10th International Workshop on
Multiple Classifier Systems
Naples, Italy, June 15-17, 2011

Introducing the Separability Matrix for ECOC coding

Miguel Angel Bautista, Sergio Escalera, Xavier Baró & Oriol Pujol

UNIVERSITAT DE BARCELONA

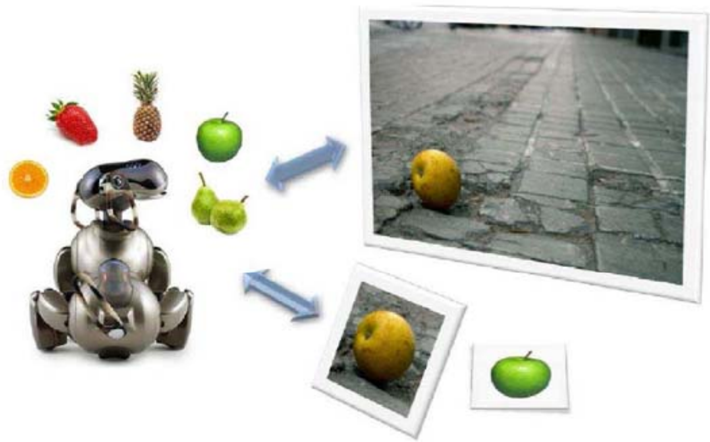
Centre de Visió per Computador

Outline

- Classification problems and the ECOC framework.
- Motivation.
- The Separability Matrix.
- An application of the Separability Matrix for coding ECOCs.
  - The Confusion-Separability Extension Coding.
- Experiments and Results.
- Conclusions and Future Work.

Introduction to the ECOC framework

- Classification tasks are a well known type of supervised learning problem. The goal is to classify and object among a certain number of possible categories.
- The ECOC framework has proven to be a powerful tool to deal with multi-class classification problems.
- This framework is composed of two different steps :
  - **Coding** : Decompose a given *N*-class problem into a set of *n* binary problems.
  - **Decoding** : Given a test sample *s*, determine its category.

