ON THE STRUCTURE OF CLASSIFICATIONS RECOGNIZABLE BY THE PATTERN RECOGNITION ALGORITHM "CORE-n"

V.A.Zheligovsky[1,2,3], S.Schwarz[1,4]

[1]School of Mathematics & Statistics,
University of Sydney, NSW 2006, Australia

[2]International Institute of Earthquake Prediction Theory and Mathematical Geophysics Ac. Sci. 79 k.2, Warshavskoe av., 113556, Moscow

To appear in "Computational Seismology" February 4, 1997

Abstract. The pattern recognition algorithm "Core-n" has been successfully applied in different areas ranging from geophysics to social sciences. We present results on the geometry in the object space of classifications recognizable by the algorithm, including those obtained in computer experiments with low-dimensional problems.

0. Introduction.

The pattern recognition algorithm "Core-n" introduced in the sixties (Bongard, 1967) has been successfully applied in different areas, ranging from political sciences (Lichtman & DeCell, 1990) and finance (Caputo & Kolari, 1990) to geophysics. The use of the algorithm for the determination of earthquake-prone territories, first done by Gelfand et al. (1972, 1976), was probably the most spectacular application. Later forecasts for all major seismic-active regions of the Earth were gradually worked out at the Department of Computational Seismology of the Institute of Earth's Physics of the Academy of Sciences of the USSR, reorganized in 1989 into the International Institute of Earthquake Prediction Theory and Mathematical Geophysics Ac. Sci. USSR (see issues 6-27 of "Computational Seismology", the annual collection of papers of the Department / Institute). An analysis of the conformity of predictions and of the subsequent seismic events has been carried out - see, e.g., numerous papers in "Computational Seismology".

The vast range of applications justifies a systematic study of the properties of the algorithm. However, to the best of our knowledge it has not been undertaken so far. The paper presents some results of such a study including those of an exhaustive computer check of low-dimensional cases. The following questions were addressed:

i. What are the characteristic properties of recognizable classifications? Which classifications cannot be recognized?

In particular, consideration of non-recognizable classifications is important because their abundance characterizes the "recognizing power" of the algorithm.

ii. What are geometric criteria for the stability of recognition?

Such criteria are of interest for applications, since they potentially allow one to avoid computationally expensive exhaustive checks of the free parameter sets when stability of a particular recognition is considered.

iii. How are results of recognition altered when the objects' coding is supplemented by "redundant" descriptors?

This question is of practical importance, since usually it is not known a priori what features used to represent objects correspond to the nature of the problem of classification at hand.

0.1. Description of the pattern recognition algorithm Core-n.

For completeness of presentation, the Core-n algorithm is defined in this Section.

The algorithm performs a *dichotomy* (i.e., a classification into two classes, C_1 and C_2) of *objects*, mapped by means of *coding* onto a set of *binary vectors* of a given dimension M (i.e.

[3]E-mail: zheligo_v@maths.usyd.edu.au
[4]E-mail: simons@maths.usyd.edu.au

vectors, each of M components of which is equal to 0 or 1). In practice, a researcher prepares a questionnaire consisting of M questions, and each vector component - a *descriptor* - represents an affirmative (e.g., $\mathbf{v}_p = 1$) or negative ($\mathbf{v}_p = 0$) answer to the corresponding question for a particular object. Thus objects are identified with binary vectors $\mathbf{v}$ representing vertices of the M-dimensional hypercube $X^M = \{0,1\}^M$. It is vital for a successful application of the algorithm that the questionnaire is complete in the sense that the totality of answers to the questions contains enough information to enable a correct assignment of an object to either class.

Unless stated otherwise it is not assumed below that the object coding is surjective, i.e. some vectors $\mathbf{v} \in X^M$ may represent no objects. The image of the objects' coding is denoted by O.

The operation of the algorithm depends on the values of six *free parameters*: n, A_1, R_1, A_2, R_2 and Δ as explained below. It is the responsibility of the user to fine-tune these parameters.

The algorithm operates in two stages: *learning* and *voting*. In the first stage two subsets, $L_1 \subset C_1$ and $L_2 \subset C_2$, are required; $L_1 \sqcup L_2$ is called the *learning material.*

Definition 1. A *pattern* $\sigma = \{(p_i, b_i) | 1 \le i \le k\}$ of rank k (rank$\sigma = k$) is a set of k pairs comprised of distinct descriptors p_i, $1 \le p_i \le M$, and associated binary values $b_i \in \{0,1\}$.

Definition 2. A vector $\mathbf{v}$ *possesses* a pattern $\sigma = \{(p_i, b_i) | 1 \le i \le k\}$ if $\mathbf{v}_{p_i} = b_i$, $1 \le i \le k$.

Definition 3. A function $\sigma : X^M \to \{0,1\}$ is *associated with a pattern* σ if $\sigma(\mathbf{v}) = 1$ when a vector $\mathbf{v}$ possesses the pattern σ, and $\sigma(\mathbf{v}) = 0$ otherwise.

During the learning stage the numbers

$$N_i(\sigma) = \sum_{\mathbf{v} \in L_i} \sigma(\mathbf{v}) = |\{\mathbf{v} \in L_i \mid \mathbf{v} \text{ possesses } \sigma\}| \quad (i = 1, 2)$$

are determined for each pattern σ of rank$\sigma \le n$. As a result of learning two lists of patterns $\mathcal{S}^i = \{\sigma^i_j\}$ $(i = 1, 2)$ characterizing the first and the second classes, respectively, are compiled:

Definition 4. A pattern σ *characterizes* the i-th class $(i = 1, 2)$ if

$$\text{rank } \sigma \le n, \quad N_i(\sigma) \ge A_i \quad \text{and} \quad N_{3-i}(\sigma) \le R_i.$$

Remark 1. Without loss of generality it may be assumed $A_i \ge 0$, $R_i \ge 0$ and either $A_i = \infty$ implying $\mathcal{S}^i = \emptyset$, or for surjective coding $2^{M-n} \le A_i + R_i \le 2^{M-1}$ (an analogous inequality for $O \ne X^M$ is more complicated).

At the second stage vectors are attributed to classes by "voting": $\mathbf{v} \in O$ is assigned to the first class if $E_1(\mathbf{v}) - E_2(\mathbf{v}) \ge \Delta$, and to the second class otherwise; here

$$E_i(\mathbf{v}) = \sum_{\sigma \in \mathcal{S}^i} \sigma(\mathbf{v}) = |\{\sigma \in \mathcal{S}^i \mid \mathbf{v} \text{ possesses } \sigma\}| \quad (i = 1, 2). \tag{1}$$

Usually the following "common sense" criteria are employed as guidelines for the choice of the free parameters:

i. No patterns should characterize both classes simultaneously. (This condition was not enforced in our tests.)

ii. Vectors from the learning material $L_1 \sqcup L_2$ should be recognized correctly.

Remarks.

2. All vectors $\mathbf{v} \in L_1$ are recognized correctly if and only if $\Delta \le \min_{\mathbf{v} \in L_1}(E_1(\mathbf{v}) - E_2(\mathbf{v}))$.

3. In applications for the recognition of earthquake-prone areas only vectors from L_1 (high-seismic areas) are required to be recognized correctly: in this context transition of objects from L_2 (low-seismic) to C_1 (high-seismic) is allowed.

0.2. Stability of pattern recognition.

The notion of stability of recognition was introduced by Gvishiani & Gurvich (1984), where simple geometric criteria of stability were constructed for the algorithm based on the introduction of a weighted metric in the object space and for the so-called algorithm of "voting by normal patterns". Stability of recognition is a consequence of the logical self-consistency of the procedure. Suppose that in some problem coding represents objects with no loss of significant information, the learning material adequately represents classes, and an application of the pattern recognition algorithm provides the correct classification. Then a further increase of information in the learning material should not alter the classification:

Definition 5. A classification $C_1 \sqcup C_2$ is called *stably recognizable* with learning material $L_1 \sqcup L_2$ if for any learning material $L'_1 \sqcup L'_2$ such that $L_1 \subset L'_1 \subset C_1$, $L_2 \subset L'_2 \subset C_2$ there exists a set of free parameters $A_1(L'_1, L'_2)$, $R_1(L'_1, L'_2)$, $A_2(L'_1, L'_2)$, $R_2(L'_1, L'_2)$ and $\Delta(L'_1, L'_2)$ yielding recognition of the initial classification $C_1 \sqcup C_2$.

Examination of stability based on the definition is computationally expensive, since it involves consideration of all subsets of $(C_1 \sqcup C_2) \backslash (L_1 \sqcup L_2)$ and loops over all possible combinations of free parameters. (In certain geometrically simple cases Theorems 1 and 2 guarantee the stability of recognition by Core-n.)

Stability of recognition of a classification $C_1 \sqcup C_2$ implies, in particular, that it can be recognized if the classification itself is used as a learning material. This justifies a special attention to self-recognizable classifications.

1. Some sufficient conditions for recognizability of classifications.

Theorems 1 - 4 can be extended by interchanging the roles of the two classes C_1 and C_2.

Definition 6. $H \subset X^M$ is a *hyperface* of codimension k (codim$H = k$) if there exists a pattern σ of rank k such that a vector $\mathbf{v}$ possesses σ if and only if $\mathbf{v} \in H$. Then $H = H_\sigma$ is called the hyperface associated with the pattern σ, and $\sigma = \sigma_H$ is called the pattern associated with the hyperface H.

