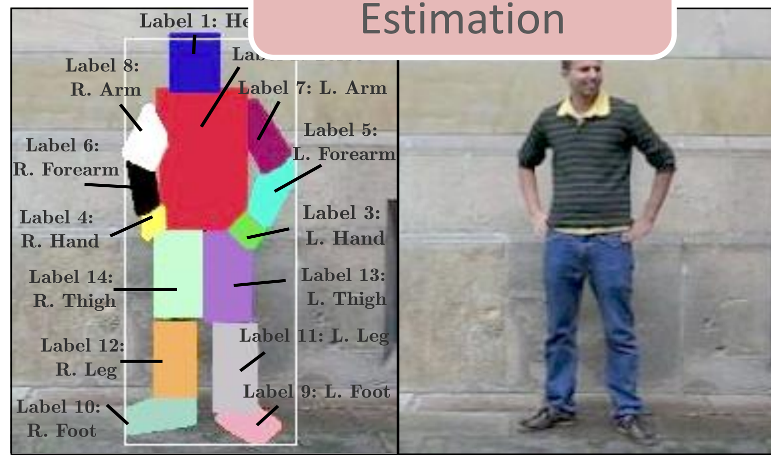


- Classification plays a central role in Computer Vision systems that teach computers how to make sense of images and videos.

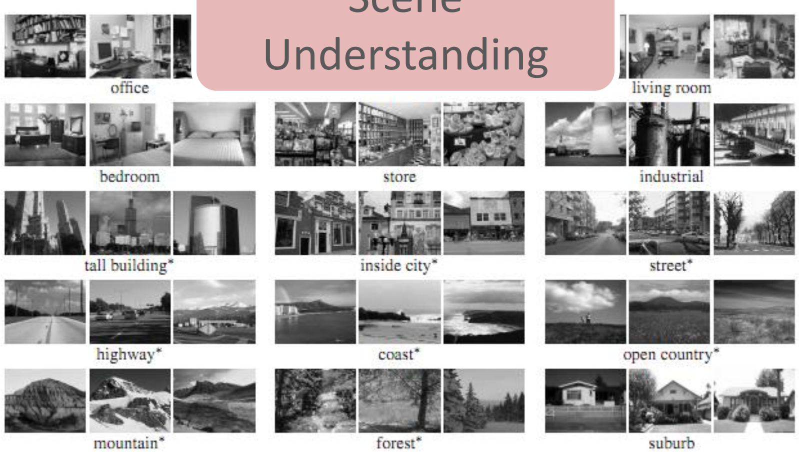
Object Recognition



Human Pose Estimation

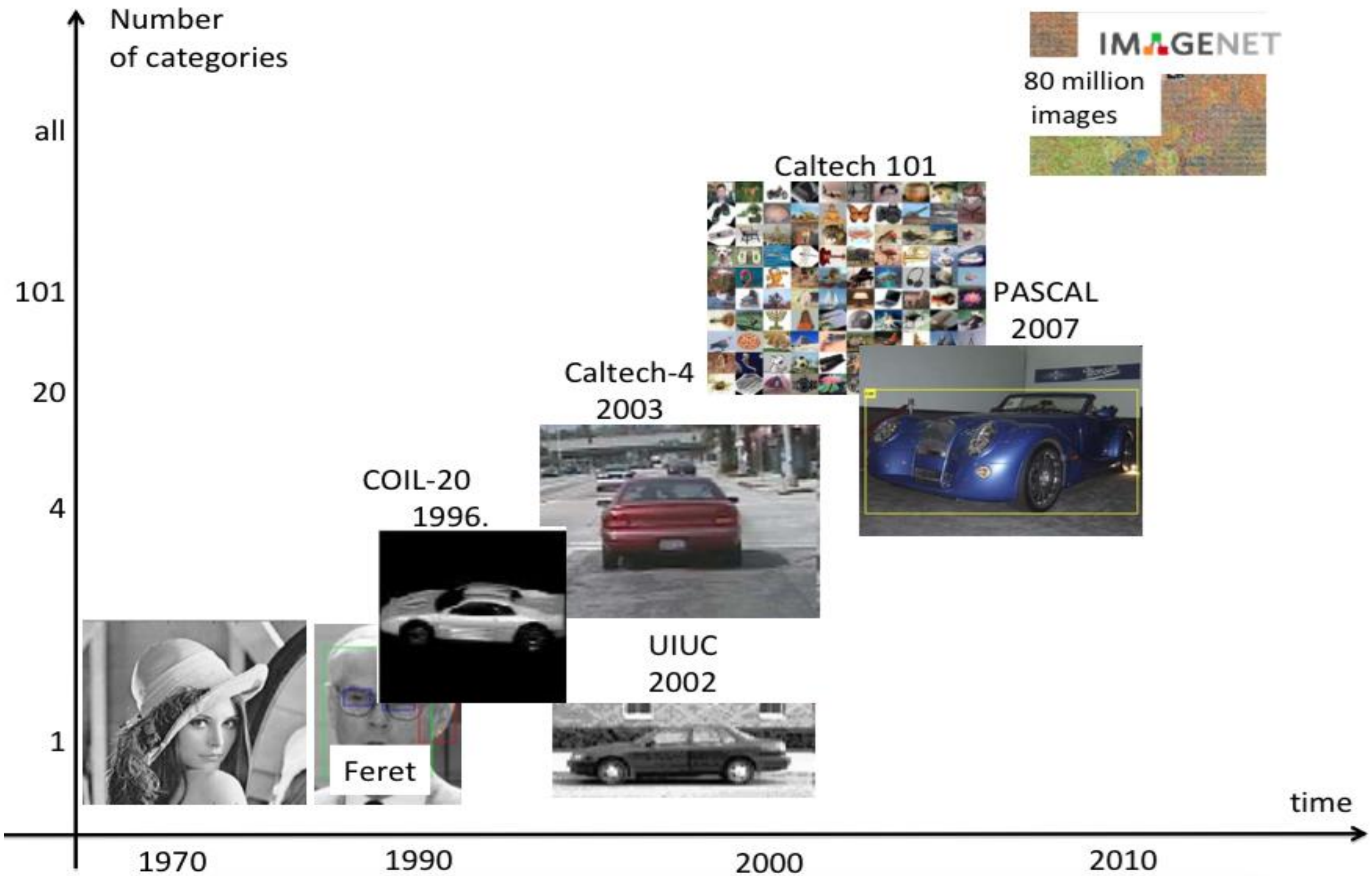


Scene Understanding



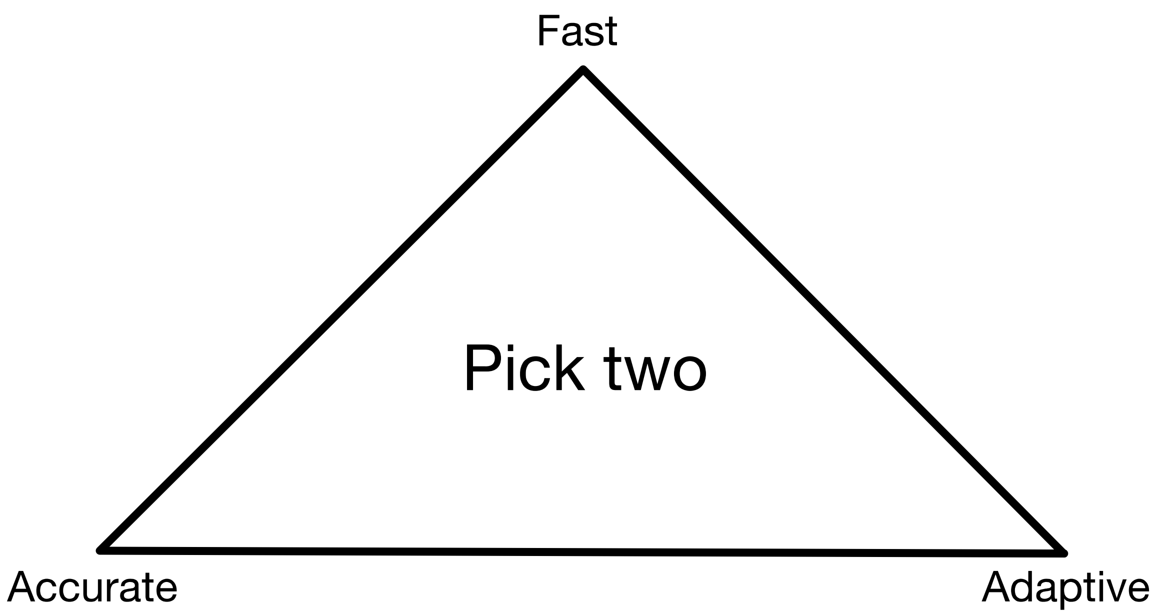
Action Recognition

- The phenomenon of data explosion in image/video databases is clearly perceivable.



The Promise and Perils of Benchmark Datasets and Challenges. David Forsyth, Alyosha Efros, Fei-Fei Li, Antonio Torralba and Andrew Zisserman. Frontiers in Computer Vision Workshop, CVPR 2011.

- This data explosion phenomenon calls for **new developments** in multi-class learning systems.
- The holy grail of classification is a classifier system that is (regardless of the scale of the problem) [1]:
  - **Accurate**
  - **Fast**
  - **Adaptive to data distribution**



[1] Polikar, R. (2006). Ensemble based systems in decision making. Circuits and Systems Magazine, IEEE, 6(3), 21-45.

PROBLEM

- Classification in the traditional **supervised setting**:
  - Learning + Prediction

$$\{\mathbf{x}_i, y_i\} : i \in \{1, \ldots, n\}$$
$$\mathbf{x}_i \in \mathbb{R}^f$$
$$y_i \in \{1, \ldots, k\}$$

  - With $k \leq 2$ we have binary (or one-class) classification:
    - SVMs w/o kernels (object recognition).
    - Adaboost w/o cascading (face recognition).
    - Random Forests (pose estimation).
    - Nearest Neighbors (semantic hashing).
    - Neural Networks (back to end CNNs).
  - While some algorithms naturally extend to $k \geq 2$ (i.e. RF or NN) others cannot be directly applied in the multi-class case.

PROBLEM

- Standard multi-class extensions of binary classifiers share the same spirit:
  - Train one classifier per class.
  - Choosing the classifier with highest score as the prediction.
- What if classifier misses its prediction?
- Should classifiers be trained on groups of classes?
  - **Pros:**
    - Balancing data on classifiers.
    - Leverage the loss in performance of noisy categories.
    - Recover from errors in classifiers!
  - **Cons**:
    - How to obtain the final prediction?

Error-Correcting Output Codes (ECOC)

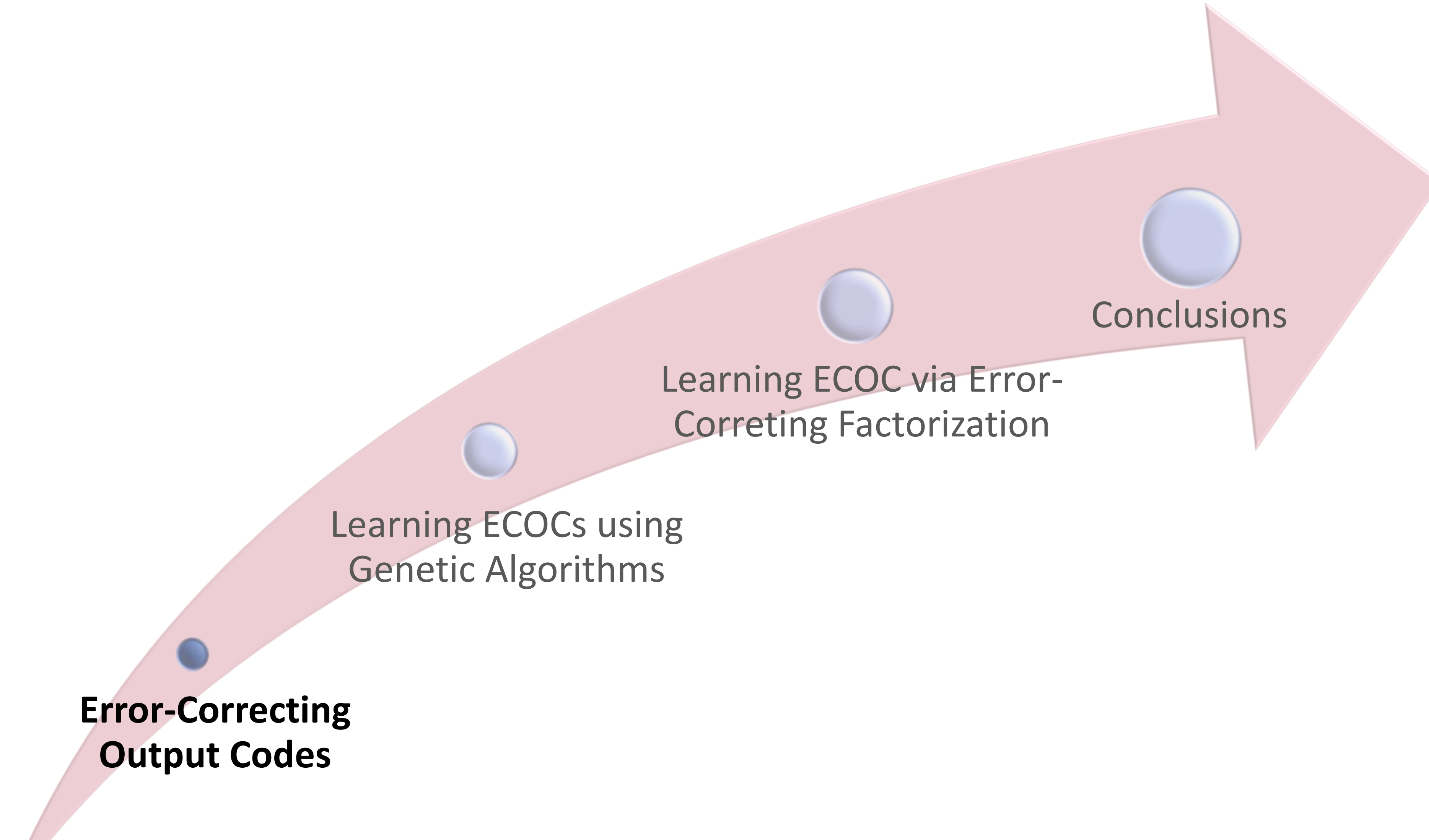
OBJECTIVES

- Develop multi-class classifiers that are accurate, fast and adaptive, within the framework of the Error-Correcting Output Codes:
  - **Accurate**: by using powerful binary classifiers.
  - **Fast**: minimizing the number of classifiers used.
  - **Adaptive**: exploiting multi-class data distribution.

- Deepen into open questions which call for further **study of Error-Correction capabilities** of ECOCs.

- Evaluate our approaches in several Multi-class classification tasks:
  - Localization sites of proteins, Japanese vowel sounds, written letters, etc.
  - Face Recognition, Traffic sign recognition, symbol recognition, etc.

1. **Error-Correcting Output Codes.**
    1. ECOC Introduction.
    2. ECOC Coding.
    3. ECOC Decoding.
    4. ECOC Properties.
2. **Learning ECOCs using Genetic Algorithms.**
    1. Minimal Error-Correcting Output codes.
    2. On the design of an ECOC-compliant Genetic Algorithm.
3. **Learning ECOCs via Error-Correcting Factorization.**
    1. Error-Correcting Capabilities.
    2. Error-Correcting Factorization.
4. **Conclusions.**



Error-Correcting Output Codes (ECOC)

- The **ECOC framework** is a powerful tool to tackle multi-class classification problems.
  - Based on **Error-Correcting principles** of Communications Theory [1].
  - Generalizes standard multi-class decompositions [2].
  - Reduces **both bias and variance** errors [3].