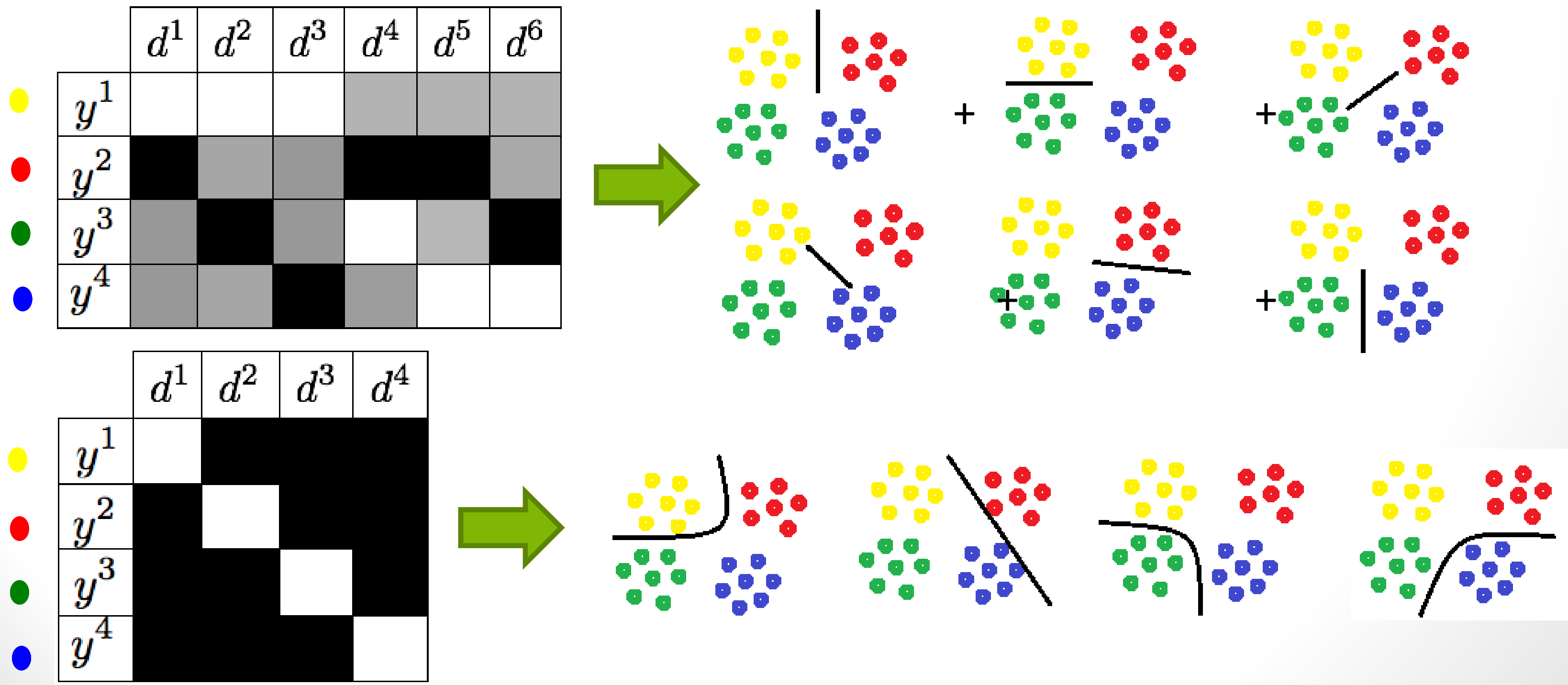


Introduction to the ECOC framewo

- At the coding step a decomposition of the *N*-class problem into *n* binary problems is build and represented into a matrix.

$$M_{N\times n} \in \{-1, +1, 0\}$$

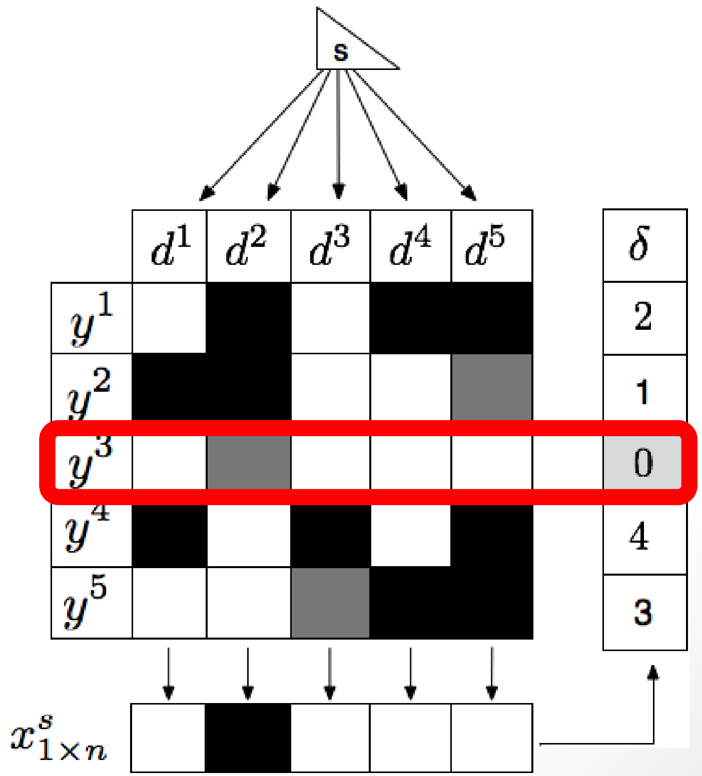
- The columns of the matrix represent the binary problems.
- The rows of the matrix represent the codes of the *N* classes.