Definition 6 agrees with the standard definition of a hyperface in geometry. The correspondence between hyperfaces of codimension k and patterns of rank k is bijective. A hyperface of codimension k is comprised of 2^{M-k} vertices of the hypercube X^M (vectors). A hyperface of codimension M is a vertex (a vector).

Theorem 1. Let $C_1 = O \cap H$, where H is a hyperface. If

$$|L_1| \geq \tfrac{1}{2}(|H| + 1), \tag{2}$$

then the learning material $L_1 \sqcup L_2$ provides a stable recognition of the classification $O = C_1 \sqcup C_2$.

Remark 4. If codim$H \geq M - 1$, (2) implies $L_1 = C_1 = H$.

Proof. Consider an arbitrary learning material $L'_1 \sqcup L'_2$, where $L_1 \subset L'_1 \subset C_1$, $L_2 \subset L'_2 \subset C_2$. Assume $A_2 = \infty$, $R_2 = 0$; hence no patterns characterize C_2.

Select

$$A_1 = |L_1|, \quad R_1 = \infty. \tag{3}$$

Then due to (2) the rank of any pattern characterizing C_1 cannot exceed $k =$codim$H$. Assuming $\sigma_H = \{(p_i, b_i) | 1 \leq i \leq k\}$ denote $\mathcal{P}_H = \{p_i,\ 1 \leq i \leq k\}$. Consider a pattern $\sigma = \{(p'_i, b'_i) | 1 \leq i \leq k'\}$ of rank $k' \leq k$ and denote $\mathcal{P}_\sigma = \{p'_i,\ 1 \leq i \leq k'\}$. Then evidently $|L'_1 \cap H_\sigma| \leq |H \cap H_\sigma| \leq 2^{-|\mathcal{P}_\sigma \backslash \mathcal{P}_H|}|H|$. Hence (2) and (3) imply that a pattern σ characterizes C_1 if and only if $H \subset H_\sigma$ and rank$\sigma \leq n$. There are

$$r = \begin{cases} 2^k - 1 & \text{if } k \leq n \\ \sum_{s=1}^{n} \binom{k}{s} & \text{if } k > n \end{cases}$$

such patterns. Clearly $H = \cap_{m=1}^{r} H^m$, where the intersection is over all hyperfaces $H^m \supset H$ of codim$H^m \leq \min(n, k)$. Thus $\Delta = r$ furnishes the classification $C_1 \sqcup C_2$.$\triangleright$

Theorem 2. Suppose

$$C_1 = O \bigcap \bigcup_{i=1}^{p} H^i, \tag{4}$$

where each H^i is a hyperface of codimension $k \leq n$. Consider $L_1 \subset C_1,\ L_2 \subset C_2 = O \backslash C_1$. Denote

$$Q = \min_{1 \leq i \leq p} |L_1 \cap H^i|. \tag{5}$$

If

$$k = n \quad \text{or} \quad Q \geq \tfrac{1}{2}(2^{M-k} + 1) \tag{6}$$

and

$$|L_2 \cap H| > 0 \tag{7}$$

for any hyperface H of codimension k such that $H \cap C_2 \neq \emptyset$, then the learning material $L_1 \sqcup L_2$ yields a stable recognition of the classification $C_1 \sqcup C_2$.

Remarks.

5. Representation (4) of C_1 is in general non-unique.

6. In (4) the H^i may intersect.

7. For $n = M$ and $k = M - 1$, (6) implies $L_1 = C_1$.

8. If $k = M$, then $n = M$ and (7) implies $L_2 = C_2$. C_1 and C_2 can be arbitrary, but as the proof shows, in this case the algorithm operates "objectwisely"; it is an a priori precise knowledge of the classes and not a genuine identification of their structure which is responsible for the successful recognition.

Proof. Consider a learning material $L_1' \sqcup L_2'$, where $L_1 \subset L_1' \subset C_1,\ L_2 \subset L_2' \subset C_2$. Select $A_2 = \infty,\ R_2 = 0$; thus no patterns characterize C_2.

Assume $A_1 = Q,\ R_1 = 0$. Due to (6) the rank of any pattern characterizing C_1 will not exceed k. Consider a pattern σ, rank$\sigma \leq k$. Either $O \cap H_\sigma \subset C_1$, or there exists a hyperface $H \subset H_\sigma$ of codimension k such that $H \cap C_2 \neq \emptyset$. In the latter case $|L_2' \cap H_\sigma| \geq |L_2 \cap H| > 0$ by (7) and hence σ does not characterize C_1. Therefore if a pattern σ characterizes C_1, then $\sigma(\mathbf{v}) = 1 \Rightarrow \mathbf{v} \notin C_2$.

Because of (5), each pattern σ_{H^i} characterizes C_1. Thus (4) implies that any $\mathbf{v} \in C_1$ possesses at least one pattern σ characterizing C_1. Hence $\Delta = 1$ provides the desired classification $C_1 \sqcup C_2$.$\triangleright$

Theorem 3. If a classification $C_1 \sqcup C_2 = X^M$ is self-recognizable by Core-n with

$$A_i \geq \min(2^{M-n-1},\ |C_i|,\ 1 + \max_H |C_i \cap H|) \quad \forall i = 1, 2 \tag{8}$$

(here the maxima are over all hyperfaces H of codimension $n+1$), then it is also self-recognizable by Core-n' $\forall n' > n$ with the original values of A_i, R_i.

Proof. A classification is recognized with some values of A_i, R_i if and only if

$$\min_{\mathbf{v} \in C_1} (E_1(\mathbf{v}) - E_2(\mathbf{v})) > \max_{\mathbf{v} \in C_2} (E_1(\mathbf{v}) - E_2(\mathbf{v})) \tag{9}$$

(cf. Remark 2). In particular, (9) holds for the original self-recognition by Core-n.

Consider a recognition by Core-n', $n' > n$, with the original values of A_i, R_i. A pattern will characterize a class either if it did so in the original recognition, or, possibly, if it is a new pattern of rank strictly larger than n. Denote by $E_i'(\mathbf{v})$ the number of patterns possessed by a vector $\mathbf{v}$ characterizing the first class in the new recognition. We first show that

$$\min_{\mathbf{v} \in C_1} (E_1'(\mathbf{v}) - E_2(\mathbf{v})) > \max_{\mathbf{v} \in C_2} (E_1'(\mathbf{v}) - E_2(\mathbf{v})). \tag{10}$$

Due to (8) three possibilities arise:

i. $A_1 > \max_{\,\mathrm{codim} H = n+1} |C_1 \cap H|$. Then no new patterns will characterize C_1 in the recognition with $n' > n$. Hence $E_1'(\mathbf{v}) = E_1(\mathbf{v})\ \forall \mathbf{v}$, and (10) is a consequence of (9).

ii. $A_1 = 2^{M-n-1}$. Any new pattern characterizing C_1 by virtue of this identity cannot be possessed by any $\mathbf{v} \in C_2$. (Only new patterns of rank $n+1$ characterizing C_1 can emerge in this case.) Therefore

$$\min_{\mathbf{v}\in C_1}(E_1'(\mathbf{v})-E_2(\mathbf{v})) \geq \min_{\mathbf{v}\in C_1}(E_1(\mathbf{v})-E_2(\mathbf{v})), \quad \max_{\mathbf{v}\in C_2}(E_1'(\mathbf{v})-E_2(\mathbf{v})) = \max_{\mathbf{v}\in C_2}(E_1(\mathbf{v})-E_2(\mathbf{v})),$$

and (10) follows from (9).

iii. $A_1 = |C_1|$. Suppose that e new patterns characterizing C_1 emerge. By virtue of the equality, they are possessed by each vector in C_1. Thus

$$\min_{\mathbf{v}\in C_1}(E_1'(\mathbf{v}) - E_2(\mathbf{v})) = e + \min_{\mathbf{v}\in C_1}(E_1(\mathbf{v}) - E_2(\mathbf{v})),$$

$$\max_{\mathbf{v}\in C_2}(E_1'(\mathbf{v}) - E_2(\mathbf{v})) \leq e + \max_{\mathbf{v}\in C_2}(E_1(\mathbf{v}) - E_2(\mathbf{v})),$$

and (9) again implies (10).

The analogue of (9) for the new recognition can be derived from (10) considering similarly the three possibilities for A_2.$\triangleright$

2. Some sufficient conditions for non-recognizability of classifications.

Theorem 4. Let $\mathcal{P} = \{p_i|\ 1 \leq p_i \leq M\}$ be a set of descriptors, $M \geq |\mathcal{P}| > n$. Suppose

$$C_1 = \left\{ \mathbf{v} \in X^M \left| \sum_{p\in\mathcal{P}} \mathbf{v}_p \text{ is odd} \right. \right\}, \quad C_2 = X^M \backslash C_1.$$

Then the classification $C_1 \sqcup C_2$ cannot be recognized with any learning material $L_1 \sqcup L_2$.