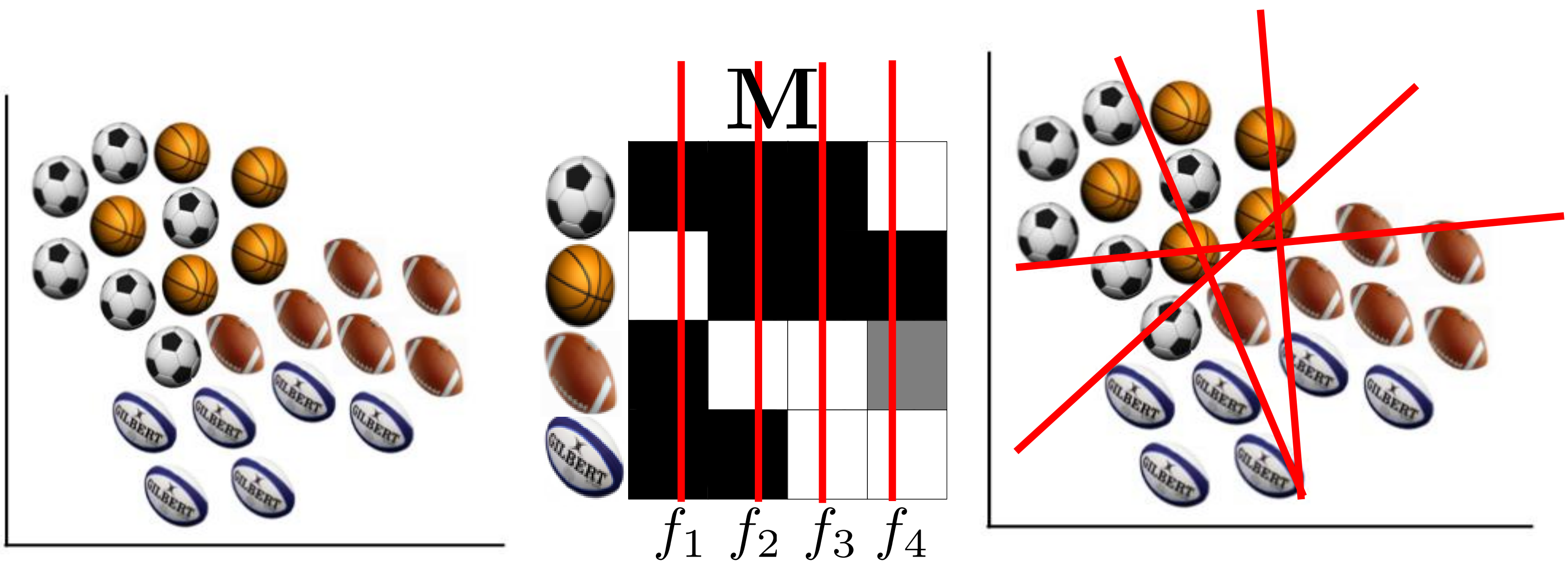
- This framework is composed of two different steps:
  - **Coding**: Decompose a given multiclass problem into a set of binary problems.
  - **Decoding:** Given a test sample, use a decoding measure to determine the prediction.

[1] Dietterich, T. G., & Bakiri, G. (1995). Solving multiclass learning problems via error-correcting output codes. Journal of artificial intelligence research, 263-286.
[2] Rifkin, R., & Klautau, A. (2004). In defense of one-vs-all classification. The Journal of Machine Learning Research, 5, 101-141.
[3] Kong, E. B., & Dietterich, T. G. (1995, July). Error-Correcting Output Coding Corrects Bias and Variance. In ICML (pp. 313-321).

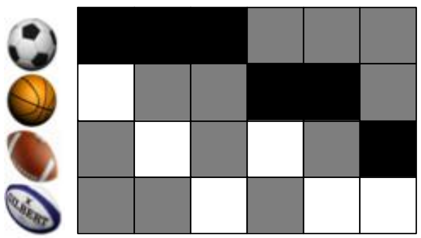
- At the **coding step** a decomposition of the $k-$class problem into l binary problems is computed:

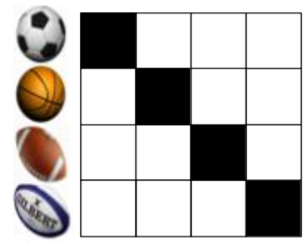
$$\mathbf{M}^{k \times l} \in \{-1, 0, +1\}$$

- **Rows** of the matrix represent the **codewords** of the classes.
- **Columns** of the matrix represent the **binary problems** to be learnt.

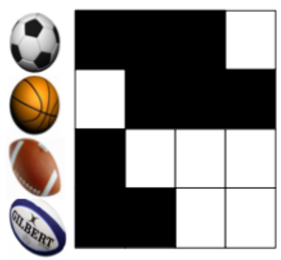


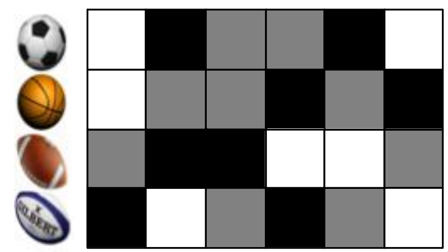
- One vs. One and One vs. All.





- Random (Dense and Sparse).

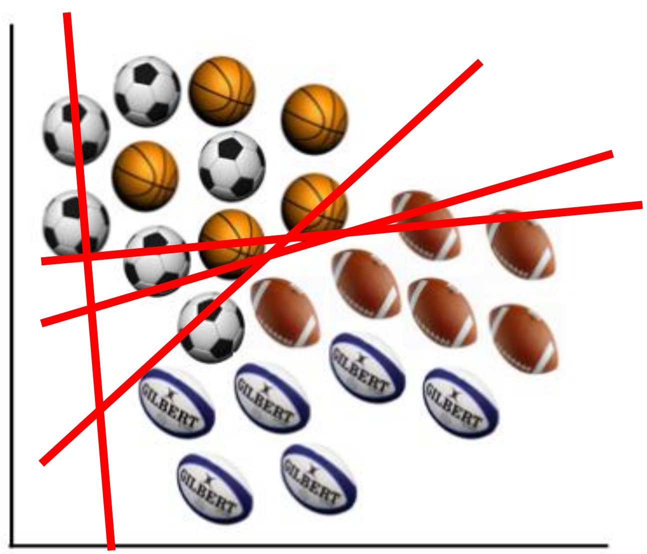


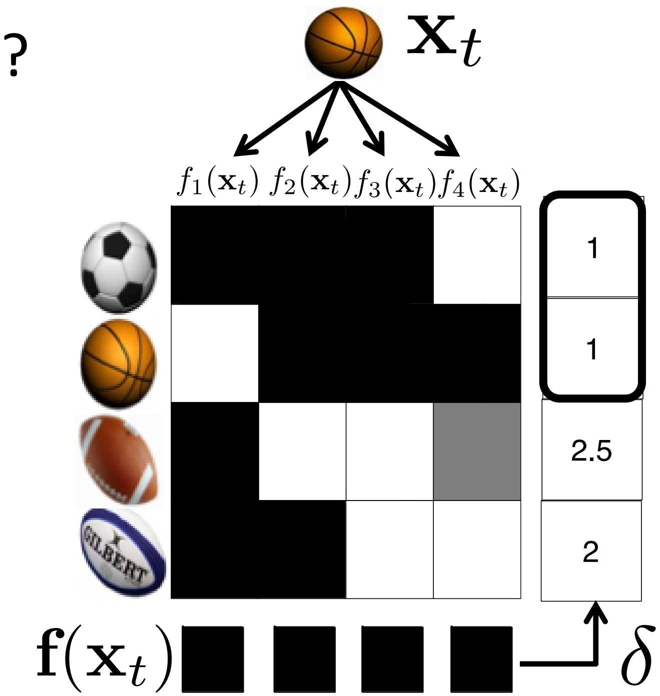


- Problem dependent (Discriminative ECOC [1], Spectral ECOC [2], Genetic Algorithms [3]).

[1] Pujol, O., Radeva, P., & Vitria, J. (2006). Discriminant ecoc: A heuristic method for application dependent design of error correcting output codes. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 28(6), 1007-1012.
[2] Zhang, X., Liang, L., & Shum, H. Y. (2009, September). Spectral error correcting output codes for efficient multiclass recognition. In Computer Vision, 2009 IEEE 12th International Conference on (pp. 1111-1118). IEEE.

- How to perform testing (decoding)?



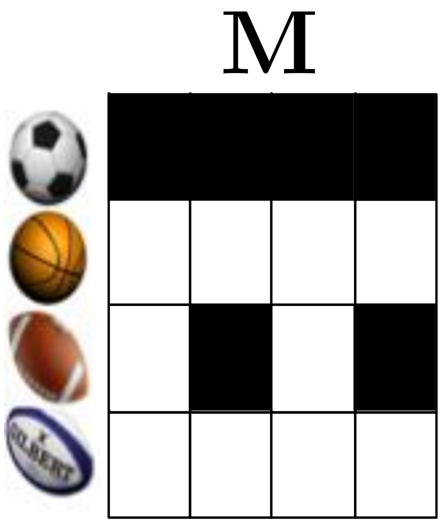


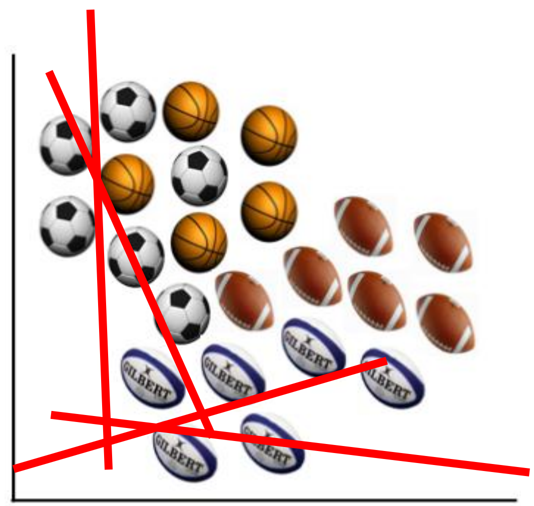
$$\arg\min_i \delta(\mathbf{m}^i, \mathbf{f}(\mathbf{x}_t))$$

- Choices for δ are unlimited [1]:
  - Hamming, Euclidean, Loss-based, probabilistic, etc.
- Even though a classifier missed its prediction, we were able to recover the correct prediction.
  - **Error-Correction!**

[1] Escalera, S., Pujol, O., & Radeva, P. (2010). On the decoding process in ternary error-correcting output codes. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 32(1), 120-134.

- There are some good practices to be followed when building the ECOC coding matrix [1].
  - M must **univocally define all the classes** in the problem (i.e. all the rows of the ECOC matrix must be different).
  - The **binary problems should be uncorrelated** in order to take profit from Error-Correcting principles.
  - **Powerful** (well tuned) **binary classifiers** should be used in order to obtain good classification accuracy.





  - M should **maximize the minimum distance between rows** to profit from Error-Correcting capabilities.

- Formalize the constraints of an ECOC coding matrix [1]:

1. The distance between any pair of rows should be greater or equal than 1.

$$\min(\delta_{AHD}(\mathbf{m}^i, \mathbf{m}^j)) \geq 1 \; \forall_{i,j} \in \{1, \ldots, k\}, i \neq j$$

2. The distance between any pair of columns should be greater or equal than 1.

$$\min(\delta_{HD}(\mathbf{m}_i, \mathbf{m}_j)) \geq 1 \; \forall_{i,j} \in \{1, \ldots, l\}, i \neq j$$

1. A column and its negation are equivalent.

$$\min(\delta_{HD}(\mathbf{m}_i, -\mathbf{m}_j)) \geq 1 \; \forall_{i,j} \in \{1, \ldots, l\}, i \neq j$$

[1] Bautista, M. Á., Escalera, S., Baró, X., & Pujol, O. (2014). On the design of an ECOC-Compliant Genetic Algorithm. Pattern Recognition, 47(2), 865-884.

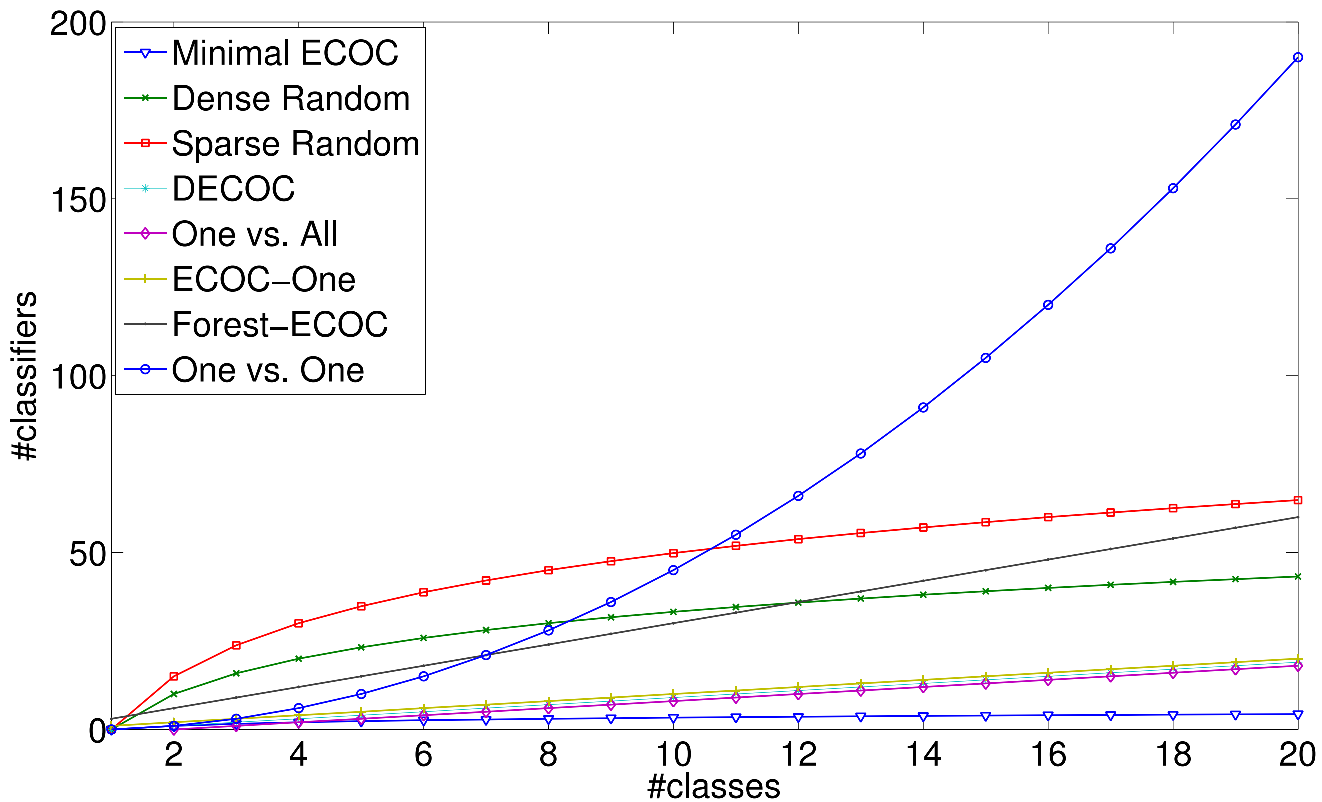
- **Error-Correction** is **extremely important:**
  - Number of binary classifiers that can miss without affecting the final prediction [1].

$$\frac{\min(\delta(\mathbf{m}^i, \mathbf{m}^j)) - 1}{2}, \forall_{i,j} \in \{1, \ldots, k\}$$

  - Measured as a single scalar [1].
  - Increasing it means that the number of classifiers (columns of ECOC matrix) will increase, and thus training complexity will increase.

[1] Dietterich, T. G., & Bakiri, G. (1995). Solving multiclass learning problems via error-correcting output codes. Journal of artificial intelligence research, 263-286.

- ECOC codings define a **super-linear** number of classifiers.
  - In this dissertation we are interested in **sub-linear** designs.
  - Given the reduce number of classifiers used by sub-linear designs, they should **exploit the multi-class data distribution**.