Introduction to the ECOC framework

- At the decoding step a new sample *s* is classified by comparing the binary responses to the rows of *M* by means of a decoding measure δ.
- Different types of decoding based on the distance used (i.e. Hamming Decoding, Euclidean Decoding, etc.)

$$\arg\min_i \delta(x^s, y^i)$$



Motivation

- Standard predefined or random strategies may not be suitable for a given problem.
- Find an optimum coding matrix *M* for a given problem was proved to be an *NP-Complete* problem [1].
- In [2] we show how reduced codes can perform has well as standard designs with far less number of dichotomizers.
- Those reduced codes can be extended in a problem-dependent way to benefit from error-correcting principles.

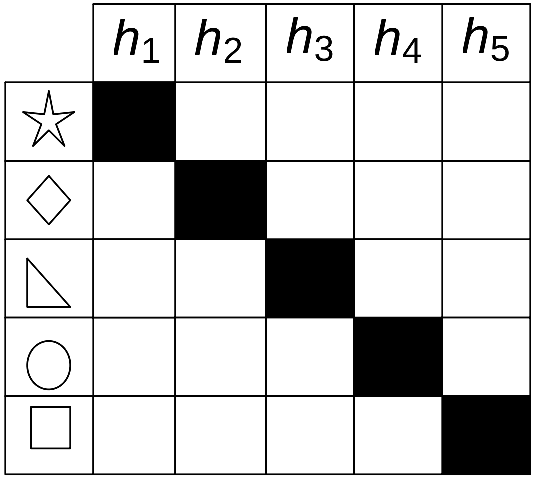
[1] On the learnability and design of Error Correcting Output Codes, K.Crammer & Y. Singer
[2] Compact Evolutionary Design of Error Correcting Output Codes, M. Bautista, S. Escalera, X. Baró, O. Pujol, J. Vitrià and P. Radeva

The Separability Matrix

- The Separability matrix *S* contains the **pairwise distance** δ between the codewords in *M*.
- With this matrix we can analyze the **correction capability** ρ of *M* since,

$$\rho = \frac{\min(S) - 1}{2}$$

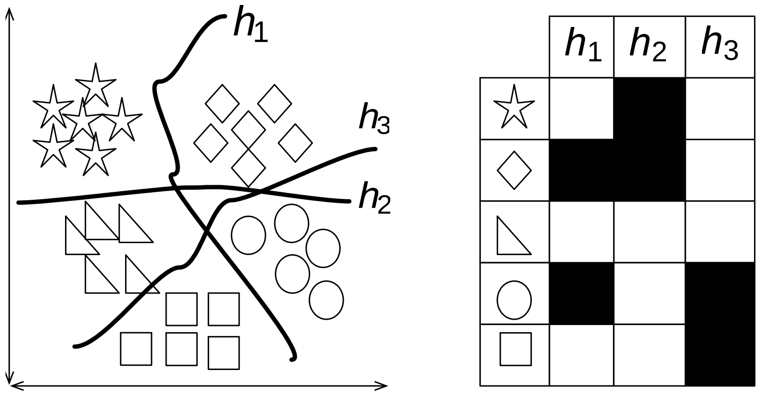
- Standard ECOC designs shown constant Separability matrices.



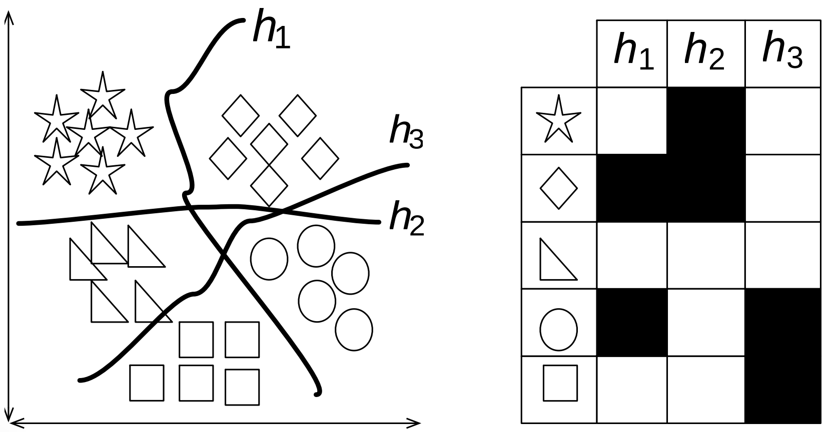
	☆	◇	◺	○	□
☆	0	2	2	2	2
◇	2	0	2	2	2
◺	2	2	0	2	2
○	2	2	2	0	2
□	2	2	2	2	0

An application of the Separability Matrix for coding ECOCs

- In [2] we show how reduced codes can perform has well as standard designs with far less number of dichotomizers.