Proof. Let a set of free parameters furnish the classification. Then $E_1(\mathbf{v}) - E_2(\mathbf{v}) \geq \Delta$ $\forall \mathbf{v} \in C_1$, and hence

$$\sum_{\mathbf{v}\in C_1} E_1(\mathbf{v}) - E_2(\mathbf{v}) \geq \Delta|C_1|. \tag{11}$$

Similarly $\forall \mathbf{v} \in C_2,\ E_1(\mathbf{v}) - E_2(\mathbf{v}) < \Delta$ implies

$$\sum_{\mathbf{v}\in C_2} E_2(\mathbf{v}) - E_1(\mathbf{v}) > -\Delta|C_2|. \tag{12}$$

For the classification under consideration $|C_1| = |C_2|$, and adding (11) and (12) yields

$$\sum_{\mathbf{v}\in C_1}(E_1(\mathbf{v}) - E_2(\mathbf{v})) + \sum_{\mathbf{v}\in C_2}(E_2(\mathbf{v}) - E_1(\mathbf{v})) > 0,$$

or

$$\sum_{\sigma\in\mathcal{S}^1}\left(\sum_{\mathbf{v}\in C_1}\sigma(\mathbf{v}) - \sum_{\mathbf{v}\in C_2}\sigma(\mathbf{v})\right) - \sum_{\sigma\in\mathcal{S}^2}\left(\sum_{\mathbf{v}\in C_1}\sigma(\mathbf{v}) - \sum_{\mathbf{v}\in C_2}\sigma(\mathbf{v})\right) > 0$$

(see (1)). This provides a contradiction, since under the conditions of the Theorem both sums in the left hand side of this inequality vanish.

To demonstrate this it is enough to show that

$$\sum_{\mathbf{v}\in C_1}\sigma(\mathbf{v}) - \sum_{\mathbf{v}\in C_2}\sigma(\mathbf{v}) = 0 \tag{13}$$

for any pattern $\sigma = \{(p_i', b_i')|1 \leq i \leq k\}$ of rank $k \leq n$. Denote $\mathcal{P}_\sigma = \{p_i'|1 \leq i \leq k\}$. A vector $\mathbf{v}$ possessing σ belongs to C_1 if and only if the sums $\sum_{p\in\mathcal{P}\backslash\mathcal{P}_\sigma}\mathbf{v}_p$ and $\sum_{p\in\mathcal{P}\cap\mathcal{P}_\sigma}\mathbf{v}_p$ have the same parity. Since $|\mathcal{P}| > n$, $\mathcal{P}\backslash\mathcal{P}_\sigma \neq \emptyset$. Hence an equal number of such vectors belong to each class, which implies (13).$\triangleright$

Classifications considered in Theorem 4 have a well-defined structure, the failure to identify which is a clear drawback of the algorithm. This example calls for a careful design of the object coding.

Theorem 5. Let $\varphi : X^M \to X^M$ be a bijection of the hypercube X^M such that $\varphi(H)$ is a hyperface for any hyperface H (for any set C, $\varphi(C)$ denotes the image of C under the mapping φ), O is φ-invariant, and $|\varphi(H) \cap L_i| = |H \cap L_i|$ $(i = 1, 2)$ for any hyperface H, $\mathrm{codim} H = n$. If $\varphi(C_1) \neq C_1$ ($\Leftrightarrow C_2 \neq \varphi(C_2)$), the classification $C_1 \sqcup C_2$ cannot be recognized with the learning material $L_1 \sqcup L_2$.

Proof. Any hyperface of codimension not exceeding n is a union of non-intersecting hyperfaces of codimension n. Thus under the conditions of the Theorem $N_i(\sigma) = N_i(\sigma_{\varphi(H_\sigma)})$ $(i = 1, 2)$ for any pattern σ, $\mathrm{rank}\sigma \leq n$. Hence a pattern σ characterizes C_i if and only if $\sigma_{\varphi(H_\sigma)}$ does. Therefore any vector $\mathbf{v} \in O$ will be attributed to the same class as $\varphi(\mathbf{v})$. This contradicts the assumption $C_i \neq \varphi(C_i)$ for both classes.$\triangleright$

3. Recognition of hypercylinders.

In this Section we establish a geometric criterion for recognizability of classifications by Core-1.

Definition 7.

$$C = \left\{ \mathbf{v} \in X^M \;\middle|\; \sum_{s=1}^{m} \left(\frac{\mathbf{v}_{p_s} - \mathbf{c}_{p_s}}{a_s} \right)^2 \leq 1 \right\} \tag{14}$$

is called a *hypercylinder* with the *semiaxes* $a_s > 0$, whose *hyperaxis* passes through the vertex $\mathbf{c}$; $\{p_s | 1 \leq p_1 < ... < p_s < ... < p_m \leq M\}$ are called *basic* descriptors of the hypercylinder.

Remarks.

9. Strictly speaking, the set C defined in Definition 7 is the intersection of a hypercylinder with the hypercube X^M. However, due to the nature of the problem under consideration, the terminology of the Definition can be admitted, ignoring this inaccuracy.

10. The hyperaxis of the hypercylinder is a hyperface $\{\mathbf{v} | \mathbf{v}_{p_s} = \mathbf{c}_{p_s} \forall s, 1 \leq s \leq m\}$ of codimension m. For $m = M$ the hypercylinder degenerates into an ellipsoid and the hyperaxis consisting of a single vertex, $\mathbf{c}$, is the center of the ellipsoid. Since a hypercylinder is a finite set its hyperaxis may be non-unique (e.g., any vertex is a center of a sphere of radius $\sqrt{M}$).

11. If strict inequality is assumed in (14), the resultant set is also a hypercylinder: since both the hypercylinder and its compliment in X^M are finite, the same set would be defined by (14) with $a_i' = a_i(1 - \epsilon)$ for any small enough $\epsilon > 0$.

12. Vertices in the complement in X^M to the hypercylinder (14) satisfy the inequality

$$\sum_{s=1}^{m} \left(\frac{\mathbf{v}_{p_s} - (1 - \mathbf{c}_{p_s})}{a_s} \right)^2 < \sum_{s=1}^{m} \frac{1}{a_s^2} - 1.$$

Thus by the previous Remark the complement is a hypercylinder with the same set of basic descriptors and the semiaxes $a_s' = a_s(1 - \epsilon) \left(\sum_{q=1}^{m} \frac{1}{a_q^2} - 1 \right)^{-1/2}$, where $\epsilon > 0$ is small, and with the hyperaxis symmetrical about the center of the hypercube X^M to the hyperaxis of the original hypercylinder. Since only binary vectors are considered, an equation of a hyperplane dividing X^M into the two hypercylinders follows directly from (14).

Theorem 6.

i. A classification $O = C_1 \sqcup C_2$ is recognizable by Core-1 only if one of the classes is an intersection of the image of the objects' coding O and of a hypercylinder (and hence by Remark 12 so is the other class), and $\exists \rho \geq 0$ such that any semiaxis of the hypercylinder is equal either to ρ, or to $\sqrt{2}\rho$.

ii. If coding is surjective, any classification considered in *i* is self-recognizable by Core-n $\forall n \leq M$.

Proof.

i. Suppose a classification $O = C_1 \sqcup C_2$ can be recognized by Core-1 with some learning material. Descriptors can be categorized into 16 sets:

$$\mathcal{P}, \mathcal{P}^1, \mathcal{P}^2, \mathcal{P}^{12}, \mathcal{P}_1, \mathcal{P}_1^1, \mathcal{P}_1^2, \mathcal{P}_1^{12}, \mathcal{P}_2, \mathcal{P}_2^1, \mathcal{P}_2^2, \mathcal{P}_2^{12}, \mathcal{P}_{12}, \mathcal{P}_{12}^1, \mathcal{P}_{12}^2, \mathcal{P}_{12}^{12}.$$

A descriptor p is attributed to that set whose upper index is a list of classes characterized by the pattern $(p, 1)$ and whose lower index is a list of classes characterized by the pattern $(p, 0)$ (for maximal generality allowing patterns to characterize both classes simultaneously). Interchanging the values 0 and 1 encoding[1] descriptors from $\mathcal{P}_1 \sqcup \mathcal{P}^2$, the sets $\mathcal{P}_1, \mathcal{P}^2, \mathcal{P}_1^2, \mathcal{P}^{12}, \mathcal{P}_1^{12}, \mathcal{P}_{12}^2$ can be transformed into subsets of $\mathcal{P}^1, \mathcal{P}_2, \mathcal{P}_2^1, \mathcal{P}_{12}, \mathcal{P}_{12}^1, \mathcal{P}_2^{12}$, respectively. Thus, the list of possible sets of descriptors is reduced to

$$\mathcal{P}, \mathcal{P}^1, \mathcal{P}_1^1, \mathcal{P}_2, \mathcal{P}_2^1, \mathcal{P}_2^2, \mathcal{P}_2^{12}, \mathcal{P}_{12}, \mathcal{P}_{12}^1, \mathcal{P}_{12}^{12}.$$