SUMMARY

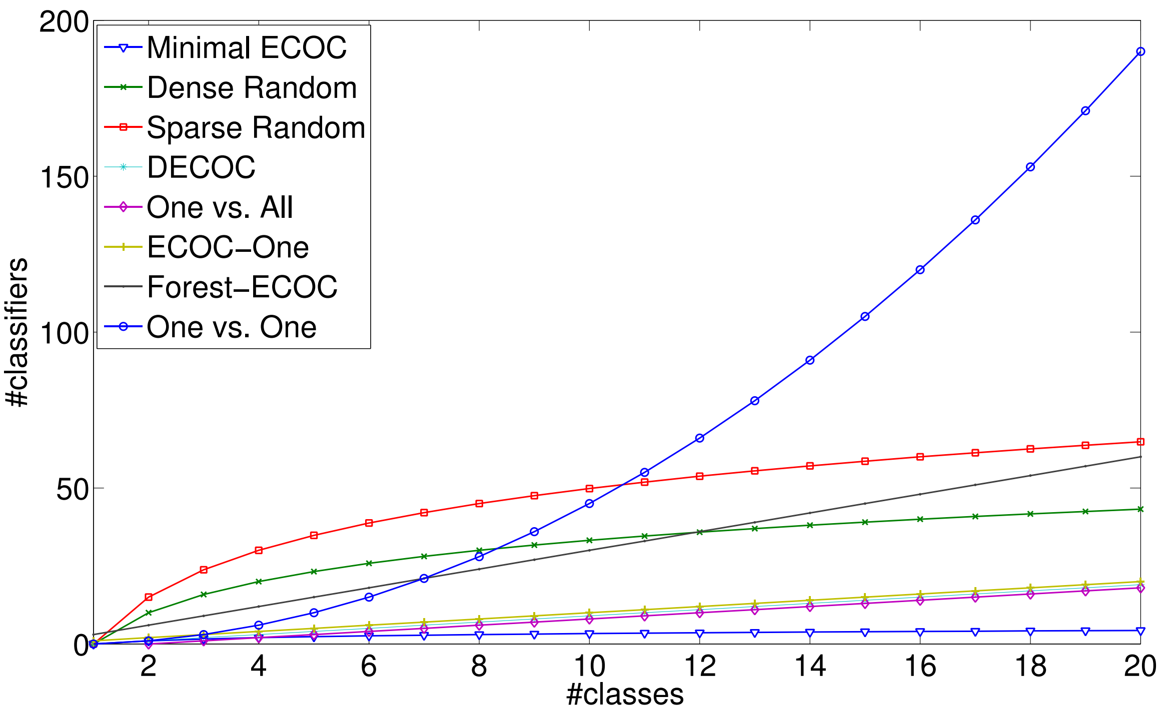
- ECOCs are a **powerful tool** to deal with multi-class problems**.**
- **Several coding designs** predefined, random and problem-dependent.
- Unlimited number of decoding designs.
- **Good practices of ECOC** are redefined as constraints.
- We define the ECOC of **Minimal length**.
- **Sub-linear ECOCs** should be optimized to compensate for the reduced number of classifiers.

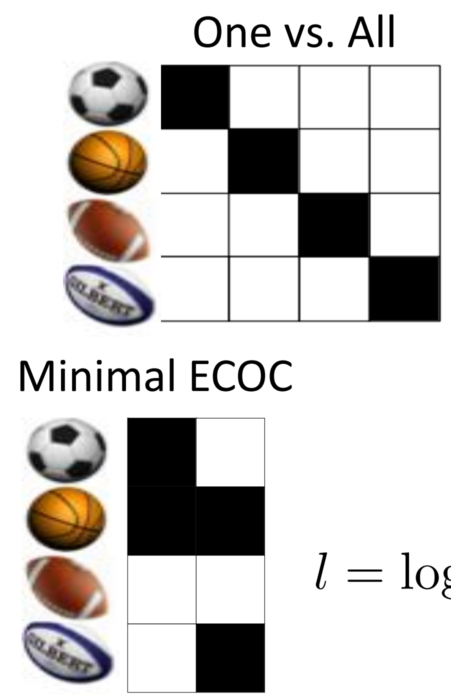
- Error-Correcting Output Codes
- **Learning ECOCs using Genetic Algorithms**
- Learning ECOC via Error-Correting Factorization
- Conclusions

Learning ECOCs using Genetic Algorithms (ECOC-GA)

- Exploit **multi-class data distribution** to find **sub-linear ECOC** designs with high performance.

- What is the **shortest code length** that can be defined [1,2]?





$$l = \log_2(k)$$

[1] Bautista, M. A., Baró, X., Pujol, O., Radeva, P., Vitrià, J., & Escalera, S. (2010). Compact evolutive design of error-correcting output codes. In Proceedings of the Supervised and Unsupervised Methods and their Applications (SUEMA), European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases.
[2] Bautista, M. Á., Escalera, S., Baró, X., Radeva, P., Vitriá, J., & Pujol, O. (2012). Minimal design of error-correcting output codes. Pattern Recognition Letters, 33(6), 693-702.

- Optimize the ECOC matrix given the data distribution and the classifiers is a **NP-complete problem** [1].
  - The ECOC search space is **extremely large**.

$$\binom{k}{2^l} \qquad \frac{2^l!}{2(2^l - k)!}$$

  - The search space is **not continuous** (ECOC coding matrices are discrete).
  - The search space is **not differentiable**.

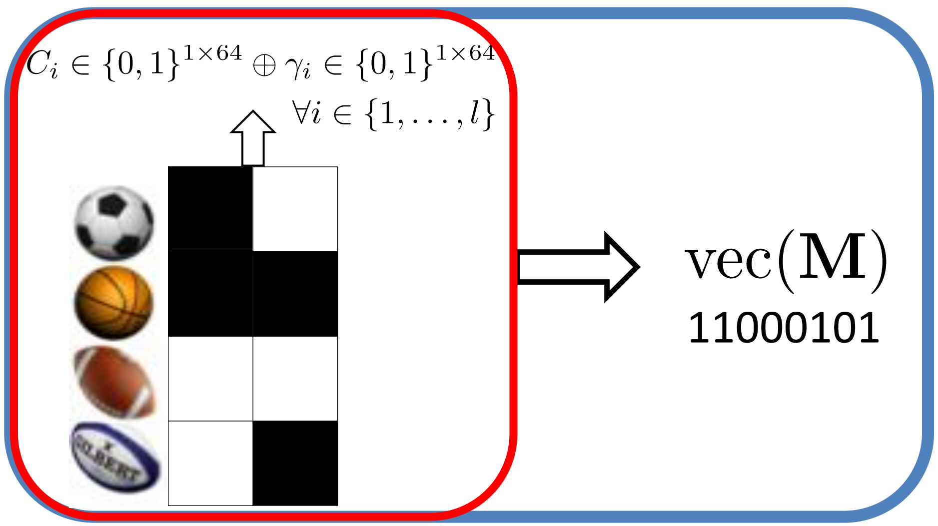
- **Genetic Algorithms** (GA) are often applied in this setting with **benefitial results**.

[1]Utschick, W., & Weichselberger, W. (2001). Stochastic organization of output codes in multiclass learning problems. Neural Computation, 13(5), 1065-1102. Chicago

- Genetic Algorithms are stochastic optimization methods based on **Darwin's Evolution** Theory [1].
- The fitness of indivudals is improved over generations by using **crossover and mutation operators**.
- **Our proposal [2,3]**:
  - **Optimize the ECOC** matrix using **Genetic Algorithms:**
    - Define how good an individual is (fitness function): classification accuracy.
    - Define how to represent a solution as a binary string (encoding): already binary.
  - **Optimize** the parameters of **binary classifiers** (SVM-RBF) using **Genetic Algorithms:**
    - Define how good a solution is: classification accuracy.
    - Define how to represent a solution as a binary string: binary representation of C, γ .
  - Use **standard crossover and mutation operators**

[1] Baluja, S., & Caruana, R. (1995, May). Removing the genetics from the standard genetic algorithm. In Machine Learning: Proceedings of the Twelfth International Conference (pp. 38-46).
[2] Bautista, M. A., Baró, X., Pujol, O., Radeva, P., Vitrià, J., & Escalera, S. (2010). Compact evolutive design of error-correcting output codes. In Proceedings of the, European Conference on Machine Learning Workshops.
[3] Bautista, M. Á., Escalera, S., Baró, X., Radeva, P., Vitriá, J., & Pujol, O. (2012). Minimal design of error-correcting output codes. Pattern Recognition Letters, 33(6), 693-702.

- An **ECOC individual** is represented as a **binary vector** and evaluated by means of its classification error.
- **Iterative** 2-step procedure:



1) Optimize the SVMs parameters.

2) Optimize the coding matrix and return to step 1.

- **Standard genetic operators** are used: scattered crossover and gaussian mutation:
  - Scattered crossover: randomly selects a set of points for each parent.
  - Gaussian mutation: adds a random number taken from a Gaussian distribution with mean 0 to each entry of the parent vector.

- We compare our proposal with Binary Minimal ECOC, PBIL [1] Minimal ECOC, One vs. All, One vs. One, Discriminant ECOC [2] and Forest-ECOC [2] approaches.
- Experimental settings:
  - We generated $10 \times k$ individuals per problem in the first generation.
  - We used SVMs with RBF kernel as our binary classifier.
    - Parameters were tuned using either GAs or PBIL for all methods, using two-fold cross-validation.
  - We used the Hamming Decoding distance.
  - We report the average classification accuracy over a stratified 10 fold-cross validation.
- A cache of dichotomizers is stored to leverage training time.

[1] Baluja, S. (1994). Population-based incremental learning. a method for integrating genetic search based function optimization and competitive learning (No. CMU-CS-94-163). Carnegie-Mellon Univ Pittsburgh Pa Dept Of Computer Science.

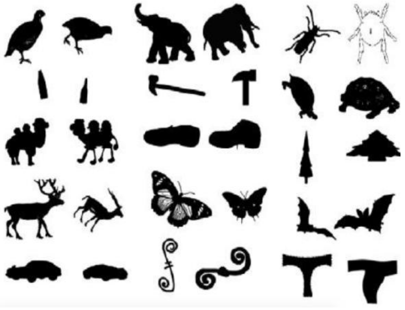
[2] Pujol, O., Radeva, P., & Vitria, J. (2006). Discriminant ecoc: A heuristic method for application dependent design of error correcting output codes. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 28(6), 1007-1012.

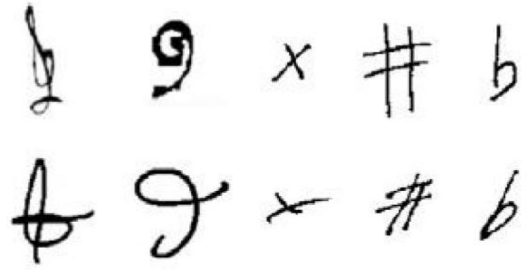
[3] Baró, X., Escalera, S., Vitrià, J., Pujol, O., & Radeva, P. (2009). Traffic sign recognition using evolutionary adaboost detection and forest-ECOC classification. Intelligent Transportation Systems, IEEE Transactions on, 10(1), 113-126.

- We evaluated experiments on 12 UCI datasets.

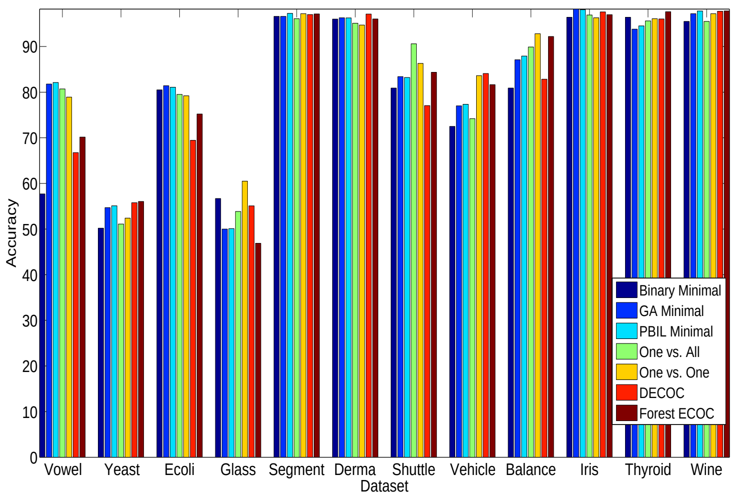
Problem	#Training samples	#Features	#Classes
Dermathology	366	34	6
Iris	150	4	3
Ecoli	336	8	8
Vehicle	846	18	4
Wine	178	13	3
Segmentation	2310	19	7
Glass	214	9	7
Thyroid	215	5	3
Vowel	990	10	11
Balance	625	4	3
Shuttle	14500	9	7
Satimage	4435	36	6
Yeast	1484	8	10

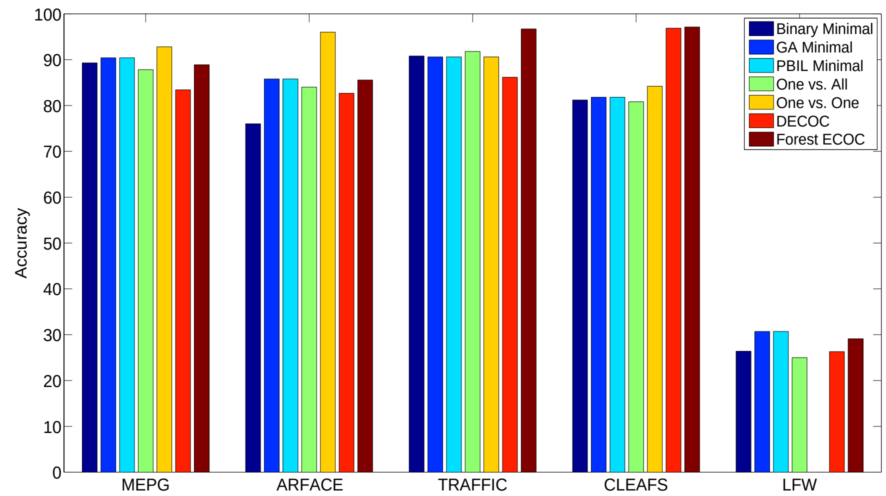
- We tackled 5 Computer Vision problems:
  - Labeled Faces in the wild: 610 categories
  - MPEG visual objects: 70 categories
  - Traffic sign categorization: 36 classes
  - ARFace dataset: 20 classes
  - Old music scores: 7 classes

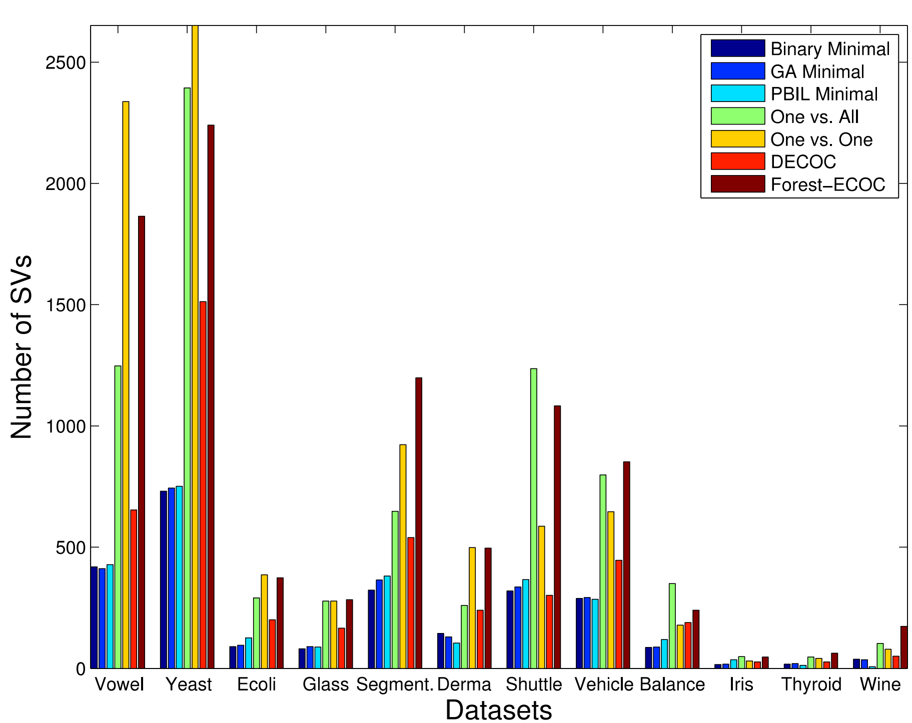
- Classification accuracies for each method.