- Identify those classes that need an increment of distance in order to benefit from error-correcting principles.



	☆	◇	◺	○	□
☆	0	1	1	3	2
◇	1	0	2	2	3
◺	1	2	0	2	1
○	3	2	2	0	1
□	2	3	1	1	0

The Confusion-Separability Extension coding (CSE Coding)

- Use the **Confusion matrix** over a validation subset to find the most confused classes.
- Use the **Separability matrix** to find the classes that need at increment of distance in order to benefit from error-correcting principles.
- Compute and **Extension Matrix** of a Binary Compact ECOC matrix which is focused on the classes that show both **high confusion and low separation**.
- Extend the coding matrix until a **maximum length of *N* dichotomies**.

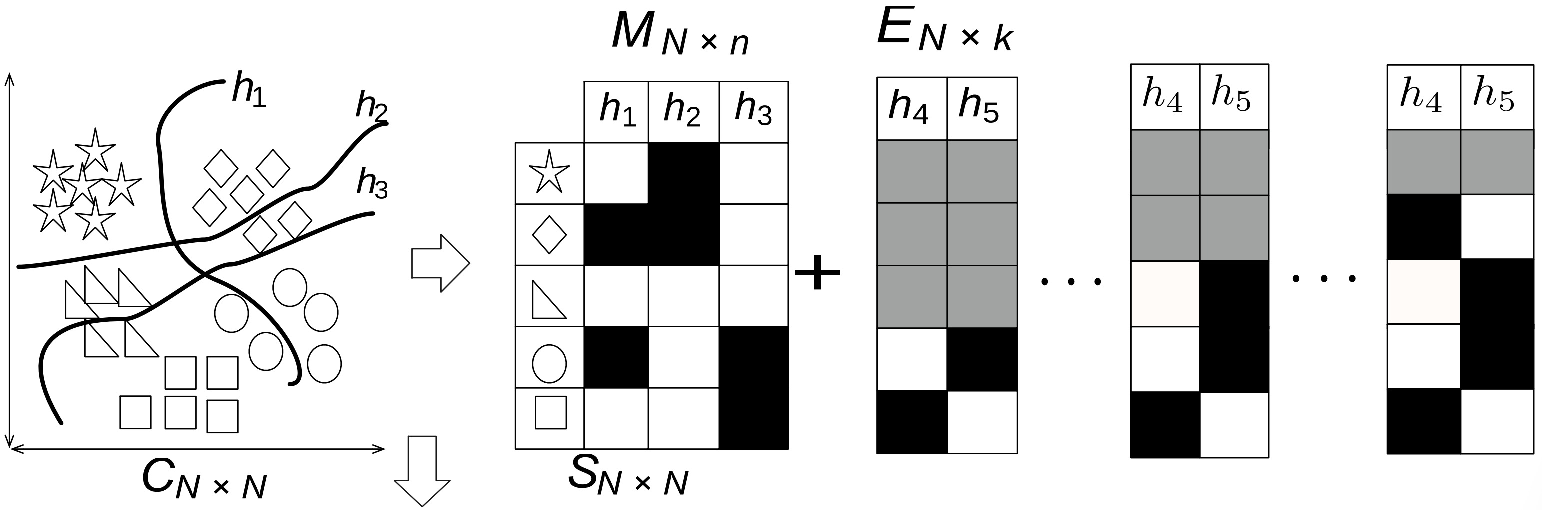
The CSE Coding Algorithm

① Find the most confused classes : $\arg\max_{c^i, c^j}(C_{i,j} + C_{j,i})$

② Compute and Extension matrix *E* which increments $\delta(y^i, y^j) \geq \varrho$

③ Fill the empty Extension codes taking into account the confusion of c^i, c^j with the rest of the classes.