For a vector $\mathbf{v}$ denote $e_\beta^\alpha = |\{p \in \mathcal{P}_\beta^\alpha | \mathbf{v}_p = 1\}|$ and $n_\beta^\alpha = |\{p \in \mathcal{P}_\beta^\alpha | \mathbf{v}_p = 0\}|$; evidently, $e_\beta^\alpha + n_\beta^\alpha = |\mathcal{P}_\beta^\alpha| \quad \forall \mathbf{v} \in X^M$. Then

$$E_1(\mathbf{v}) = e^1 + e_1^1 + n_1^1 + e_2^1 + e_2^{12} + n_{12} + e_{12}^1 + n_{12}^1 + e_{12}^{12} + n_{12}^{12} = e^1 + |\mathcal{P}_1^1| + e_2^1 + e_2^{12} + n_{12} + |\mathcal{P}_{12}^1| + |\mathcal{P}_{12}^{12}|,$$

and

$$E_2(\mathbf{v}) = n_2 + n_2^1 + e_2^2 + n_2^2 + e_2^{12} + n_2^{12} + n_{12} + n_{12}^1 + e_{12}^{12} + n_{12}^{12} = n_2 + n_2^1 + |\mathcal{P}_2^2| + |\mathcal{P}_2^{12}| + n_{12} + n_{12}^1 + |\mathcal{P}_{12}^{12}|.$$

A vector is thus attributed to C_1 if and only if

$$\begin{aligned} E_1(\mathbf{v}) - E_2(\mathbf{v}) &= e^1 + |\mathcal{P}_1^1| + e_2^1 + e_2^{12} + n_{12} + |\mathcal{P}_{12}^1| + |\mathcal{P}_{12}^{12}| - (n_2 + n_2^1 + |\mathcal{P}_2^2| + |\mathcal{P}_2^{12}| + n_{12} + n_{12}^1 + |\mathcal{P}_{12}^{12}|) \\ &= e^1 + |\mathcal{P}_1^1| + 2e_2^1 + e_2^{12} - |\mathcal{P}_2| + e_2 - |\mathcal{P}_2^1| - |\mathcal{P}_2^2| - |\mathcal{P}_2^{12}| + e_{12}^1 \geq \Delta. \end{aligned}$$

This is identical to (14) for $\mathbf{c}_p = 1, \; 1 \leq p \leq M$,

$$\rho^2 = \tfrac{1}{2}(\Delta - |\mathcal{P}_1^1| + |\mathcal{P}_2| + |\mathcal{P}_2^1| + |\mathcal{P}_2^2| + |\mathcal{P}_2^{12}|),$$

$$a_s = \begin{cases} \rho & \text{if } p_s \in \mathcal{P}_2^1, \\ \sqrt{2}\rho & \text{if } p_s \in \mathcal{P}^1 \sqcup \mathcal{P}_2^{12} \sqcup \mathcal{P}_2 \sqcup \mathcal{P}_{12}^1. \end{cases}$$

Therefore C_1 is the required intersection of the image of the objects' coding with the identified hypercylinder.

ii. Let the first class C_1 be a hypercylinder (14) whose m_1 semiaxes are $\sqrt{2}\rho$, and m_2 semiaxes are ρ. Denote by $\mathcal{P}_1$ the set of descriptors associated with the semiaxes $\sqrt{2}\rho$, by $\mathcal{P}_2$ - the set of descriptors associated with the semiaxes ρ, and by $\mathcal{P}_0$ - the remaining $m_0 = M - m_1 - m_2$ descriptors. Suppose without loss of generality that the center of the hypercylinder is $\mathbf{c}_p = 1, \; 1 \leq p \leq M$. Then $\mathbf{v} \in C_1$ if and only if

$$e_1 + 2e_2 \geq \delta, \tag{15}$$

where $e_i = |\{p \in \mathcal{P}_i | \mathbf{v}_p = 1\}|$ $(i = 1, 2)$ and $\delta = m_1 + 2m_2 - 2\rho^2$. An integer $\delta > 0$ can be assumed in (15) (the case $\delta = 0$ is trivial). Theorem 1 proves *ii* for $\delta = 1$ (C_2 is a hyperface) and for $\delta = m_1 + 2m_2$ (C_1 is a hyperface). Consequently, further arguments are provided for

$$2 \leq \delta \leq m_1 + 2m_2 - 1. \tag{16}$$

Without loss of generality it may be supposed $m_1 > 0$ (the case $m_1 = 0, \; m_2 > 0$ is equivalent to $m_1 > 0, \; m_2 = 0$ with $\sqrt{2}\rho$ regarded as a new value for the radius ρ; the case $m_1 = m_2 = 0$ is degenerate). Furthermore, for $m_1 = 1, \; m_2 = 0$, (14) represents a hyperface

[1] This is a coordinate transformation which does not change the intrinsic geometry of classes; see also Section 5.1 .

and thus Theorem 1 proves *ii*. For $m_1 = 1$, $m_2 > 0$ and an odd δ the set of integer solutions to (15) satisfying $0 \le e_i \le m_i$ $(i = 1, 2)$ coincides with the set of such integer solutions to $e_1 + e_2 \ge \frac{1}{2}(\delta + 1)$, and hence this case is equivalent to $m'_0 = m_0$, $m'_1 = m_1 + 1$, $m'_2 = 0$. Finally, for $m_1 = 1$, $m_2 > 0$ and an even δ integer solutions to (15) are $e_1 = 0$ or 1, $m_2 \ge e_2 \ge \frac{\delta}{2}$, and this case is equivalent to $m'_0 = m_0 + 1$, $m'_1 = m_2$ and $m'_2 = 0$. Therefore it is supposed below that $m_1 > 1$.

The proof of *ii* is split into 4 Lemmas.

Lemma 1. Suppose the coding is surjective, $m_1 > 1$, $m_2 > 0$ and (16) holds. Then any classification considered in part *i* of the Theorem is self-recognizable by Core-1.

Proof. Let $N_i(\mathcal{P}, b)$ denote the number of vectors $\mathbf{v} \in C_i$ possessing a pattern of rank one (p, b) for some $p \in \mathcal{P} \in \{\mathcal{P}_0, \mathcal{P}_1, \mathcal{P}_2\}$ (the number is independent of the specific $p \in \mathcal{P}$). Then

$$N_i(\mathcal{P}_0, 0) = N_i(\mathcal{P}_0, 1) = \tfrac{1}{2}|C_i| \quad \forall i = 1, 2$$

(if $m_0 = 0$ assume $N_i(\emptyset, 0) = N_i(\emptyset, 1) = \frac{1}{2}|C_i|$ for $i = 1, 2$ by definition),

$$N_1(\mathcal{P}_1, 0) = 2^{m_0} \sum_{e_1 \ge 0, e_2 \ge 0:\ e_1 + 2e_2 \ge \delta} \binom{m_1 - 1}{e_1} \binom{m_2}{e_2},$$

$$N_1(\mathcal{P}_1, 1) = 2^{m_0} \sum_{e_1 \ge 0, e_2 \ge 0:\ e_1 + 2e_2 \ge \delta - 1} \binom{m_1 - 1}{e_1} \binom{m_2}{e_2},$$

$$N_1(\mathcal{P}_2, 0) = 2^{m_0} \sum_{e_1 \ge 0, e_2 \ge 0:\ e_1 + 2e_2 \ge \delta} \binom{m_1}{e_1} \binom{m_2 - 1}{e_2},$$

$$N_1(\mathcal{P}_2, 1) = 2^{m_0} \sum_{e_1 \ge 0, e_2 \ge 0:\ e_1 + 2e_2 \ge \delta - 2} \binom{m_1}{e_1} \binom{m_2 - 1}{e_2}.$$

From the above combinatorial results it follows that

$$N_1(\mathcal{P}_1, 0) < N_1(\mathcal{P}_1, 1), \quad N_1(\mathcal{P}_2, 0) < N_1(\mathcal{P}_2, 1)$$

(while all positive terms in the sum representing $N_1(\mathcal{P}_1, 0)$ are present in the sum representing $N_1(\mathcal{P}_1, 1)$, the latter sum contains extra positive terms corresponding to $e_1 + 2e_2 = \delta - 1$, e.g. $e_1 = \min(\delta - 1, m_1 - 1 - \mathrm{mod}_2(\delta - m_1))$, $e_2 = \frac{1}{2}(\delta - 1 - e_1)$; similarly, all positive terms in the sum representing $N_1(\mathcal{P}_2, 0)$ are present in the sum representing $N_1(\mathcal{P}_2, 1)$, and the latter contains extra positive terms corresponding to $\delta - 2 \le e_1 + 2e_2 = \delta - 1$, e.g. $e_1 = \min(\delta - 2, m_1 - \mathrm{mod}_2(\delta - m_1))$, $e_2 = \frac{1}{2}(\delta - 2 - e_1)$). Since

$$N_2(\mathcal{P}_i, 0) = 2^{M-1} - N_1(\mathcal{P}_i, 0), \quad N_2(\mathcal{P}_i, 1) = 2^{M-1} - N_1(\mathcal{P}_i, 1) \quad \forall i = 0, 1, 2$$

this implies

$$N_2(\mathcal{P}_1, 0) > N_2(\mathcal{P}_1, 1), \quad N_2(\mathcal{P}_2, 0) > N_2(\mathcal{P}_2, 1).$$

Therefore, since

$$N_i(\mathcal{P}_1, 0) = |C_i| - N_i(\mathcal{P}_1, 1), \quad N_i(\mathcal{P}_2, 0) = |C_i| - N_i(\mathcal{P}_2, 1) \quad \forall i = 1, 2$$

one finds

$$N_1(\mathcal{P}_1, 1) > N_1(\mathcal{P}_0, 1) = N_1(\mathcal{P}_0, 0) > N_1(\mathcal{P}_1, 0),$$
$$N_1(\mathcal{P}_2, 1) > N_1(\mathcal{P}_0, 1) = N_1(\mathcal{P}_0, 0) > N_1(\mathcal{P}_2, 0),$$

and

$$N_2(\mathcal{P}_1, 0) > N_2(\mathcal{P}_0, 0) = N_2(\mathcal{P}_0, 1) > N_2(\mathcal{P}_1, 1),$$
$$N_2(\mathcal{P}_2, 0) > N_2(\mathcal{P}_0, 0) = N_2(\mathcal{P}_0, 1) > N_2(\mathcal{P}_2, 1).$$