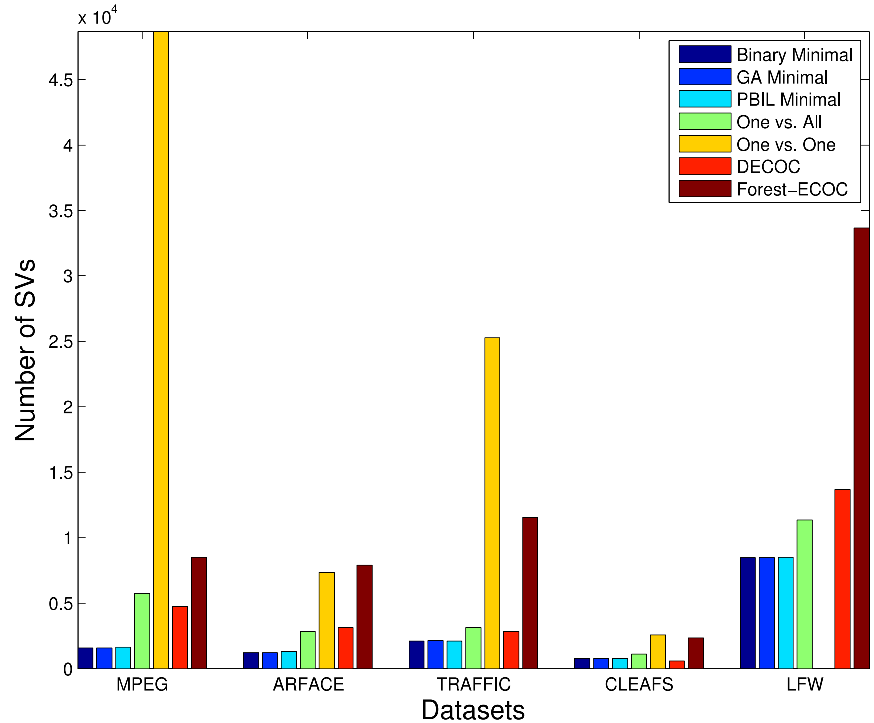




Binary Minimal	GA Minimal	PBIL Minimal	OVA	OVO	DECOC	Forest ECOC
5,2	3,6	3	4,8	3,7	4,2	3,1

- Number of Support Vectors for each method.





x2.9 over OVA

x3.2 over OVO

x1.8 over OVA

x7.4 over OVO

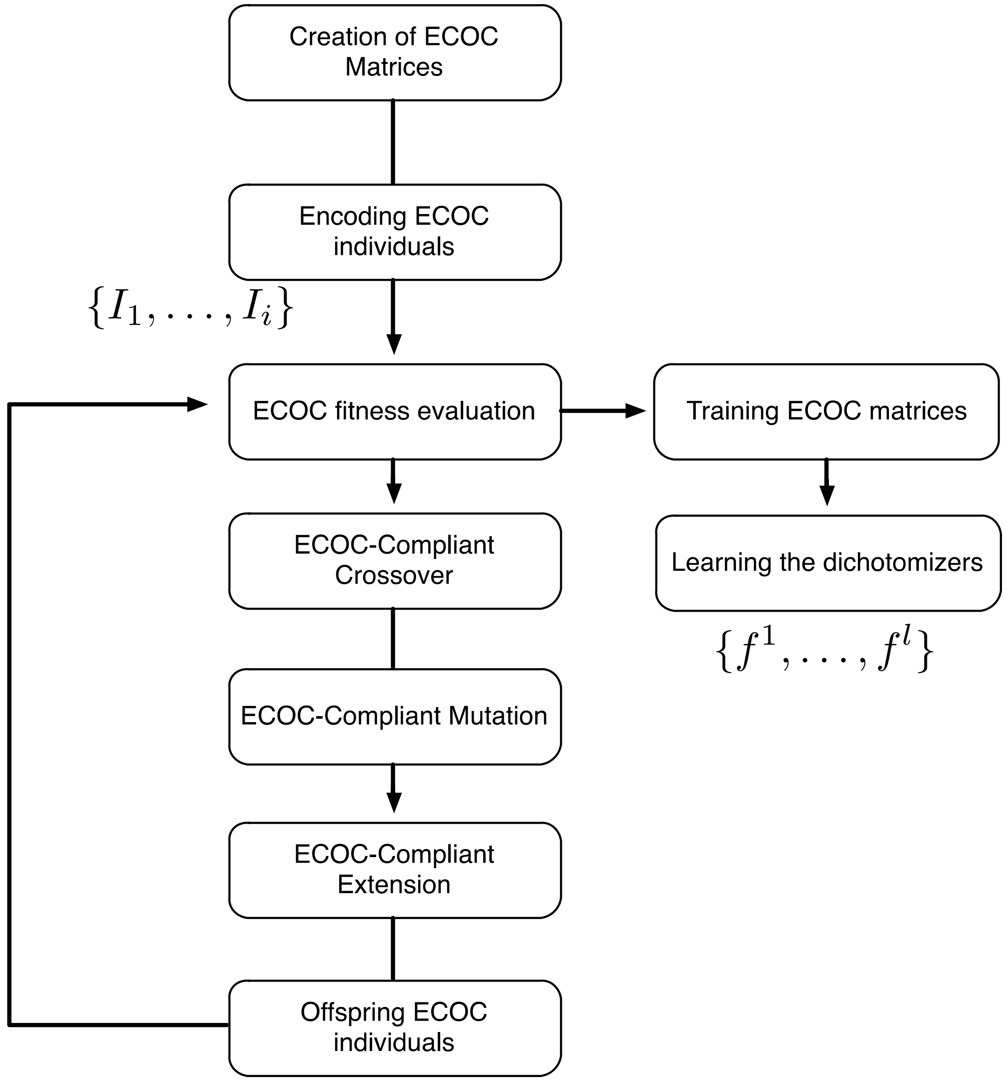
SUMMARY

- **Minimal ECOCs** are suitable for multi-class classification, when the **coding matrix is optimized** using Genetic Algorithms.
- With the **minimal number of classifiers** we obtain comparable or even better classification accuracies than state-of-the-art works.
- To obtain a high performance **we optimize the parameters of the binary classifiers** using **GAs**.
- **Large-scale tasks** can be tackled with **Minimal ECOCs**.

- Standard operators overlook **individuals structure**.
  - Generation of vast number of non-valid individuals makes the **algorithm ineffective**.

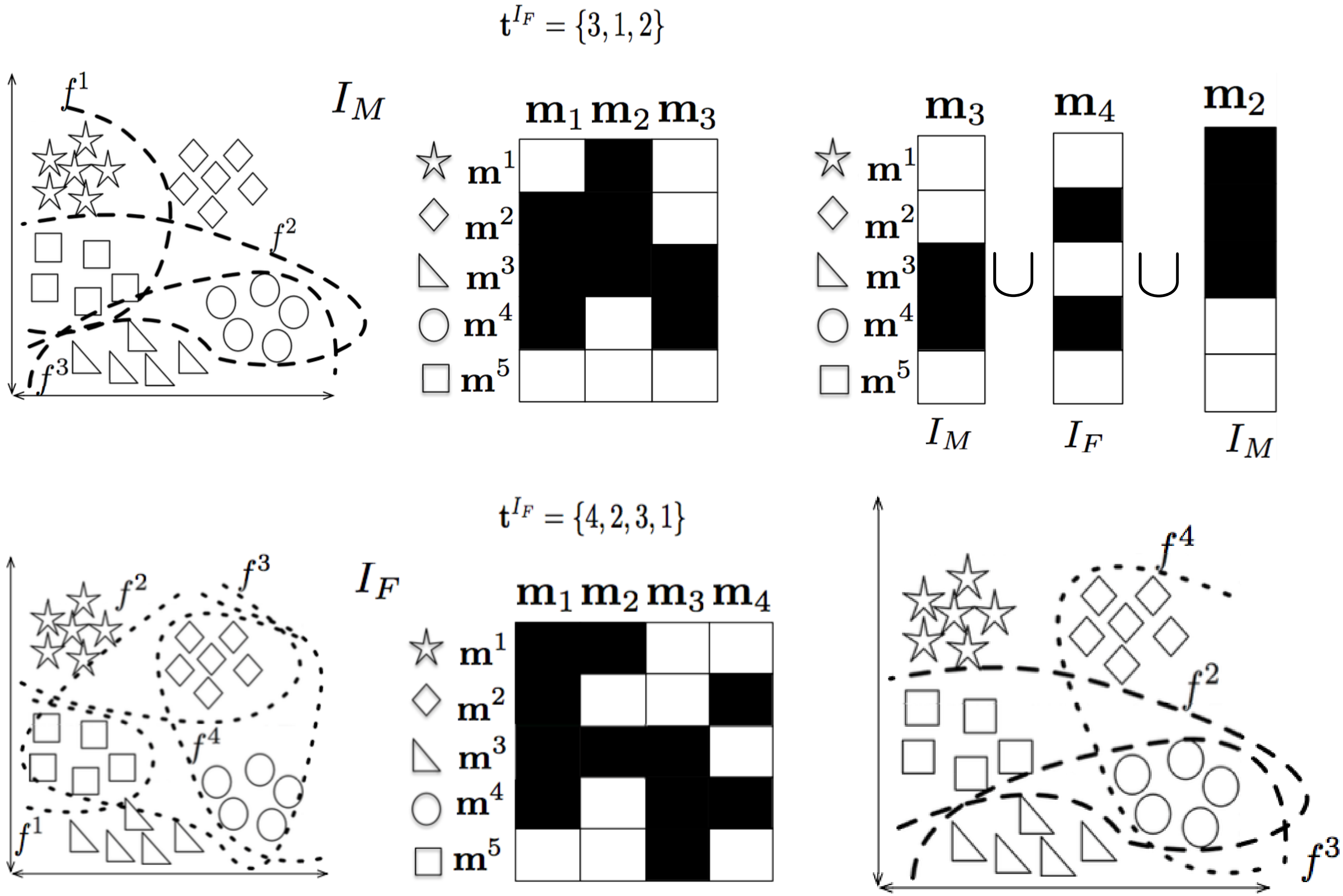
- **New proposal: ECOC-Compliant Genetic Algorithm**
  - Redefine **crossover and mutation** operators in order to take into account **ECOC properties**.
  - Possibility of including new operators?
    - Controlling the number of classifiers.
    - Adding and removing classifiers when needed.
  - Operators should be **fast and simple**.

- General pipeline for the **ECOC-Compliant Genetic Algorithm** [1].

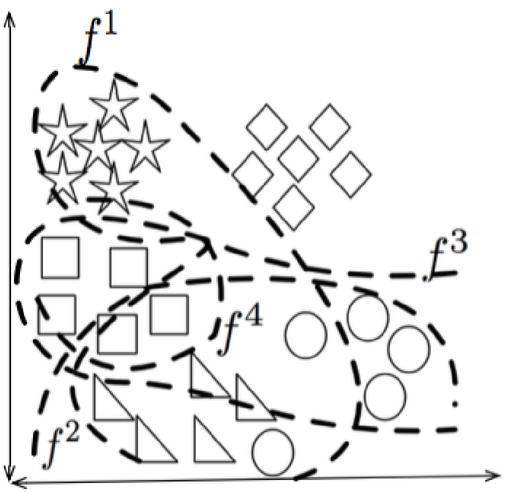


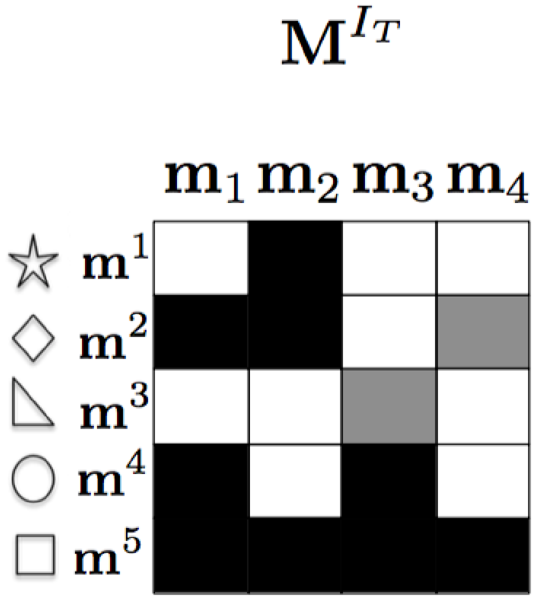
[1] Bautista, M. Á., Escalera, S., Baró, X., & Pujol, O. (2014). On the design of an ECOC-Compliant Genetic Algorithm. Pattern Recognition, 47(2), 865-884.

- ECOC-Compliant **crossover** operator.



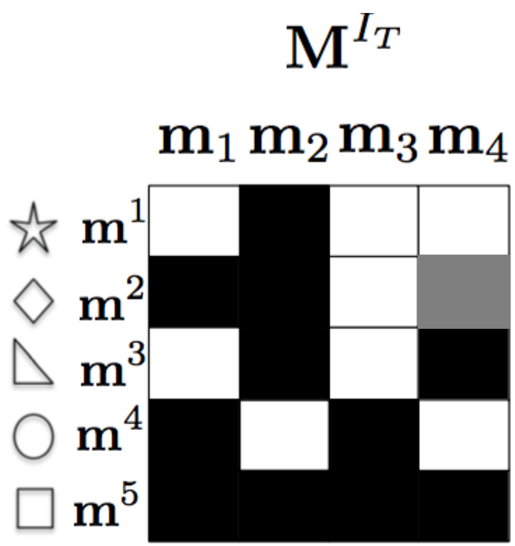
- ECOC-Compliant **mutation** operator.

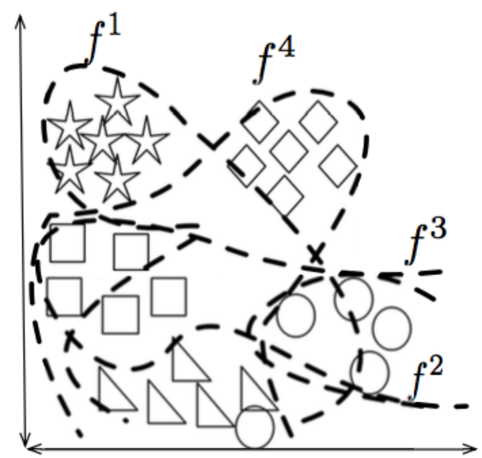




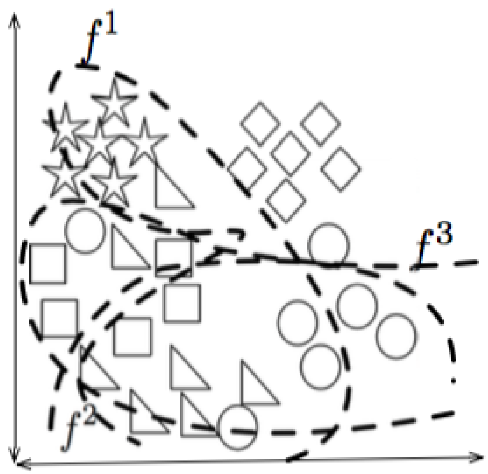
$\mathbf{C}^{I_T}$

	$\mathbf{c}^1$	$\mathbf{c}^2$	$\mathbf{c}^3$	$\mathbf{c}^4$	$\mathbf{c}^5$
$\mathbf{c}^1$	6	0	0	0	0
$\mathbf{c}^2$	0	6	0	0	0
$\mathbf{c}^3$	0	0	3	1	0
$\mathbf{c}^4$	0	0	2	5	2
$\mathbf{c}^5$	0	0	0	0	3

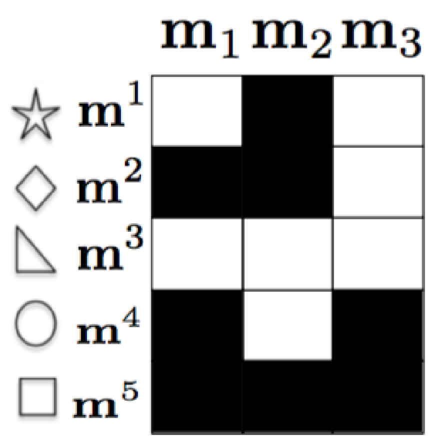




- ECOC-Compliant **extension** operator.