④ Update Confusion and Separability matrices.



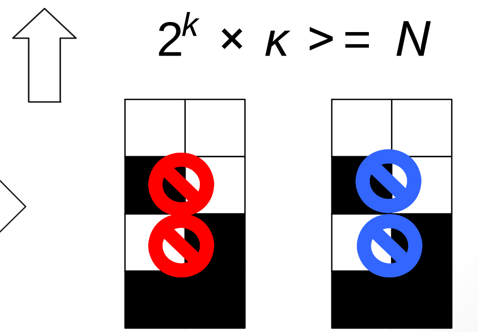
$C_{N\times N}$

	☆	◇	◺	○	□
☆	6	0	0	0	0
◇	0	4	0	2	0
◺	0	0	3	0	2
○	0	0	0	3	2
□	0	0	0	0	5

$S_{N\times N}$

	☆	◇	◺	○	□
☆	0	1	1	3	2
◇	1	0	2	2	3
◺	1	2	0	2	1
○	3	2	2	0	1
□	2	3	1	1	0

$2^k \times \kappa >= N$



$\varrho = 3$

Experiments: Data

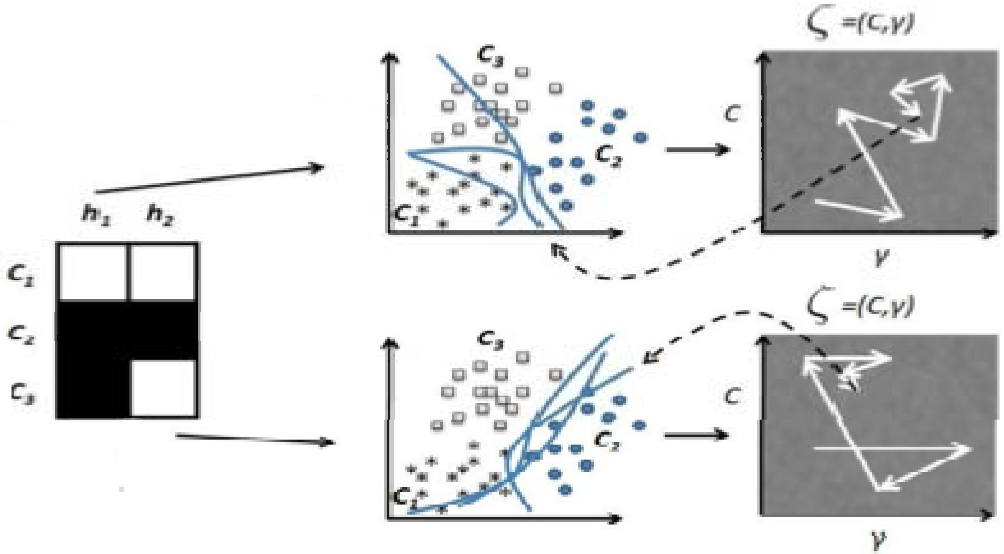
- We tested the novel methodology on several public datasets from the UCI Machine Learning Repository.

Problem	#Training samples	#Features	#Classes
Dermathology	366	34	6
Ecoli	336	8	8
Vehicle	846	18	4
Segmentation	2310	19	7
Glass	214	9	7
Vowel	990	10	11
Yeast	1484	8	10

- In addition we perform experiments with 3 public Computer Vision problems.
  - The ARFace dataset with 20 classes.
  - The Traffic Sign dataset with 36 classes.
  - The MPEG dataset with 70 classes.

Experiments: Methods and Settings

- We compare the One vs. All and Dense Random with the CSE coding with $\varrho \in \{3, 5\}$
- Classification results are the average over a Stratified 10 fold CV.
- We use the SVM-RBF and AdaBoost as our base classifier.
- An optimization process is carried out to tune the parameters of the SVMs.
- SVM-RBF classifiers have 2 parameters to optimize (C & Y).