Adding

$$\sum_{e_1\geq 0, e_2\geq 0:\ e_1+2e_2\geq\delta-2} \frac{1}{e_1!(m_1-e_1-1)!}\frac{1}{e_2!(m_2-e_2-1)!}$$
$$> \sum_{e_1\geq 0, e_2\geq 0:\ e_1+2e_2\geq\delta-1} \frac{1}{e_1!(m_1-e_1-1)!}\frac{1}{e_2!(m_2-e_2-1)!}$$

(in consequence of (15) the sum at the left hand side of this inequality contains extra positive terms, as above) with

$$\sum_{e_1\geq 0, e_2\geq 0:\ e_1+2e_2\geq\delta-2} \frac{1}{(e_1-1)!(m_1-e_1)!}\frac{1}{e_2!(m_2-e_2-1)!}$$
$$= \sum_{e_1\geq 0, e_2\geq 0:\ e_1+2e_2\geq\delta-1} \frac{1}{e_1!(m_1-1-e_1)!}\frac{1}{(e_2-1)!(m_2-e_2)!}$$

(to demonstrate this identity use a new index of summation $e_2' = e_2+1$ at the left hand side, and $e_1' = e_1+1$ at the right hand side; as usual, $e! = \infty$ for $e<0$) yields

$$\sum_{e_1\geq 0, e_2\geq 0:\ e_1+2e_2\geq\delta-2} \frac{m_1}{e_1!(m_1-e_1)!e_2!(m_2-e_2-1)!}$$
$$> \sum_{e_1\geq 0, e_2\geq 0:\ e_1+2e_2\geq\delta-1} \frac{m_2}{e_1!(m_1-1-e_1)!e_2!(m_2-e_2)!}.$$

Multiplication of this inequality by $(m_1-1)!(m_2-1)!$ establishes $N_1(\mathcal{P}_2,1) > N_1(\mathcal{P}_1,1)$, finally leading to

$$N_1(\mathcal{P}_2,1) > N_1(\mathcal{P}_1,1) > N_1(\mathcal{P}_0,1) = N_1(\mathcal{P}_0,0) > N_1(\mathcal{P}_1,0) > N_1(\mathcal{P}_2,0) \tag{17.1}$$

and

$$N_2(\mathcal{P}_2,0) > N_2(\mathcal{P}_1,0) > N_2(\mathcal{P}_0,0) = N_2(\mathcal{P}_0,1) > N_2(\mathcal{P}_1,1) > N_2(\mathcal{P}_2,1). \tag{17.2}$$

In view of (17.1) for $A_1 = N_1(\mathcal{P}_1,1)$, $R_1=\infty$ a pattern (p,b) characterizes the first class if and only if $p\in\mathcal{P}_1\sqcup\mathcal{P}_2$ and $b=1$. Similarly, due to (17.2) for $A_2 = N_2(\mathcal{P}_2,0)$, $R_2=\infty$ the second class will be characterized by a pattern (p,b) if and only if $p\in\mathcal{P}_2$ and $b=0$. For a vector $\mathbf{v}$ denote $n_i = |\{p\in\mathcal{P}_i|\mathbf{v}_p=0\}|$, $i=1,2$. For the chosen free parameters a vector is thus attributed to C_1 if and only if

$$E_1(\mathbf{v}) - E_2(\mathbf{v}) = e_1+e_2-n_2 = e_1+2e_2-m_2 \geq \Delta,$$

$\Delta = \delta - m_2$ thus guaranteeing a correct recognition by Core-1.$\triangleright$

Remark 13. C_2 consists of vectors for which $n_1+2n_2\geq\delta' = m_1+2m_2-\delta+1$ (see Remark 12) and (15) implies $2\leq\delta'\leq m_1+2m_2-1$. Thus due to the interchangeability of classes under the conditions of Lemma 1 for $A_1=N_1(\mathcal{P}_2,1)$, $R_1=\infty$, $A_2=N_2(\mathcal{P}_1,0)$, $R_2=\infty$ a vector $\mathbf{v}$ is assigned to C_2 if and only if $E_1(\mathbf{v})-E_2(\mathbf{v}) = e_2-n_1-n_2 = m_1-n_1-2n_2 < \Delta$, and $\Delta=\delta-2m_2$ again furnishes a correct recognition by Core-1.

Lemma 2. Let the coding be surjective, $m_1>1$, $m_2>0$ and

$$2<\delta<m_1+2m_2-1. \tag{18}$$

Then any classification considered in part i of the Theorem is self-recognizable by Core-n $\forall n>1$.

Proof. Denote by $N_i(\mathcal{P}_{n_1},\mathcal{P}_{n_2},b_1,b_2)$ the number of vectors possessing a pattern of rank two $\{(p_1,b_1),(p_2,b_2)\}$, where $p_i\in\mathcal{P}_{n_i}$, $0\leq n_i\leq 2$, $i=1,2$ and $p_1\neq p_2$. Obviously $\forall b_1,b_2\in\{0,1\}$

$$N_1(\mathcal{P}_0,\mathcal{P}_i,b_1,b_2)\leq N_1(\mathcal{P}_0,b_1)\quad (0\leq i\leq 2,\ \text{or } i=1,2 \text{ if } m_0=1), \tag{19.1}$$

$$N_1(\mathcal{P}_i, \mathcal{P}_j, 0, b_2) \le N_1(\mathcal{P}_i, 0) \quad (i = 1, 2 \text{ and } j = 1, 2), \tag{19.2}$$

$$N_1(\mathcal{P}_1, \mathcal{P}_1, 1, 1) < N_1(\mathcal{P}_1, 1) \tag{19.3}$$

((19.3) could be violated if $e_1 + 2e_2 \ge \delta$ implied $e_1 = m_1$; this occurs only for $\delta = m_1 + 2m_2$, incompatible with (18)),

$$N_1(\mathcal{P}_1, \mathcal{P}_2, 1, 1) < N_1(\mathcal{P}_1, 1) \tag{19.4}$$

((19.4) could be violated if $e_1 + 2e_2 \ge \delta$ implied $e_2 = m_2$; this occurs only for $\delta \ge m_1 + 2m_2 - 1$, incompatible with (18)).

Two possibilities arise.

1. $m_2 = 1$, or

$$\sum_{e_1 \ge 0, e_2 \ge 0:\ e_1 + 2e_2 \ge \delta - 4} \binom{m_1}{e_1}\binom{m_2 - 2}{e_2} < \sum_{e_1 \ge 0, e_2 \ge 0:\ e_1 + 2e_2 \ge \delta - 1} \binom{m_1 - 1}{e_1}\binom{m_2}{e_2},$$

which is equivalent to

$$N_1(\mathcal{P}_2, \mathcal{P}_2, 1, 1) < N_1(\mathcal{P}_1, 1). \tag{19.5}$$

In view of (17.1), the five inequalities (19) show that any pattern of rank 2 or higher is possessed by strictly less than $N_1(\mathcal{P}_1, 1)$ vectors from C_1.

Similarly, unless $i = j = 2$ and $b_1 = b_2 = 0$ by (17.2)

$$N_2(\mathcal{P}_i, \mathcal{P}_j, b_1, b_2) < N_2(\mathcal{P}_2, 0) \quad \forall b_1, b_2 \in \{0, 1\} \text{ and } 0 \le i, j \le 2.$$

If $m_2 > 1$, the inequality $N_2(\mathcal{P}_2, \mathcal{P}_2, 0, 0) < N_2(\mathcal{P}_2, 0)$ could be violated if $e_1 + 2e_2 < \delta$ implied $e_2 = 0$; this occurs only for $\delta < 2$, incompatible with (18). Hence any pattern of rank 2 or higher is possessed by strictly less than $N_2(\mathcal{P}_2, 0)$ vectors from C_2.

Therefore by Theorem 3 classifications are self-recognizable by Core-n $\forall n > 1$ for this possibility with A_1, R_1, A_2, R_2 and Δ the same as in Lemma 1.

2. $m_2 > 1$ and

$$\sum_{e_1 \ge 0, e_2 \ge 0:\ e_1 + 2e_2 \ge \delta - 4} \binom{m_1}{e_1}\binom{m_2 - 2}{e_2} \ge \sum_{e_1 \ge 0, e_2 \ge 0:\ e_1 + 2e_2 \ge \delta - 1} \binom{m_1 - 1}{e_1}\binom{m_2}{e_2}. \tag{20}$$

In this case the role of the two classes is reversed in the proof.