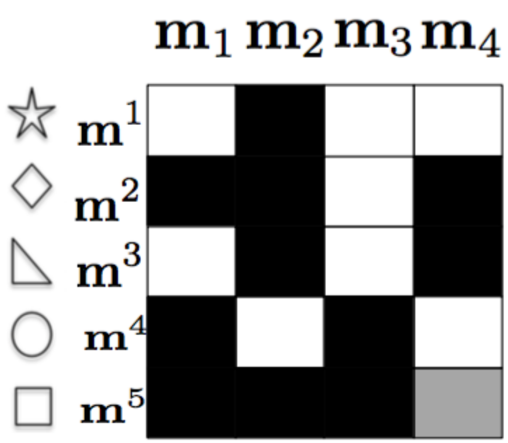
$\mathbf{M}^{I_T}$

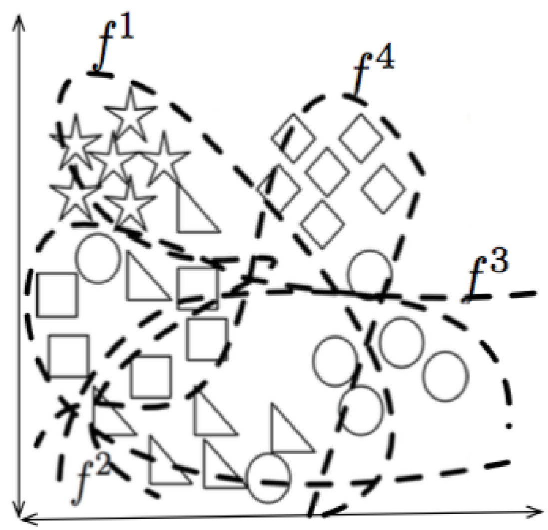


$\mathbf{C}^{I_T}$

	$\mathbf{c}^1$	$\mathbf{c}^2$	$\mathbf{c}^3$	$\mathbf{c}^4$	$\mathbf{c}^5$
$\mathbf{c}^1$	6	0	1	0	0
$\mathbf{c}^2$	0	6	0	1	0
$\mathbf{c}^3$	0	0	3	1	0
$\mathbf{c}^4$	0	0	2	5	2
$\mathbf{c}^5$	0	0	1	1	3

$\mathbf{M}^{I_T}$

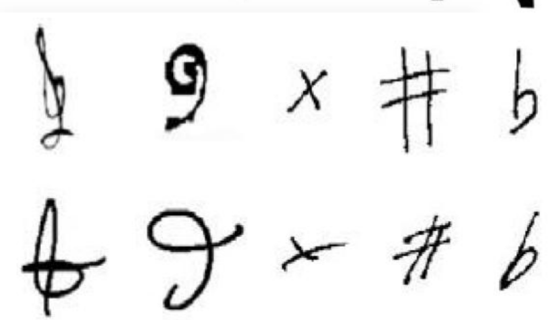




- We run experiments on 9 UCI datasets.

Problem	#Training samples	#Features	#Classes
Vowel	990	10	11
Yeast	1484	8	10
Ecoli	336	8	8
Glass	214	9	7
Segmentation	2310	19	7
Dermathology	366	34	6
Shuttle	14500	9	7
Vehicle	846	18	3
Satimage	4435	36	6

- We tackled 4 Computer Vision problems:
  - MPEG visual objects: 70 categories
  - Traffic sign categorization: 36 classes
  - ARFace dataset: 20 classes
  - Old music scores: 7 classes



- We compare our proposal with Binary Minimal ECOC, Lorena et. al [1], Pedrajas et. al [2], One vs. All, One vs. One, Discriminant ECOC and Forest-ECOC approaches.

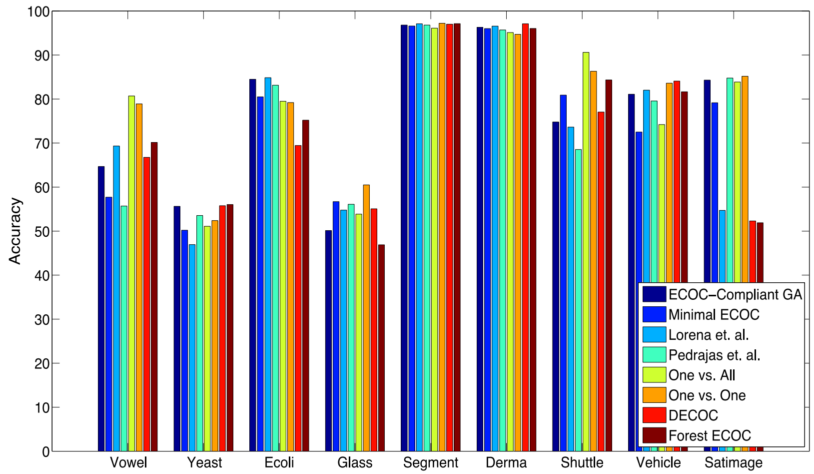
- Experimental settings:
  - We generated $10 \times k$ individuals per problem.
  - We used SVMs with RBF kernel as our binary classifier.
    - Parameters were tuned using Gas for all methods, using two-fold cross-validation.
  - We used the Loss-Weighted decoding.
  - We report the average classification accuracy over a stratified 10 fold-cross validation.

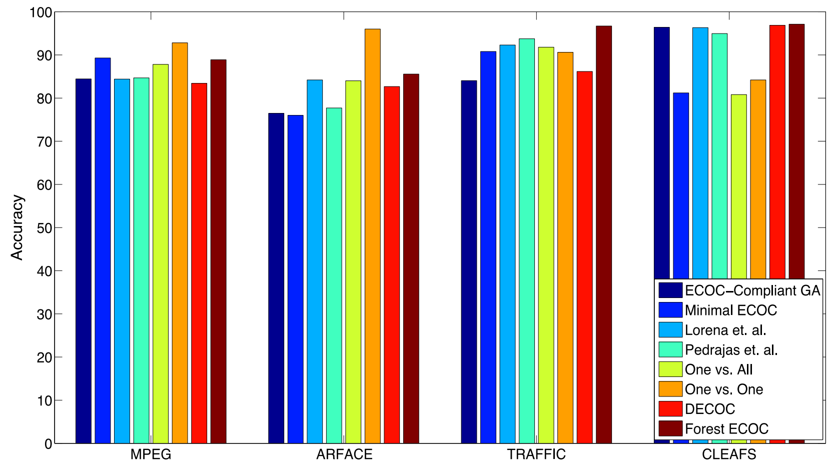
- A cache of dichotomizers is stored to leverage training time.

[1] Lorena, A. C., & de Carvalho, A. C. (2006, October). Multiclass SVM design and parameter selection with genetic algorithms. In Neural Networks, 2006. SBRN'06. Ninth Brazilian Symposium on (pp. 131-136). IEEE. Chicago Lorena, Ana Carolina, and André CPLF d

[2] Garcia-Pedrajas, N., & Fyfe, C. (2008). Evolving output codes for multiclass problems. Evolutionary Computation, IEEE Transactions on, 12(1), 93-106.

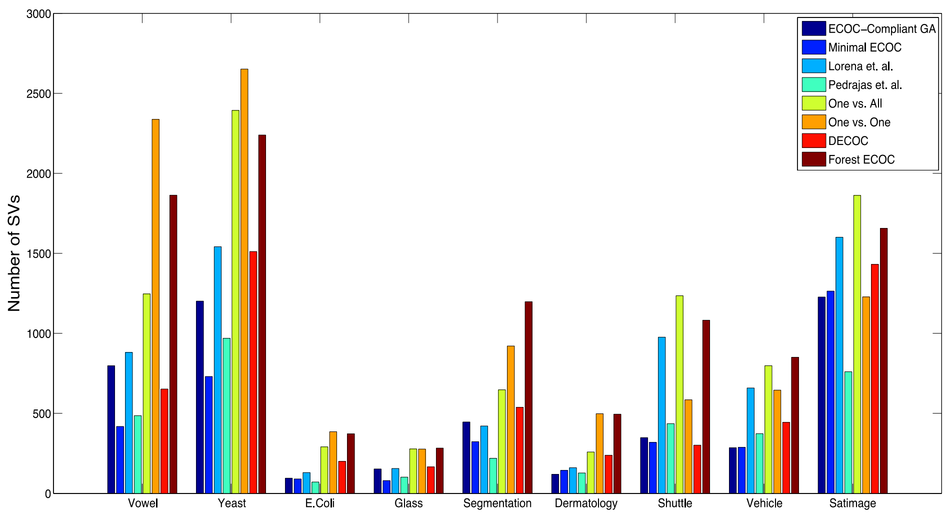
- Classification accuracies for each method on UCI datasets.

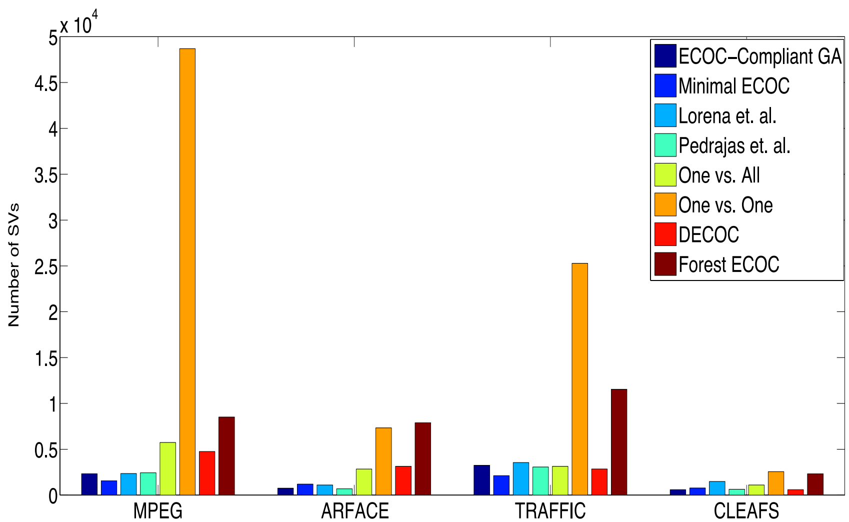




ECOC-Compliant GA	Minimal ECOC	Lorena et. al	Pedrajas et. al	OVA	OVO	DECOC	Forest ECOC
4.5	6.2	4.4	4,8	4.7	3.2	4.3	3.4

- Number of support vectors for each method on UCI datasets.





x1.9 over OVA

x2.0 over OVO

x1.8 over OVA

x12.0 over OVO

SUMMARY

- We propose to **redefine the operators** to take into account the **properties of Error-Correcting Output codes.**
- The novel genetic operators **avoid the generation of non-valid individuals**.
- To obtain a high performance we optimize the parameters of the binary classifiers SVM-RBF using GA.
- Results show that we obtain comparable or even better results than state-of-the-art ECOC design while **reducing drastically the number of Support Vectors.**

CONCLUSIONS

- **Genetic Algorithms** can be powerful tools to **optimize ECOC coding matrices**.
- Avoid the generation of non-valid individuals.
- The training computational cost can be leveraged using simple speed up tricks.
- The **Error-Correcting properties** of an ECOC **cannot be exploited by the proposed GAs**.

Error-Correcting Output Codes

Learning ECOCs using Genetic Algorithms

Learning ECOC via Error-Correting Factorization

Conclusions

Learning ECOCs via Error-Correcting Factorization (ECOC-ECF)

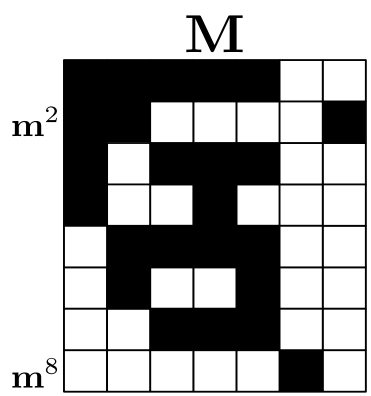
- In this dissertation we aim to deepen into open questions which call for further study of **Error-Correction capabilities of ECOCs**:
  1. How do **Minimal ECOC matrices** behave? ☑
  2. Is Error-Correction distributed evenly on all classes? ☒
  3. Can problem-dependent designs profit from the distribution of Error-Correcting capabilities? ☒
  4. Is it better to allocate Error-Correction to classes prone to error or to classes not prone to error? ☒
  5. Is there a problem-dependent definition of the minimum number of classifiers needed for an ECOC matrix? ☒

- **Error-Correction capability** of an ECOC **is key**:
  - It measures the **number of binary classifiers that can miss** its prediction without affecting the final multi-class prediction.

$$\frac{\min(\delta(\mathbf{m}^i,\mathbf{m}^j))-1}{2}$$

  - The correcting capability is measured as a **single scalar** for the whole ECOC matrix in state-of-the-art works.
  - A single **scalar does not provide information** about how the Error-Correction is distributed.
  - We can compute the **pair-wise distance matrix between codewords,** aka the Separability Matrix.

$$\mathbf{H} \in \mathbb{R}^{k\times k}$$



$$\lfloor \frac{\min(\mathbf{H})-1}{2} \rfloor = 0$$

H

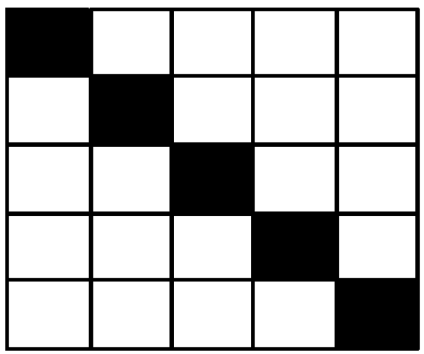
	0	4	1	3	1	3	2	6
$\mathbf{h}^2$	4	**0**	5	3	5	3	6	4
	1	5	**0**	2	2	4	1	5
	3	3	2	**0**	4	4	3	3
	1	5	2	4	**0**	2	1	5
	3	3	4	4	2	**0**	3	3
	2	6	1	3	1	3	**0**	4
$\mathbf{h}^8$	6	4	5	3	5	3	4	**0**

$$\lfloor \frac{\min(\mathbf{h}^2 \oplus \mathbf{h}^8)-1}{2} \rfloor = 1$$

- Analyzing the **pair-wise distance** enables a **deeper understanding** of how the **correction** is distributed among classes (how separated codewords are).

- Every ECOC matrix has its pair-wise distance matrix.