Experiments and Results

- Results for UCI and Computer Vision experiments with SVM as the base classifier.

	One vs. All ECOC		CSE ECOC $\varrho = 3$		CSE ECOC $\varrho = 5$		Dense Random ECOC	
Data set	Perf.	Classif.	Perf.	Classif.	Perf.	Classif.	Perf.	Classif.
Vowel	55.0±10.5	11	66.9±7.8	9.2	**69.8±6.3**	10.6	67.9±8.3	11
Yeast	41.0±7.3	10	54.7±11.8	5.7	53.0±9.3	9.5	**54.9±6.4**	10
Ecoli	**78.9±3.5**	8	76.4±4.4	7	78.6±3.9	7.4	72.1±2.7	8
Glass	51.6±10.2	7	**55.5±7.6**	6	52.7±8.4	3	42.8±11.02	7
Segment	**97.3±0.7**	7	96.9±0.8	6.6	96.6±1.0	6.2	96.6±1.3	7
Derma	97.1±1.2	6	**97.1±0.9**	5.2	95.9±1.2	3	95.7±0.8	6
Vehicle	80.1±4.0	4	**81.1±3.5**	3	70.6±3.4	3	81.1±3.6	4
MPEG7	83.2±5.1	70	88.5±4.5	15	89.6±4.9	20.4	**90.0±6.4**	70
ARFaces	76.0±7.22	50	80.7±5.2	13.8	84.6±5.3	20.2	**85.0±6.3**	50
Traffic	91.3±1.1	36	95.7±0.92	12.2	**96.6±0.8**	19	93.3±1.0	36
Rank & #	3.0	20.8	**2.2**	**8.8**	2.3	10.3	2.5	20.8

Experiments and Results

- Results for UCI and Computer Vision experiments with AdaBoost as the base classifier.

	One vs. All ECOC		CSE ECOC $\varrho = 3$		CSE ECOC $\varrho = 5$		Dense Random ECOC	
Data set	Perf.	Classif.	Perf.	Classif.	Perf.	Classif.	Perf.	Classif.
Vowel	40.6±1.3	11	44.7±0.8	10	46.5±1.2	10.6	**47.0±1.2**	11
Yeast	36.8±1.1	10	**45.6±0.4**	9.6	42.9±1.0	9.5	40.8±1.3	10
Ecoli	71.5±10.9	8	68.1±8.3	7.4	63.3±9.2	7.4	**75.0±7.8**	8
Glass	**53.8±12.1**	7	52.8±13.5	6	44.5±10.8	6	49.5±10.9	7
Segment	**96.4±0.7**	7	95.0±0.3	6.8	94.8±0.9	6.2	95.3±1.0	7
Derma	**89.3±4.9**	6	77.6±6.3	5.4	76.0±5.3	3	76.7±5.3	6
Vehicle	**73.6±1.3**	4	72.7±1.9	4	62.9±1.4	3	72.7±1.5	4
MPEG7	54.4±7.2	70	65.5±9.5	15	73.7±8.3	24.3	**86.5±6.4**	70
ARFaces	36.3±7.2	50	53.8±5.2	13.8	62.8±8.3	20.4	**81.5±6.3**	50
Traffic	80.6±6.2	36	81.3±8.1	12.2	87.4±7.9	20.6	**91.2±5.3**	36
Rank & #	2.6	20.8	2.4	9.16	3.0	10.89	**1.9**	**20.8**

Conclusions and Future Work

- The **Separability Matrix** is introduced as a novel tool to analyze and enhance ECOC coding designs.
- The **Extension Algorithm** proposed can be applied to any existing ECOC scheme.
- A new coding design based on the Separability matrix is introduced **obtaining significant performance improvements** over state-of-the-art ECOC designs.
- The proposed methodology **reduces the number of base classifiers** needed in comparison with state-of-the-art designs.
- A possible improvement will be **to optimize the initial Compact ECOC** coding matrix.

Thank you!