C_2 consists of vectors for which $n_1 + 2n_2 \ge \delta' = m_1 + 2m_2 - \delta + 1$, and (18) implies $2 < \delta' < m_1 + 2m_2 - 1$. By the same arguments as above,

$$N_2(\mathcal{P}_0, \mathcal{P}_i, b_1, b_2) \le N_2(\mathcal{P}_0, b_1) \quad (0 \le i \le 2, \text{ or } i = 1, 2 \text{ if } m_0 = 1), \tag{21.1}$$

$$N_2(\mathcal{P}_i, \mathcal{P}_j, 1, b_2) \le N_2(\mathcal{P}_i, 1) \quad (i = 1, 2 \text{ and } j = 1, 2), \tag{21.2}$$

$$N_2(\mathcal{P}_1, \mathcal{P}_i, 0, 0) < N_2(\mathcal{P}_1, 0) \quad \forall i = 1, 2. \tag{21.3}$$

Also,

$$\sum_{n_1 \ge 0, n_2 \ge 0:\ n_1 + 2n_2 \ge \delta' - 4} \binom{m_1}{n_1}\binom{m_2 - 2}{n_2}$$

$$= 2^{m_1 + m_2 - 2} - \sum_{n_1 \ge 0, n_2 \ge 0:\ n_1 + 2n_2 < \delta' - 4} \binom{m_1}{n_1}\binom{m_2 - 2}{n_2}$$

(employing new indices of summation $e_1 = m_1 - n_1,\ e_2 = m_2 - 2 - n_2$)

$$= 2^{m_1 + m_2 - 2} - \sum_{e_1 \ge 0, e_2 \ge 0:\ e_1 + 2e_2 \ge \delta} \binom{m_1}{e_1}\binom{m_2 - 2}{e_2}$$

$$= 2^{m_1 + m_2 - 2} - \sum_{e_1 \ge 0, e_2 \ge 0:\ e_1 + 2e_2 \ge \delta - 4} \binom{m_1}{e_1}\binom{m_2 - 2}{e_2}$$

$$
\begin{aligned}
&+ \sum_{e_1\geq 0, e_2\geq 0:\ \delta-4\leq e_1+2e_2<\delta} \binom{m_1}{e_1}\binom{m_2-2}{e_2}\\
&\leq 2^{m_1+m_2-1} - \sum_{e_1\geq 0, e_2\geq 0:\ e_1+2e_2\geq\delta-4} \binom{m_1}{e_1}\binom{m_2-2}{e_2}\\
&\leq 2^{m_1+m_2-1} - \sum_{e_1\geq 0, e_2\geq 0:\ e_1+2e_2\geq\delta-1} \binom{m_1-1}{e_1}\binom{m_2}{e_2} \qquad \text{(by (20))}\\
&= \sum_{e_1\geq 0, e_2\geq 0:\ e_1+2e_2<\delta-1} \binom{m_1-1}{e_1}\binom{m_2}{e_2}
\end{aligned}
$$

(substituting new indices of summation $n_1 = m_1 - 1 - e_1,\ n_2 = m_2 - e_2$)

$$
= \sum_{n_1\geq 0, n_2\geq 0:\ n_1+2n_2>\delta'-1} \binom{m_1-1}{n_1}\binom{m_2}{n_2} < \sum_{n_1\geq 0, n_2\geq 0:\ n_1+2n_2\geq\delta'-1} \binom{m_1-1}{n_1}\binom{m_2}{n_2},
$$

which is equivalent to

$$
N_2(\mathcal{P}_2, \mathcal{P}_2, 0, 0) < N_2(\mathcal{P}_1, 0). \tag{21.4}
$$

Together with (17.2) inequalities (21) imply that any pattern of rank 2 or higher is possessed by strictly less than $N_2(\mathcal{P}_1, 0)$ vectors from C_2. As before, any pattern of rank 2 or higher is possessed by strictly less than $N_1(\mathcal{P}_2, 1)$ vectors from C_1. Hence by Theorem 3 any classification is again self-recognizable by Core-n $\forall n > 1$ for this possibility with A_1, R_1, A_2, R_2 and Δ the same as in Remark 13.$\triangleright$

Lemma 3. Suppose the coding is surjective, $m_1 > 1,\ m_2 > 0$ and $\delta = 2$. Then any classification satisfying (15) is self-recognizable by Core-n $\forall n > 1$.

Proof. Set $A_2 = \infty, R_2 = 0$; hence no patterns characterize C_2. For $\delta = 2$, $\mathbf{v} \in C_2$ if and only if $\mathbf{v}_p = 0\ \forall p \in \mathcal{P}_2$ and $\mathbf{v}_p = 1$ for at most one $p \in \mathcal{P}_1$ (see (15)). Hence for $A_1 = R_1 = 0$ a pattern $\{(p_i, b_i) | 1 \leq i \leq k\}$ of rank k characterizes C_1 if and only if at least one of the following conditions is satisfied: either $\exists p_i \in \mathcal{P}_2$ such that $b_i = 1$, or $\exists p_i, p_j \in \mathcal{P}_1$ such that $p_i \neq p_j$ and $b_i = b_j = 1$. Since $n \geq 2$, any $\mathbf{v} \in C_1$ thus possesses at least one pattern characterizing C_1, and $\Delta = 1$ furnishes the desired recognition.$\triangleright$

Clearly, the case $m_2 > 0$ and $\delta = m_1 + 2m_2 - 1$ can be reduced to that of Lemma 3 by interchanging the classes C_1 and C_2.

Lemma 4. Suppose the coding is surjective, $m_2 = 0$ and

$$
2 \leq \delta \leq m_1 - 1. \tag{22}
$$

Then any classification considered in part i of the Theorem is self-recognizable by Core-n $\forall n \geq 1$.

Proof. The class C_1 is comprised of vectors, at most $m_1 - \delta$ of whose components from $\mathcal{P}_1$ are zeroes. Hence, using (22),

$$
N_1(\mathcal{P}_1, 1) = 2^{m_0} \sum_{n=0}^{m_1-\delta} \binom{m_1-1}{n} > 2^{m_0} \sum_{n=0}^{m_1-\delta-1} \binom{m_1-1}{n} = N_1(\mathcal{P}_1, 0).
$$

Since $N_1(\mathcal{P}_1, 1) + N_1(\mathcal{P}_1, 0) = |C_1|$ and $N_1(\mathcal{P}_0, b) = \frac{1}{2}|C_1|$ for any $b = 0, 1$, also

$$
N_1(\mathcal{P}_1, 1) > N_1(\mathcal{P}_0, 1) = N_1(\mathcal{P}_0, 0) > N_1(\mathcal{P}_1, 0). \tag{23}
$$

Hence for $A_1 = N_1(\mathcal{P}_1, 1)$, $R_1 = \infty$ any pattern of rank 1 characterizing the first class has the form $(p, 1),\ p \in \mathcal{P}_1$.

Consider a recognition of the classification by Core-1 assuming $A_2 = \infty,\ R_2 = 0$, whereby no patterns characterize C_2. Since $\mathbf{v} \in C_1 \Leftrightarrow e_1 \geq \delta$, Core-1 recognizes the desired classification with $\Delta = \delta$.

Evidently, $\forall b_1, b_2 \in \{0,1\}$

$$N_1(\mathcal{P}_0, \mathcal{P}_i, b_1, b_2) \le N_1(\mathcal{P}_0, b_1) \quad (i = 0, 1 \text{ or } i = 1 \text{ if } m_0 = 1),$$

$$N_1(\mathcal{P}_1, \mathcal{P}_i, 0, b_2) \le N_1(\mathcal{P}_1, 0) \quad (i = 0, 1)$$

and due to (22)

$$N_1(\mathcal{P}_1, \mathcal{P}_1, 1, 1) < N_1(\mathcal{P}_1, 1).$$

Thus, by (23)

$$N_1(\mathcal{P}_i, \mathcal{P}_j, b_1, b_2) < N_1(\mathcal{P}_0, 1) \quad \forall b_1, b_2 \in \{0,1\},\ i = 0, 1,\ j = 0, 1$$

and by Theorem 3 the classification under consideration is self-recognizable by Core-n $\forall n > 1$ with the same free parameters as for Core-1.▷

Remark 14. Since C_1 and C_2 considered in Lemma 4 are interchangeable (see Remark 12), the same classification can be recognized by Core-n $\forall n \ge 1$ with $A_1 = \infty$, $R_1 = 0$, $A_2 = N_2(\mathcal{P}_1, 0)$, $R_2 = \infty$ and $\Delta = m_1 - \delta$.

The proof of the Theorem is concluded.▷

4. A necessary condition for recognizability of a classification by Core-2.

If a classification $O = C_1 \sqcup C_2$ is recognizable with some learning material by Core-n, C_1 is comprised of $\mathbf{v} \in O$ such that

$$\sum_{k=1}^{n} \sum_{\{(p_i,b_i)|1\le i\le k\}} \epsilon^{b_1,\ldots,b_k}_{p_1,\ldots,p_k} \sigma^{b_1,\ldots,b_k}_{p_1,\ldots,p_k}(\mathbf{v}) \ge \Delta. \tag{24}$$

Here the internal sum is over all combinations of k binary values $b_i \in \{0,1\}$ and k positions $1 \le p_1 < \ldots < p_k \le M$; $\sigma^{b_1,\ldots,b_k}_{p_1,\ldots,p_k}(\mathbf{v})$ is the function associated with the pattern $\{(p_i,b_i)|1 \le i \le k\}$; $\epsilon^{b_1,\ldots,b_k}_{p_1,\ldots,p_k} = 1$ if the pattern $\{(p_i,b_i)|1 \le i \le k\}$ characterizes only the first class, $\epsilon^{b_1,\ldots,b_k}_{p_1,\ldots,p_k} = -1$ if the pattern $\{(p_i,b_i)|1 \le i \le k\}$ characterizes only the second class and $\epsilon^{b_1,\ldots,b_k}_{p_1,\ldots,p_k} = 0$ otherwise. In principle, in different problems the coefficients can assume any combination of these values.