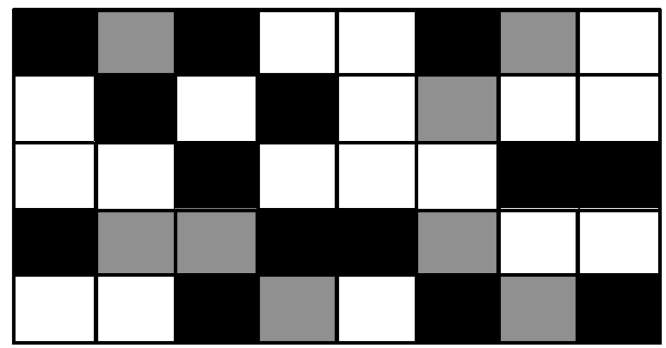
M



H

0	2	2	2	2
2	0	2	2	2
2	2	0	2	2
2	2	2	0	2
2	2	2	2	0

M

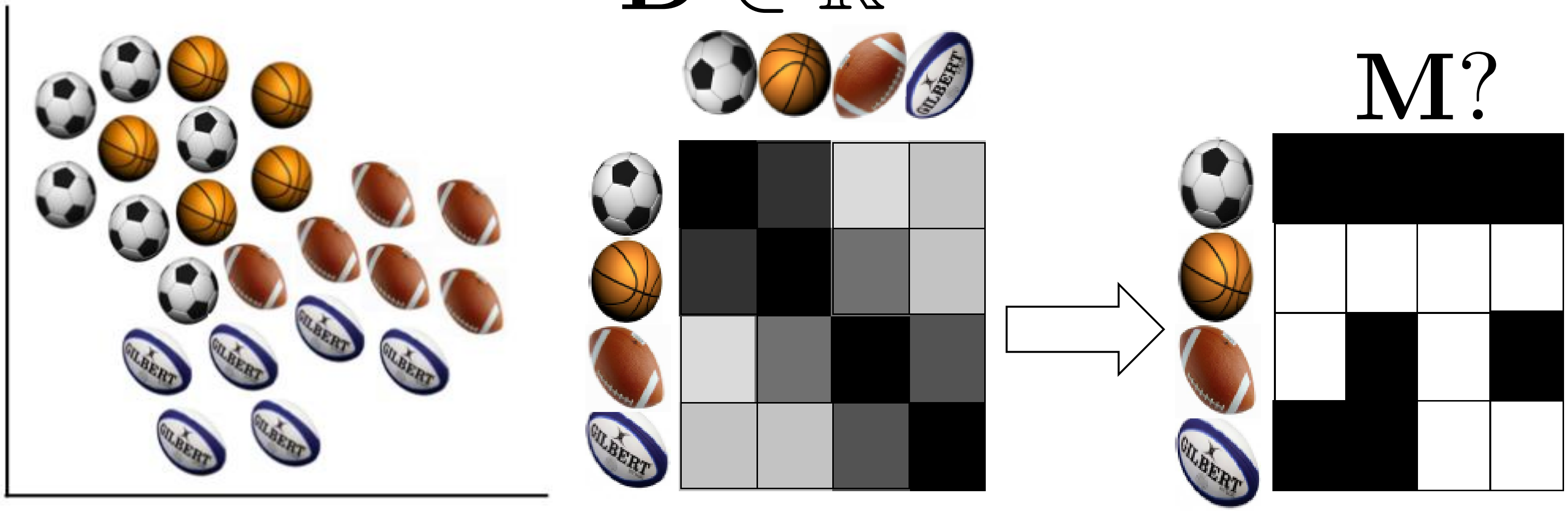


H

0	4	1	3	2
4	0	5	1	4
1	5	0	4	1
3	1	4	0	1
2	4	1	1	0

- Distance (Design) matrix → ECOC?
  - We define $\mathbf{D}$ as a design matrix that encodes the distances between pairs of codewords of a desired ECOC.

$$\mathbf{D} \in \mathbb{R}^{k \times k}$$



  - Extract $\mathbf{D}$ from multi-class data is easy (e.g heuristiscs like Mahalanobis distances between classes.)
  - Information of experts can be easily coded.
  - Analyzing $\mathbf{D}$ can assist to solve the number of classifiers problem.

- **Motivation:**
  - Is it possible to find an ECOC such that it follows the distances denotes in a design matrix?
  - What conditions should hold for an ECOC to encode an arbitrary design matrix?

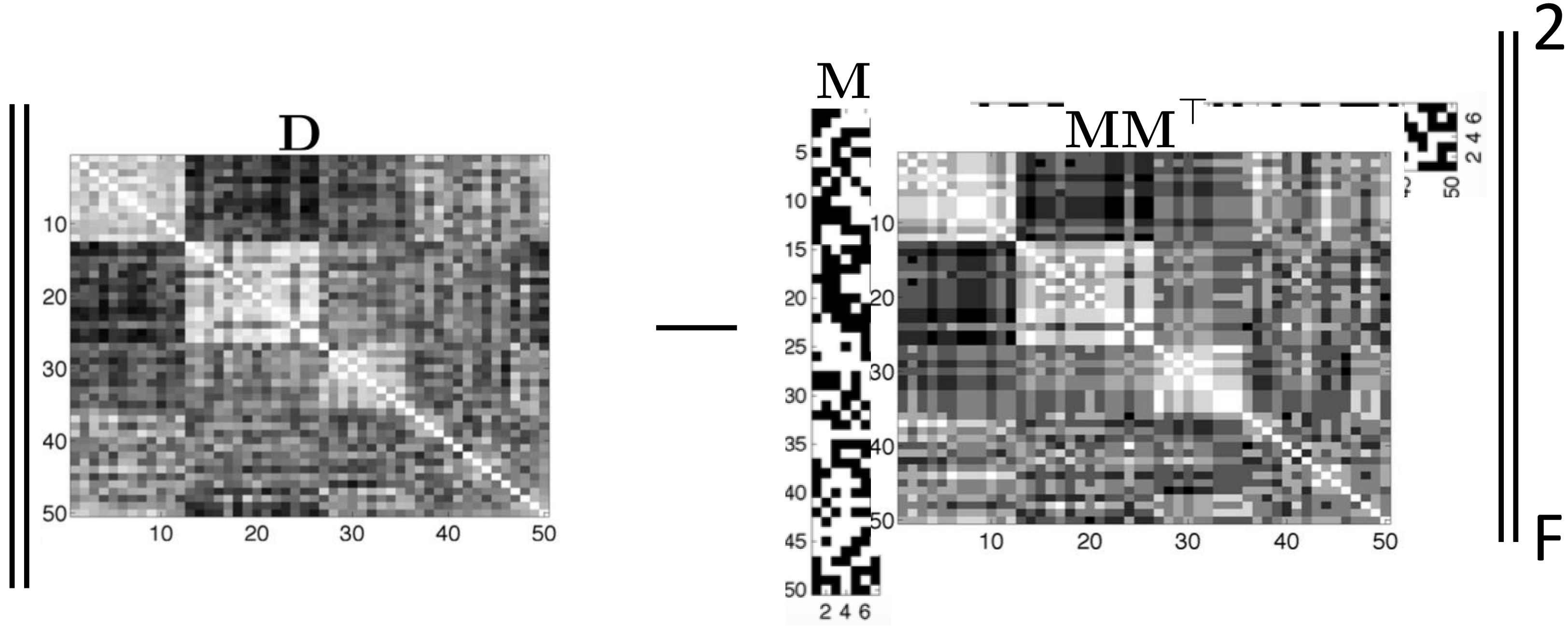
$$h_{ij} < h_{kl} \iff d_{ij} < d_{kl} \forall_{i,j,k,l}$$

  - However the l_1 distance can be seen as:

$$\|\mathbf{m}^i - \mathbf{m}^j\|_1 = \frac{-(\mathbf{m}^i {\mathbf{m}^j}^\top) + l}{2} \quad \mathbf{m} \in \{-1, +1\}$$

  - We can work with the **inner product** equivalently.
  - Factorize the design matrix!

- A visualization of the problem.



- What is the **number of classifiers** needed to **minimize the norm**?

$$\mathrm{rank}(\mathbf{D}) = \mathrm{rank}(\mathbf{M}\mathbf{M}^\top) = \mathrm{rank}(\mathbf{M})$$

- Formulating the **Error-Correcting Factorization** (ECF) [1]:

$$
\begin{aligned}
\underset{\mathbf{M}}{\text{minimize}} \quad & \|\mathbf{D} - \mathbf{M}\mathbf{M}^\top\|_F^2 \\
\text{subject to} \quad & \mathbf{M} \in \{-1,+1\}^{k \times l} \\
& \mathbf{M}\mathbf{M}^\top - \mathbf{P} \leq 0 \\
& \mathbf{M}^\top\mathbf{M} - \mathbf{1}(l-1) \leq 0 \\
& -\mathbf{M}^\top\mathbf{M} - \mathbf{1}(l-1) \leq 0
\end{aligned}
$$

- **Non-convex!**
  - Quadratic term makes the objective function non-convex.
  - Discrete constraint makes the problem NP-Complete.
- **Good news:**
  - Discrete constraint can be relaxed $\mathbf{M} \in [-1,+1]$
  - Coordinate Descent has been successfully applied in non-convex problems (convergence to stationary points if problems are uniquely solved) [2].

[1] Bautista, M. A., Pujol, O., de la Torre, F., & Escalera, S. (2015). Error-Correcting Factorization. arXiv preprint arXiv:1502.07976. Under review at TPAMI
[2] Grippo, L., & Sciandrone, M. (2000). On the convergence of the block nonlinear Gauss–Seidel method under convex constraints. Operations Research Letters, 26(3), 127-136.

- **Codeword descent approach** for ECF[1]:
  - Optimize **the *i-th* codeword** of M while **fixing the rest** of the rows.

$$\begin{aligned} \underset{\mathbf{m}^i}{\text{minimize}} \quad & \left\| \begin{bmatrix} l & \mathbf{d}_i \\ \mathbf{d}^{iT} & \mathbf{D}'^i_{/i} \end{bmatrix} - \begin{bmatrix} \mathbf{m}^i \mathbf{m}^{iT} & \mathbf{M}'^i \mathbf{m}_i \\ \mathbf{M}'^i \mathbf{m}^{iT} & \mathbf{M}'^i \mathbf{M}'^{i\top} \end{bmatrix} \right\|_F^2 \\ \text{subject to} \quad & \mathbf{m}^i \in [-1, +1]^l \\ & \begin{bmatrix} \mathbf{m}^i \mathbf{m}^{iT} & \mathbf{M}'^i \mathbf{m}_i \\ \mathbf{M}'^i \mathbf{m}^{iT} & \mathbf{M}'^i \mathbf{M}'^{i\top} \end{bmatrix} - \begin{bmatrix} l & \mathbf{p}_i \\ \mathbf{p}^{iT} & \mathbf{P}'^i_{/i} \end{bmatrix} \leq 0. \end{aligned}$$

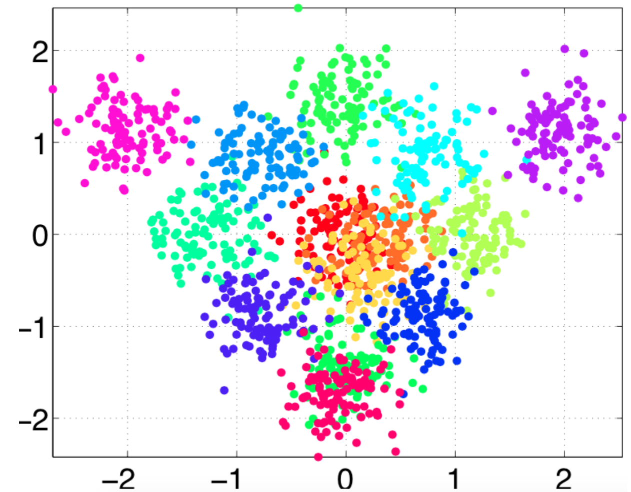
  - Terms not involve the *i-th* codeword can be drop, **reducing the ECF** problem to **least squares**:

$$\begin{aligned} \underset{\mathbf{m}^i}{\text{minimize}} \quad & \left\| \mathbf{M}'^i \mathbf{m}^i - \mathbf{d}^i \right\|_2^2 \\ \text{subject to} \quad & -1 \leq \mathbf{m}^i \leq +1 \\ & \mathbf{M}'^i \mathbf{m}^i - \mathbf{p}^i \leq 0 \end{aligned}$$

  - Least-squares can be **solved uniquely** (when not overdetermined), thus the algorithm is **guaranteed to converge** to **stationary point**.

[1] Bautista, M. A., Pujol, O., de la Torre, F., & Escalera, S. (2015). Error-Correcting Factorization. arXiv preprint arXiv:1502.07976. Under review at TPAMI

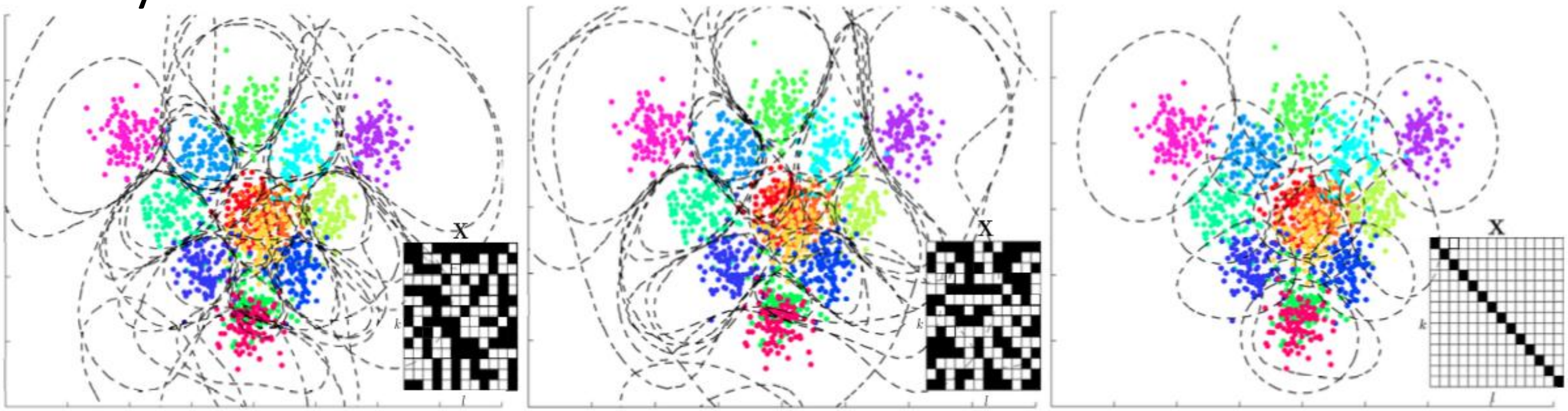
- We tested ECF on 8 UCI datasets and a Toy (synthetic) problem of 14 classes.