Since

$$\sum_{\{b_i\in\{0,1\}|1\le i\le k\}} \sigma^{b_1,\ldots,b_k}_{p_1,\ldots,p_k}(\mathbf{v}) = 1 \quad \forall \mathbf{v} \in X^M \tag{25}$$

(here the sum is over all combinations of k binary values b_i) for any set of $k \ge 1$ distinct positions p_i, (24) is equivalent to

$$\sum_{k=1}^{n} \sum_{\{(p_i,b_i)|1\le i\le k\}} \bar{\epsilon}^{b_1,\ldots,b_k}_{p_1,\ldots,p_k} \sigma^{b_1,\ldots,b_k}_{p_1,\ldots,p_k}(\mathbf{v}) \ge \Delta' \tag{26}$$

where $0 \le \bar{\epsilon}^{b_1,\ldots,b_k}_{p_1,\ldots,p_k} \le 2$. ((26) is a generalization of (15) for an arbitrary n.) Using

$$\sigma^{b_1,\ldots,b_{k-1},0}_{p_1,\ldots,p_{k-1},p_k}(\mathbf{v}) = \sigma^{b_1,\ldots,b_{k-1}}_{p_1,\ldots,p_{k-1}}(\mathbf{v}) - \sigma^{b_1,\ldots,b_{k-1},1}_{p_1,\ldots,p_{k-1},p_k}(\mathbf{v}) \quad \forall \mathbf{v} \in X^M \tag{27}$$

(valid for any $k > 1$ distinct positions p_i and any $k-1$ binary values $b_i \in \{0,1\}$) and (25) for $k = 1$, (26) can be rewritten as

$$\sum_{k=1}^{n} \sum_{\{p_i|1\le i\le k\}} \hat{\epsilon}_{p_1,\ldots,p_k} \sigma^{1,\ldots,1}_{p_1,\ldots,p_k}(\mathbf{v}) \ge \delta, \tag{28}$$

where the constants $\hat{\epsilon}_{p_1,\ldots,p_k}$ and δ can be expressed in $\epsilon^{b_1,\ldots,b_k}_{p_1,\ldots,p_k}$ and Δ. The identity

$$\sigma^{1,\ldots,1}_{p_1,\ldots,p_k}(\mathbf{v}) = \prod_{i=1}^{k} \mathbf{v}_{p_i} \tag{29}$$

transforms (28) into

$$\sum_{k=1}^{n} \sum_{\{p_i|1\le i\le k\}} \left(\hat{\epsilon}_{p_1,\ldots,p_k} \prod_{i=1}^{k} \mathbf{v}_{p_i}\right) \ge \delta. \tag{30}$$

The purpose of this Section is to explore, for $n = 2$, which combinations of coefficients $\hat{\epsilon}_{p_1,\ldots,p_k}$ are *acceptable*, i.e. for which coefficients (30) is equivalent to the original inequality (24) for some integers $\epsilon^{b_1,\ldots,b_k}_{p_1,\ldots,p_k}$, $-1 \le \epsilon^{b_1,\ldots,b_k}_{p_1,\ldots,p_k} \le 1$.

For $n = 2$, (24) is

$$\sum_{p_1=1}^{M-1} \sum_{p_2=p_1+1}^{M} (\epsilon^{0,0}_{p_1,p_2}\sigma^{0,0}_{p_1,p_2}(\mathbf{v}) + \epsilon^{0,1}_{p_1,p_2}\sigma^{0,1}_{p_1,p_2}(\mathbf{v}) + \epsilon^{1,0}_{p_1,p_2}\sigma^{1,0}_{p_1,p_2}(\mathbf{v}) + \epsilon^{1,1}_{p_1,p_2}\sigma^{1,1}_{p_1,p_2}(\mathbf{v}))$$

$$+\sum_{p=1}^{M}(\epsilon^0_p\sigma^0_p(\mathbf{v}) + \epsilon^1_p\sigma^0_p(\mathbf{v})) \ge \Delta.$$

Rearranging the left hand side of this inequality using (27) for $k = 2$, (25) for $k = 1$ and (29), yields

$$\sum_{p=1}^{M}\left(\epsilon^1_p - \epsilon^0_p + \sum_{p_2=1}^{p-1}(\epsilon^{0,1}_{p_2,p} - \epsilon^{0,0}_{p_2,p}) + \sum_{p_2=p+1}^{M}(\epsilon^{1,0}_{p,p_2} - \epsilon^{0,0}_{p,p_2})\right)\mathbf{v}_p$$

$$+\sum_{p_1=1}^{M-1}\sum_{p_2=p_1+1}^{M}(\epsilon^{0,0}_{p_1,p_2} - \epsilon^{0,1}_{p_1,p_2} - \epsilon^{1,0}_{p_1,p_2} + \epsilon^{1,1}_{p_1,p_2})\mathbf{v}_{p_1}\mathbf{v}_{p_2} \ge \Delta - \sum_{p=1}^{M}\epsilon^1_p - \sum_{p_1=1}^{M-1}\sum_{p_2=p_1+1}^{M}\epsilon^{0,0}_{p_1,p_2}, \tag{31}$$

matching the form

$$\sum_{p=1}^{M}\epsilon_p\mathbf{v}_p + \sum_{p_1=1}^{M-1}\sum_{p_2=p_1+1}^{M}\epsilon_{p_1,p_2}\mathbf{v}_{p_1}\mathbf{v}_{p_2} \ge \delta \tag{32}$$

which (30) admits for $n = 2$. Comparison of (31) and (32) shows that the coefficients of (31) are acceptable if they can be represented as

$$\epsilon_p = \epsilon^1_p - \epsilon^0_p + \sum_{p_2=1}^{p-1}(\epsilon^{0,1}_{p_2,p} - \epsilon^{0,0}_{p_2,p}) + \sum_{p_2=p+1}^{M}(\epsilon^{1,0}_{p,p_2} - \epsilon^{0,0}_{p,p_2}) \tag{33}$$

and

$$\epsilon_{p_1,p_2} = \epsilon^{0,0}_{p_1,p_2} - \epsilon^{0,1}_{p_1,p_2} - \epsilon^{1,0}_{p_1,p_2} + \epsilon^{1,1}_{p_1,p_2} \quad (p_1 < p_2) \tag{34}$$

for a set of allowed coefficients

$$-1 \le \epsilon^b_p \le 1, \quad -1 \le \epsilon^{b_1,b_2}_{p_1,p_2} \le 1 \quad (p_1 < p_2). \tag{35}$$

Denote

$$\tau_{p_1,p_2} = \begin{cases} \epsilon^{1,0}_{p_1,p_2} - \epsilon^{0,0}_{p_1,p_2} & \text{if } p_1 < p_2, \quad (36.1)\\ 0 & \text{if } p_1 = p_2, \quad (36.2)\\ \epsilon^{0,1}_{p_2,p_1} - \epsilon^{0,0}_{p_2,p_1} & \text{if } p_1 > p_2. \quad (36.3)\end{cases}$$

In the new variables

$$\epsilon_p = \epsilon^1_p - \epsilon^0_p + \sum_{p_2=1}^{M}\tau_{p,p_2}. \tag{37}$$

Thus the coefficients in (32), ϵ_p and ϵ_{p_1,p_2}, are acceptable if they are provided by (34), (36) and (37) for a set of integer $\epsilon^{b_1,\ldots,b_k}_{p_1,\ldots,p_k}$ satisfying (35). The values of $\epsilon^{1,0}_{p_1,p_2}$, $\epsilon^{0,1}_{p_1,p_2}$ and $\epsilon^{1,1}_{p_2,p_1}$ obtained from (34), (36.1) and (36.3) would be valid if and only if

$$-1 \le \epsilon^{1,0}_{p_1,p_2} = \tau_{p_1,p_2} + \epsilon^{0,0}_{p_1,p_2} \le 1, \tag{38.1}$$

$$-1 \le \epsilon^{0,1}_{p_1,p_2} = \tau_{p_2,p_1} + \epsilon^{0,0}_{p_1,p_2} \le 1, \tag{38.2}$$

$$-1 \le \epsilon^{1,1}_{p_1,p_2} = \epsilon^{0,0}_{p_1,p_2} + \tau_{p_1,p_2} + \tau_{p_2,p_1} + \epsilon_{p_1,p_2} \le 1. \tag{38.3}$$

The inequalities (38) are consistent and provide for an allowed $\epsilon^{0,0}_{p_1,p_2}$ if and only if

$$\begin{aligned} &\max(-1, -1 - \tau_{p_1,p_2}, -1 - \tau_{p_2,p_1}, -1 - \epsilon_{p_1,p_2} - \tau_{p_1,p_2} - \tau_{p_2,p_1}) \\ &\le \min(1, 1 - \tau_{p_1,p_2}, 1 - \tau_{p_2,p_1}, 1 - \epsilon_{p_1,p_2} - \tau_{p_1,p_2} - \tau_{p_2,p_1}), \end{aligned} \tag{39}$$

where $p_1 < p_2$. Therefore, a set of coefficients ϵ_p and ϵ_{p_1,p_2} is acceptable if and only if the ϵ_p are represented by (37), where the ϵ^b_p satisfy (35) and integers τ_{p_1,p_2} satisfy (36.2) and (39).