	Glass	Segment.	Ecoli	Yeast	Vowel	Toy	Traffic	ARFace
#s	214	2310	336	1484	990	400	3481	1300
#f	9	19	8	8	10	2	100	120
#c	7	7	8	10	11	14	36	50

- We also choose 2 Vision problems:
  - Traffic sign categorization: **36 classes**
  - ARFace dataset: **20 classes**

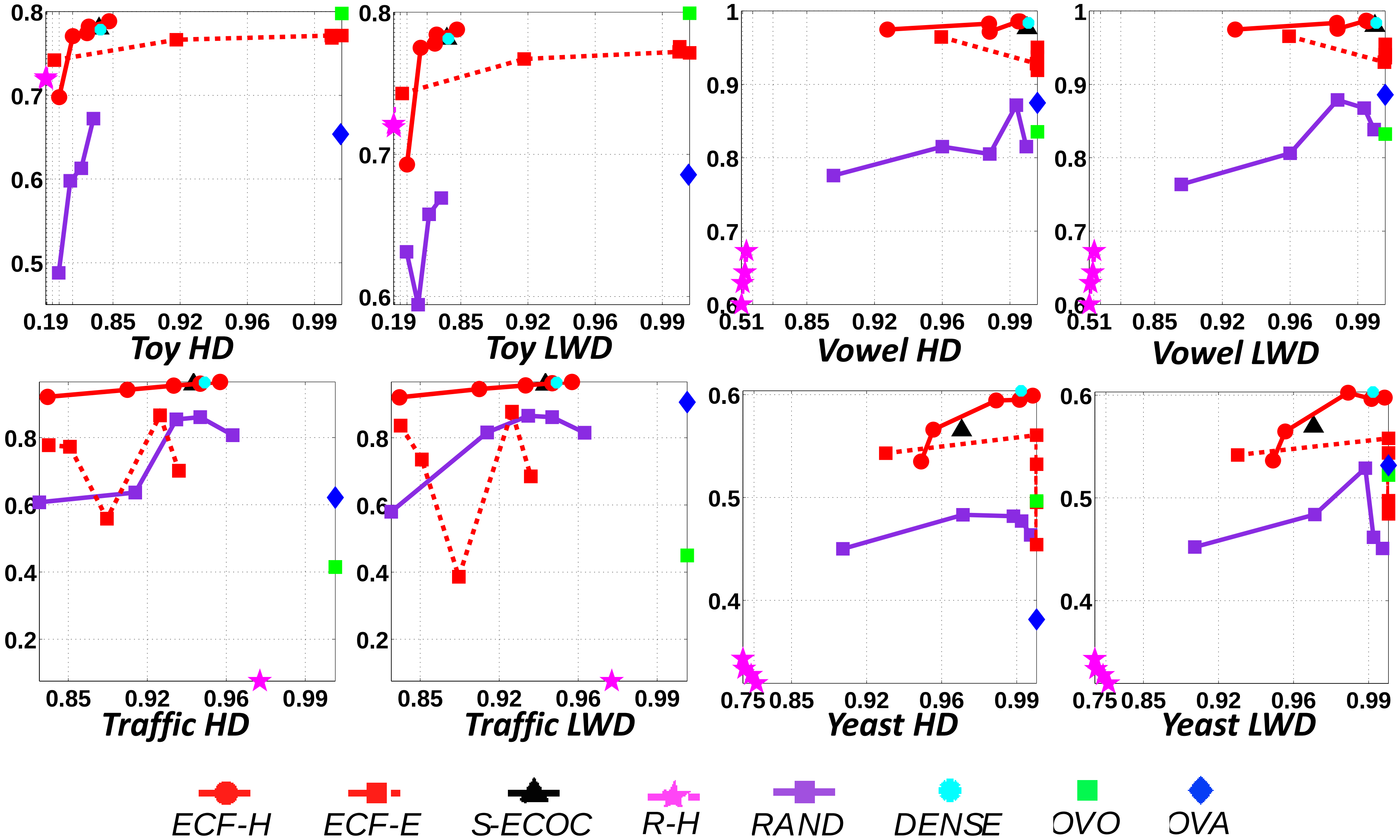
- Methods:
  - **ECF-H, where Error-Correction is allocated to classes prone to error.**
  - **ECF-E: Error-Correction is allocated to classes not prone to error.**
  - OVA, OVO, Random ECOC (RAND), Dense Random (DR), Spectral ECOC (S-ECOC)[1], Relaxed Hierarchy (R-H) [2].
- Binary classifier:
  - SVM-RBF with parameters optimized using grid-search.
  - Reported accuracy are the average over a 10-fold stratified cross-validation.
- Toy Problem



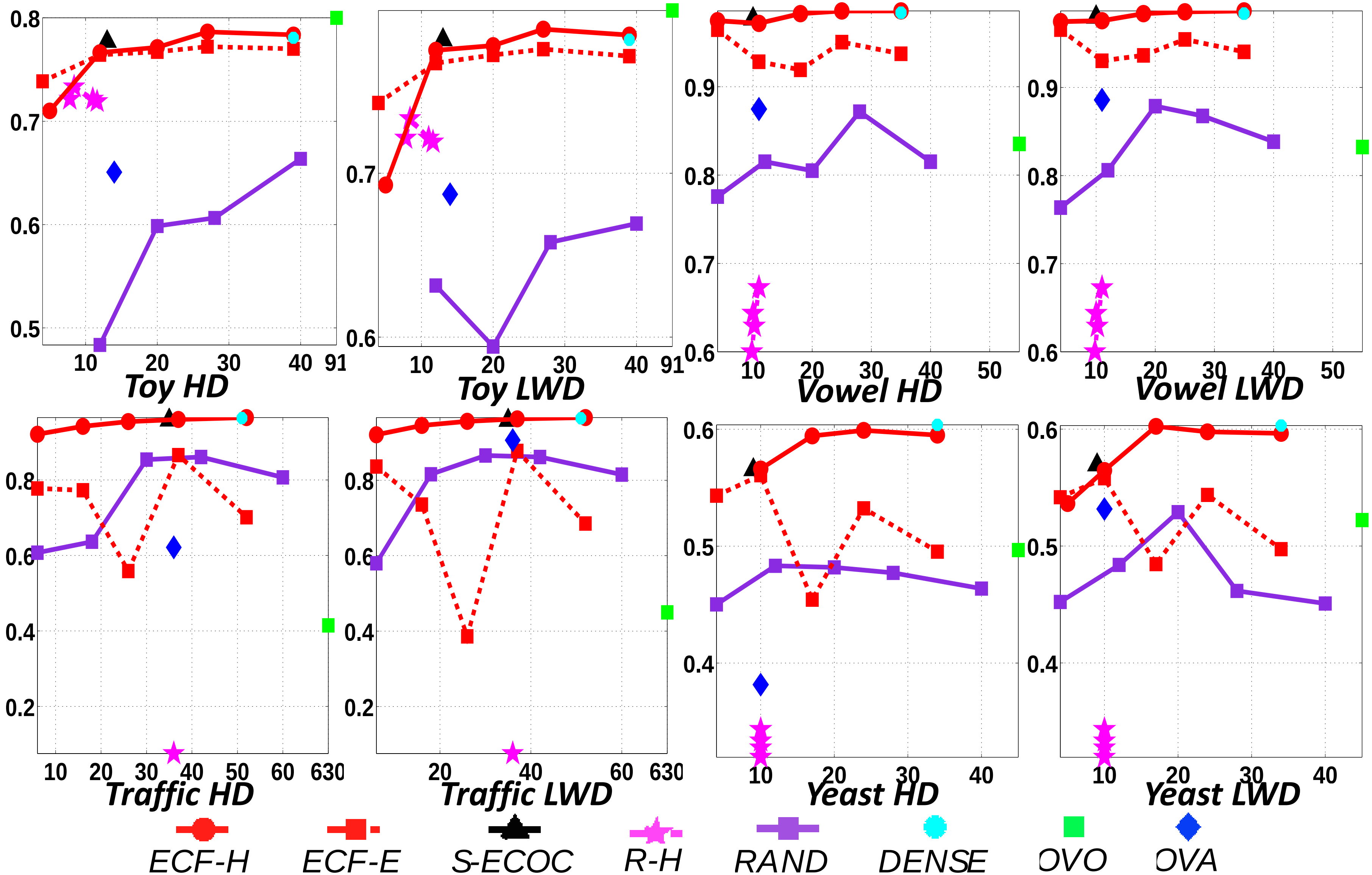
Error-Correcting Factorization 77.49% | Dense Random 66.45% | One versus All 49.53%

[1] Zhang, X., Liang, L., & Shum, H. Y. (2009, September). Spectral error correcting output codes for efficient multiclass recognition. In Computer Vision, 2009 IEEE 12th International Conference on (pp. 1111-1118). IEEE.
[2] Gao, T., & Koller, D. (2011, November). Discriminative learning of relaxed hierarchy for large-scale visual recognition. In Computer Vision (ICCV), 2011 IEEE International Conference on (pp. 2072-2079). IEEE.

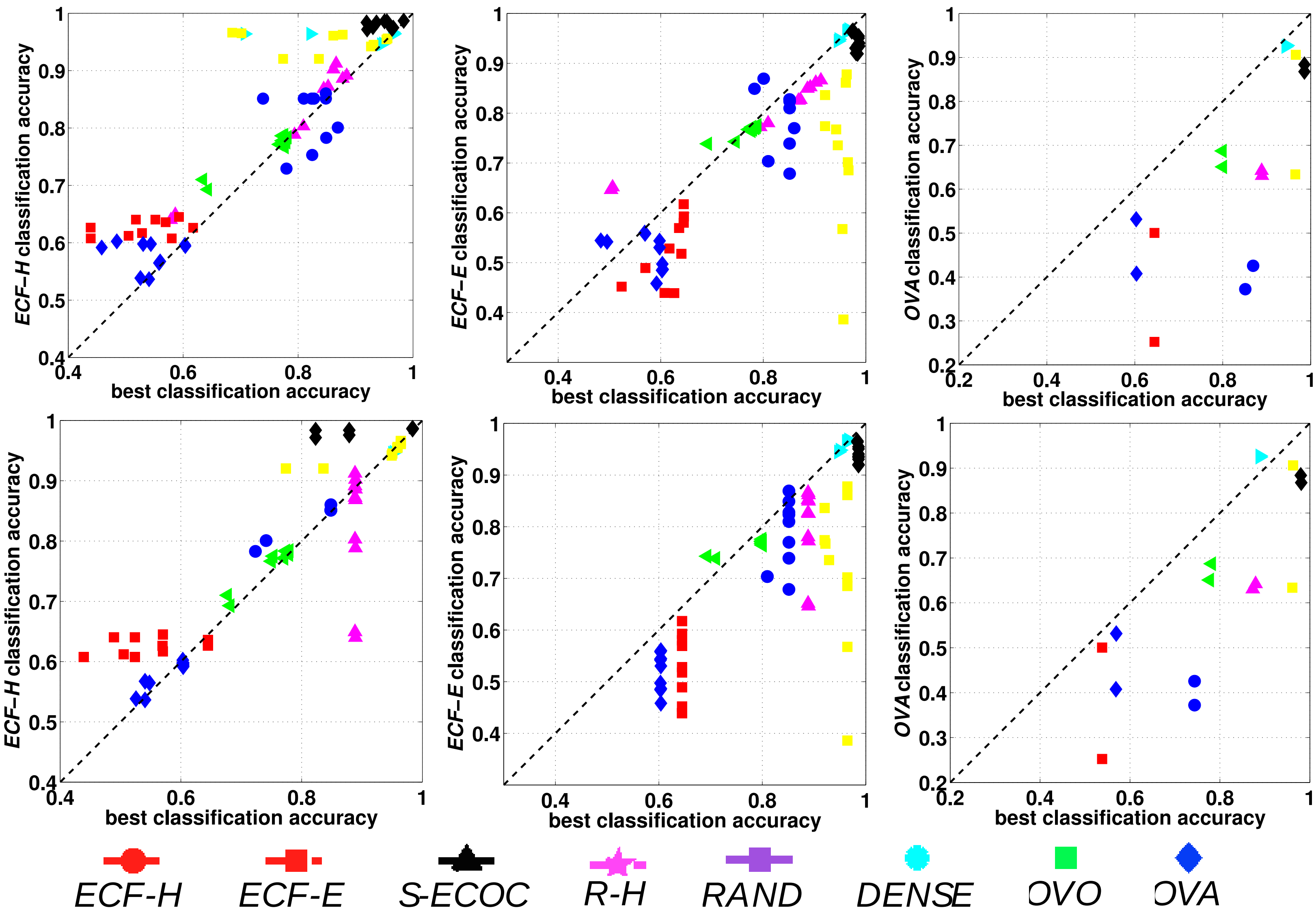
- Classification accuracy as a function of the relative complexity.



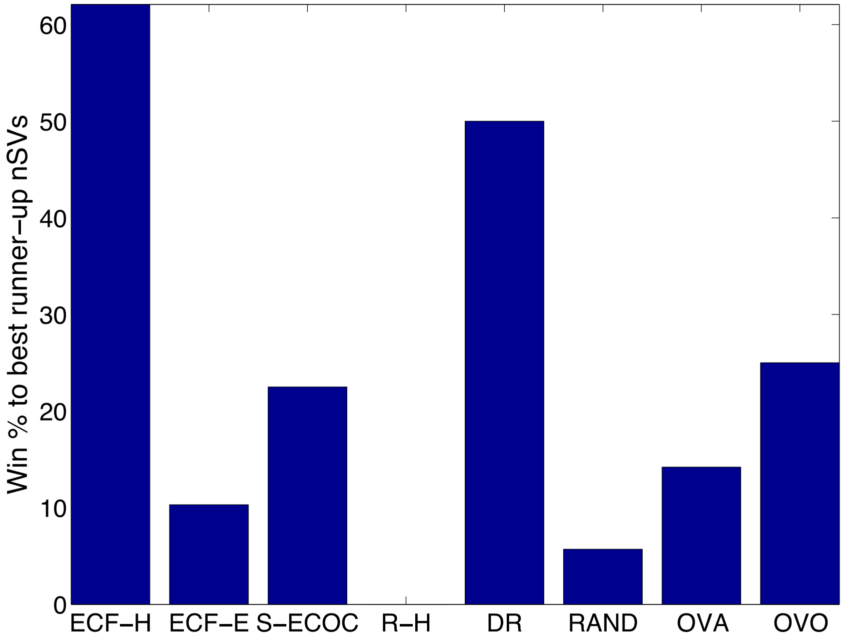
- Classification accuracy as a function of the number of classifiers.

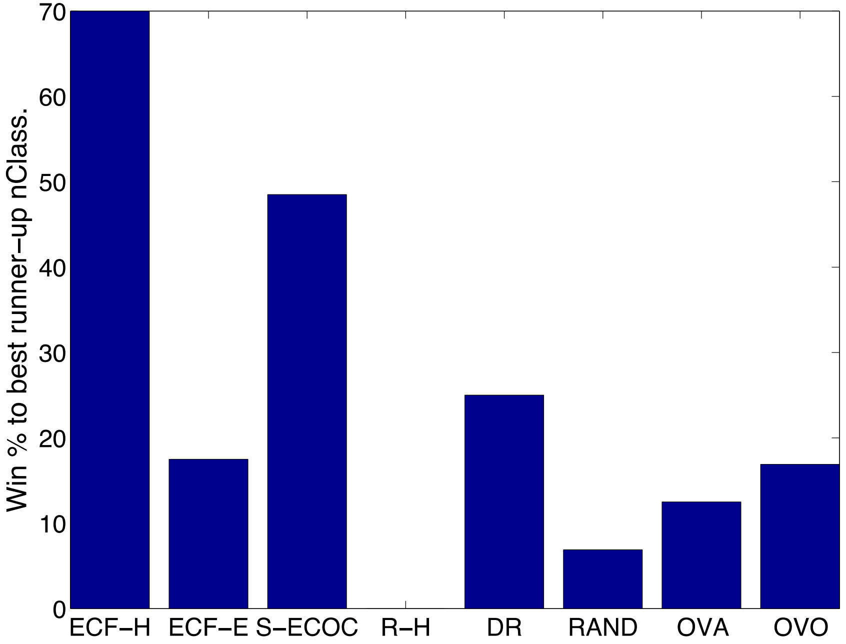


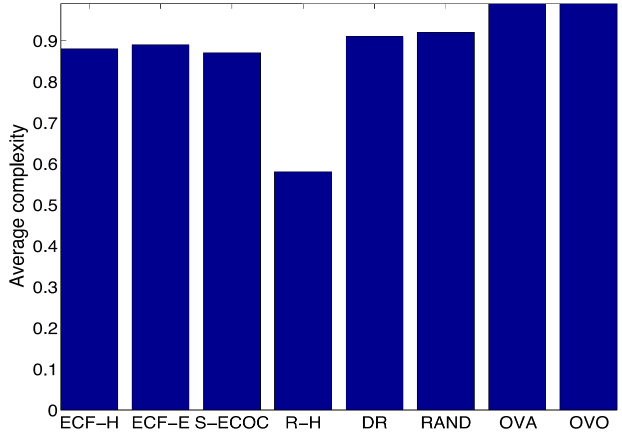
- Comparing ECF-H, ECF-E and OVA with the top performer.



- Proportion of times a method is the top performer.

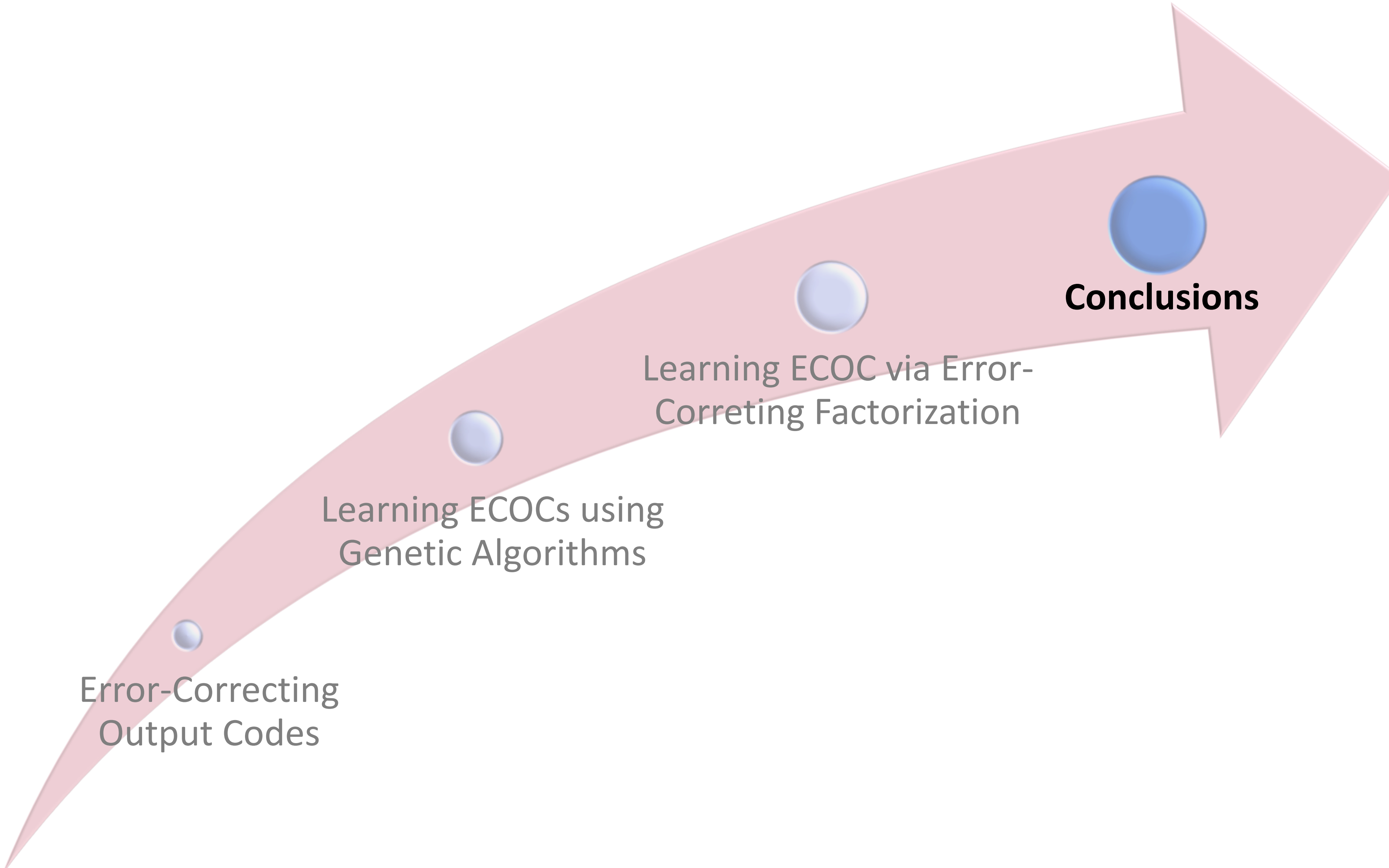






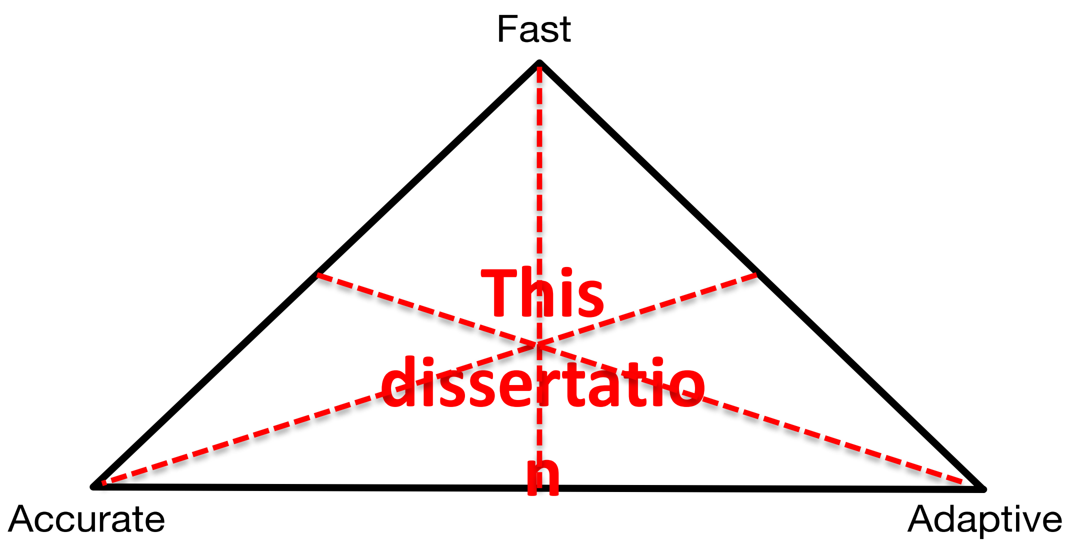
SUMMARY

- We propose a **generalized framework** to build **ECOC matrices** that follow a certain **error-correcting criterion design**.
- The **Error-Correcting Factorization** is formulated a **constrained Coordinate Descent**.
- We **allocate the correction capability** of the ECOC to those categories which are more **prone to confusion**.
- Experiments show that we obtain **higher accuracies** than state of the art methods **with more efficient models**.



Conclusions

- In this dissertation we have proposed approaches for optimizing ECOC classifiers based on various optimization methods.
    1. **Accurate**: by using powerful binary classifiers.
    2. **Adaptive**: exploiting multi-class data distribution.
    3. **Fast**: minimizing the number of classifiers used.



- We have proposed a **novel representation** of **Error-Correction** for an ECOC, enabling us to **allocate Error-Correction** in a flexible manner.

- We have tackled open questions regarding **Error-Correction capabilities** of ECOCs.

  - How do sub-linear ECOC matrices behave? ☑

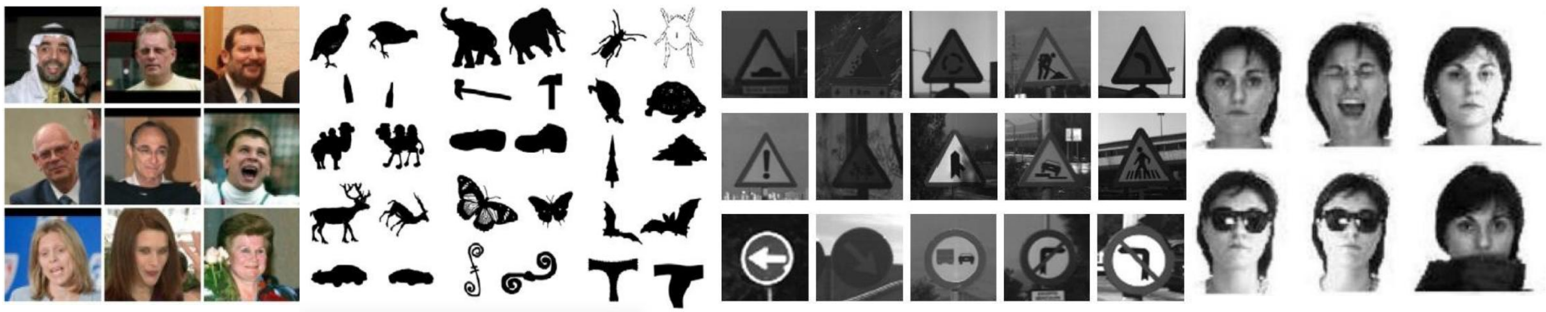
  - Is Error-Correcting distributed evenly on all classes? ☑

  - Can problem-dependent designs profit from the distribution of Error-Correcting capabilities? ☑

  - Is it better to allocate Error-Correction to classes prone to error or to classes not prone to error? ☑

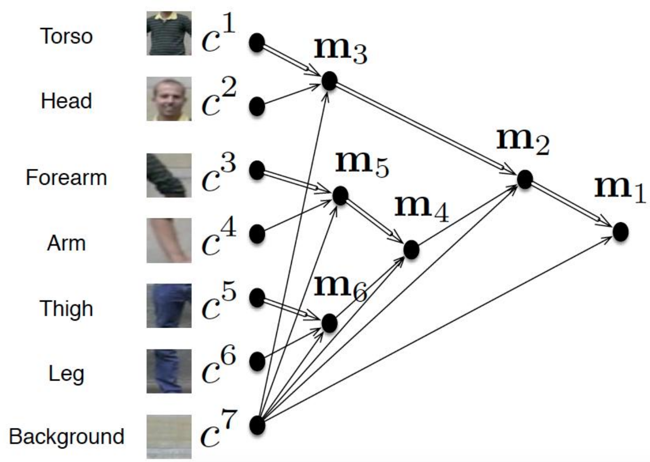
  - What is the minimum problem-dependent number of classifiers needed for an ECOC matrix? ☑

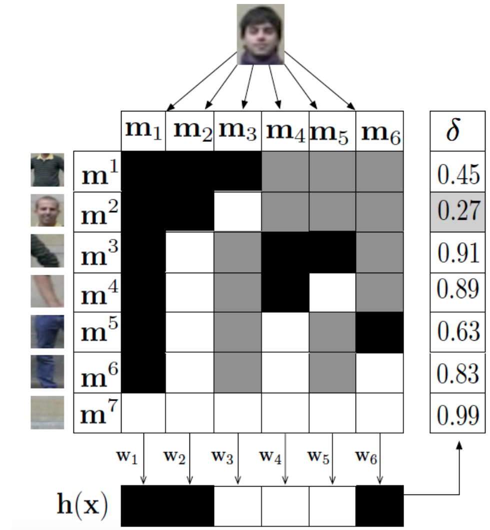
- We have **evaluated** the proposed approaches in **several Multi-class classification tasks:**
  - UCI datasets: Localization sites of proteins, Japanese vowel sounds, written letters, etc.
  - Computer Vision datasets: Face Recognition, Traffic sign recognition, symbol recognition, etc.



- Our approaches have **outperformed state-of-the-art** when analyzing accuracy as a function of the complexity.

- The proposed approaches have been evaluated in the challenging problem of **Human Pose Estimation**
  - Recognize **body limbs** of persons in an image.





- **Error-Correcting Factorization**. Bautista, Miguel Ángel and De la Torre, Fernando and Pujol, Oriol, and Escalera, Sergio. **IEEE Transactions on Pattern Analysis and Machine Intelligence.** Under review, 2015.

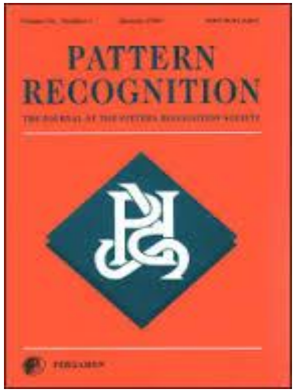
- **A Gesture Recognition System for Detecting Behavioral Patterns of ADHD**. Bautista, Miguel Ángel and Hernández-Vela, Antonio and Escalera, Sergio and Igual, Laura and Pujol, Oriol and Moya, Josep and Violant, Verónica and Anguera, María Teresa. **IEEE Transactions on System, Man and Cybernetics, Part B**. In Press, 2015.

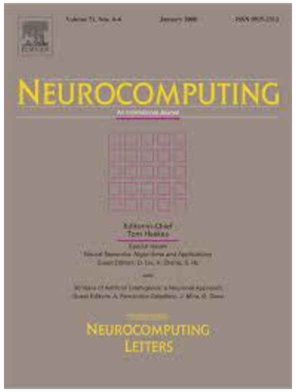
- **On the design of an ECOC-compliant genetic algorithm**. Bautista, Miguel Ángel and Escalera, Sergio and Baró, Xavier and Pujol, Oriol. **Pattern Recognition**, 47, 2, 865–884, 2013, Elsevier.

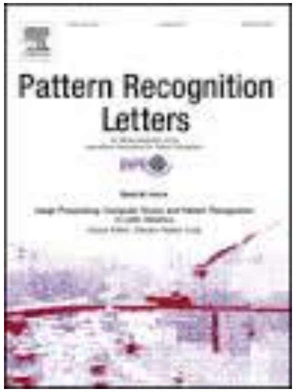
- **HuPBA 8k+: Dataset and ECOC-GraphCut based Segmentation of Human Limbs**. Sanchez, Daniel and Bautista, Miguel Ángel and Escalera, and Sergio. **Neurocomputing**, 2014, Elsevier.

- **Minimal Design of Error-Correcting Output Codes, Bautista**. Miguel Ángel and Escalera, Sergio and Baró, Xavier and Radeva, Petia and Vitriá, Jordi and Pujol, Oriol. **Pattern Recognition Letters**, 33, 6, 693–702, 2011, Elsevier.

- **Probability-based dynamic time warping and bag-of-visual-and-depth-words for human gesture recognition in rgb-d**. Hernández-Vela, Antonio and Bautista, Miguel Ángel and Perez-Sala, Xavier and Ponce-López, Víctor and Escalera, Sergio and Baró, Xavier and Pujol, Oriol and Angulo, Cecilio. **Pattern Recognition Letters**, 50, 112-121, 2014, Elsevier.







- **Introducing the separability matrix for error correcting output codes coding**. Bautista, Miguel and Pujol, Oriol and Baró, Xavier and Escalera, Sergio. Multiple Classifier Systems, 227–236, 2011, Springer.

- **On the Design of Low Redundancy Error-Correcting Output Codes.** Bautista, Miguel and Escalera, Sergio and Baró, Xavier and Pujol, Oriol and Vitria, Jordi and Radeva, Petia. Ensembles in Machine Learning Applications, 21–38, 2011, Springer.

- **Compact evolutive design of error-correcting output codes**. Bautista, MA and Baro, X and Pujol, O and Radeva, P and Vitria, J and Escalera, S. European Conference on Machine Learning Workshops, 119–128, 2010, LNCS.

- **Probability-based Dynamic TimeWarping for Gesture Recognition on RGBD data**. Bautista, Miguel Angel and Hernández-Vela, Antonio and Ponce, Victor and Perez-Sala, Xavier and Baró, Xavier and Pujol, Oriol and Angulo, Cecilio and Escalera, Sergio. International Conference on Pattern Recogntion Workshops, Advances in Depth Image Analysis and Applications, 126–135, 2012, LNCS 7854 - Springer.

- **A genetic inspired optimization for ECOC**. Bautista, Miguel Ángel and Escalera, Sergio and Baró, Xavier and Pujol, Oriol. Structural, Syntactic, and Statistical Pattern Recognition, 743–751, 2012, Springer Berlin Heidelberg.

- **BoVDW: Bag-of-Visual-and-Depth-Words for gesture recognition**. Hernández-Vela, Antonio and Bautista, Miguel Angel and Perez Sala, Xavier and Ponce, Victor and Baró, Xavier and Pujol, Oriol and Angulo, Cecilio and Escalera, Sergio. 21st International Conference on Pattern Recognition (ICPR), 2012, 449–452, 2012, IEEE.

- **Human body segmentation with multi-limb error-correcting output codes detection and graph cuts optimization**. Sánchez, Daniel and Ortega, Juan Carlos and Bautista, Miguel Ángel and Escalera, Sergio. Pattern Recognition and Image Analysis, 50–58, 2013, Springer Berlin Heidelberg.

- **Chalearn looking at people challenge 2014: Dataset and results**. Escalera, Sergio and Bar, X and Gonzlez, J and Bautista, Miguel A and Madadi, Meysam and Reyes, Miguel and Ponce, V and Escalante, Hugo J and Shotton, Jamie and Guyon, Isabelle. IEEE European Conference Computer Vision Looking At People Workshop, 2014.

- **Learning To Segment Humans By Stacking Their Body Parts**. Puertas, Eloi and Bautista, MA and Sanchez, Daniel and Escalera, Sergio and Pujol, Oriol. IEEE European Conference Computer Vision Looking At People Workshop, 2014.

Learning Error-Correcting Representations For Multi-class Problems

Miguel Ángel Bautista Martín

Thank you