The inequality (39) represents a system of 12 individual non-trivial inequalities equivalent to the following two:

$$\max(-2, -2 - \epsilon_{p_1,p_2}) \le \tau_{p_1,p_2} \le \min(2, 2 - \epsilon_{p_1,p_2}), \tag{40.1}$$

$$\begin{aligned} &\max(-2, -2 - \epsilon_{p_1,p_2}, -2 + \tau_{p_1,p_2}, -2 - \epsilon_{p_1,p_2} - \tau_{p_1,p_2}) \\ &\le \tau_{p_2,p_1} \le \min(2, 2 - \epsilon_{p_1,p_2}, 2 + \tau_{p_1,p_2}, 2 - \epsilon_{p_1,p_2} - \tau_{p_1,p_2}), \end{aligned} \tag{40.2}$$

where $p_1 < p_2$. In new variables

$$\xi_{p_1,p_2} = \begin{cases} \tau_{p_1,p_2} - \frac{1}{2}\epsilon_{p_1,p_2} & \text{if } p_1 < p_2, \\ 0 & \text{if } p_1 = p_2, \\ \tau_{p_1,p_2} - \frac{1}{2}\epsilon_{p_2,p_1} & \text{if } p_1 > p_2 \end{cases}$$

(40) reduces to an equivalent form

$$|\xi_{p_1,p_2}| \le 2 - \tfrac{1}{2}|\epsilon_{p_1,p_2}|, \tag{41.1}$$

$$|\xi_{p_2,p_1}| \le 2 - \tfrac{1}{2}|\epsilon_{p_1,p_2}|, \tag{41.2}$$

$$|\xi_{p_1,p_2}| + |\xi_{p_2,p_1}| \le 2, \tag{41.3}$$

where $p_1 < p_2$. (41.1) and (41.2) imply

$$|\epsilon_{p_1,p_2}| \le 4 \tag{42}$$

(cf. (34)). Hence a set of coefficients ϵ_p and ϵ_{p_1,p_2} is acceptable if and only if (42) holds and (expressing (39) in ξ)

$$\epsilon_p = \epsilon^1_p - \epsilon^0_p - \tfrac{1}{2}\left(\sum_{p_2=1}^{p-1} \epsilon_{p_2,p} + \sum_{p_2=p+1}^{M} \epsilon_{p,p_2}\right) + \sum_{p_2=1}^{M} \xi_{p,p_2}, \tag{43}$$

where ξ_{p_1,p_2} and ξ_{p_2,p_1} are any pairs of numbers satisfying (41), which are simultaneously integers or half-integers depending on whether ϵ_{p_1,p_2} is even or odd.

For $|\epsilon_{p_1,p_2}| \ge 2$, (41.3) follows from (41.1) and (41.2); for $\epsilon_{p_1,p_2} = 0$, (41.1) and (41.2) follow from (41.3). If $|\epsilon_{p_1,p_2}| = 1$, any half-integer solution to (41) can be represented as $\xi_{p_1,p_2} = \mu_{p_1,p_2} + \nu_{p_1,p_2}$ and $\xi_{p_2,p_1} = \mu_{p_2,p_1} + \nu_{p_2,p_1}$, where $|\mu_{p_1,p_2}| = |\mu_{p_2,p_1}| = \frac{1}{2}$ and their signs are arbitrary, ν_{p_1,p_2} and ν_{p_2,p_1} are integer and $|\nu_{p_1,p_2}| + |\nu_{p_2,p_1}| \le 1$.

In view of (35) and of the partial independence of ξ_{p_1,p_2} and ξ_{p_2,p_1} discussed in the previous paragraph, it follows from (43) that a set of coefficients of (32) is acceptable if and only if (42) holds and

$$\left|\epsilon_p + \tfrac{1}{2}\left(\sum_{p_2=1}^{p-1} \epsilon_{p_2,p} + \sum_{p_2=p+1}^{M} \epsilon_{p,p_2}\right)\right|$$

$$\leq 2 + \tfrac{1}{2}\left(\sum_{p_2=1}^{p-1} \min(4 - |\epsilon_{p_2,p}|, |\epsilon_{p_2,p}|) + \sum_{p_2=p+1}^{M} \min(4 - |\epsilon_{p,p_2}|, |\epsilon_{p,p_2}|)\right) + \sum_{p_2=1}^{M} \nu_{p,p_2}. \quad (44)$$

Here ν_{p_1,p_2} are integers satisfying (assuming $p_1 < p_2$)

$$\nu_{p_1,p_2} \geq 0, \quad \nu_{p_2,p_1} \geq 0, \quad \nu_{p_1,p_2} + \nu_{p_2,p_1} \leq \zeta_{p_1,p_2}, \quad (45)$$

where

$$\zeta_{p_1,p_2} = \begin{cases} 2 & \text{if } \epsilon_{p_1,p_2} = 0, \\ 1 & \text{if } |\epsilon_{p_1,p_2}| = 1, \\ 0 & \text{if } |\epsilon_{p_1,p_2}| \geq 2 \text{ or } p_1 = p_2. \end{cases} \quad (46)$$

Lemma 5. Integers $\Gamma_p \geq 0,\ 1 \leq p \leq M$ admit a representation

$$\Gamma_p = \sum_{p_2=1}^{M} \nu_{p,p_2}, \quad (47)$$

where ν_{p_1,p_2} are integers satisfying (45), if and only if

$$\sum_{p=1}^{M} \Gamma_p \mathbf{v}_p \leq \sum_{p_1=1}^{M-1} \sum_{p_2=p_1+1}^{M} \min(1, \mathbf{v}_{p_1} + \mathbf{v}_{p_2})\zeta_{p_1,p_2} \quad \forall \mathbf{v} \in X^M \ (\mathbf{v} \neq \mathbf{0}). \quad (48)$$

Proof. The inequalities (48) are an evident consequence of (45) and (47).

The problem of representation of the integers Γ_p will be iteratively reduced in such a way that at each step $\sum_{p=1}^{M} \Gamma_p$ and/or the dimension M strictly decrease. Since in the cases $\Gamma_p = 0 \ \ \forall p$ or $M \leq 2$ the problem is trivial, this procedure provides a complete demonstration of the Lemma.

Consider two possibilities.

i. The inequalities (48) are strict $\forall \mathbf{v} \neq \mathbf{0}$, except maybe for $\mathbf{v}_p = 1\ \forall p$. Unless the reductions have saturated to the final state $\sum_{p=1}^{M} \Gamma_p = 0$, there exists a pair of indices $p_1' < p_2'$ such that $\zeta_{p_1',p_2'} > 0$ and $\Gamma_{p_1'} > 0$ and/or $\Gamma_{p_2'} > 0$. Then the problem can be reduced to an equivalent one by the substitutions

$$\Gamma_p' = \begin{cases} \Gamma_p & \text{if } p \neq p_1', \\ \Gamma_{p_1'} - 1 & \text{if } p = p_1' \end{cases}$$

(if $\Gamma_{p_1'} > 0$) or

$$\Gamma_p' = \begin{cases} \Gamma_p & \text{if } p \neq p_2', \\ \Gamma_{p_2'} - 1 & \text{if } p = p_2' \end{cases}$$

(otherwise) and

$$\zeta_{p_1,p_2} = \begin{cases} \zeta_{p_1,p_2} & \text{if } p_1 \neq p_1' \text{ or } p_2 \neq p_2', \\ \zeta_{p_1',p_2'} - 1 & \text{if } p_1 = p_1' \text{ and } p_2 = p_2'. \end{cases}$$

Clearly, the inequalities (48) remain valid for the new quantities, $\Gamma_p' \geq 0$, and $\sum_{p=1}^{M} \Gamma_p$ decreases by 1.

ii. $\exists \mathbf{v}^* \neq \mathbf{0}$ for which (48) is an equality and $\mathbf{v}_p^* = 0$ for at least one descriptor. Suppose $\mathbf{v}_p^* = 1$ for $1 \leq p \leq p^*$ and $\mathbf{v}_p^* = 0$ for $p^* + 1 \leq p \leq M$ (if necessary, permute indices). Note that $\Gamma_p \geq \sum_{p_2=p^*+1}^{M} \zeta_{p,p_2} \ \ \forall p,\ 1 \leq p \leq p^*$. (Let this be false for some index $p',\ 1 \leq p' \leq p^*$. Add the inequality $\Gamma_{p'} < \sum_{p_2=p^*+1}^{M} \zeta_{p',p_2}$ with (48), assuming there $\mathbf{v}_p = 1$ if and only if $1 \leq p \leq p^*$ and $p \neq p'$. The resultant inequality would imply that (48) for $\mathbf{v}^*$ is also a strict inequality.) For $p_1 \leq p^* < p_2$ select $\nu_{p_1,p_2} = \zeta_{p_1,p_2},\ \nu_{p_2,p_1} = 0$. Then the problem splits into two reduced subproblems of smaller dimensions $M_1' = p^*$ for $\Gamma_p' = \Gamma_p - \sum_{p_2=p^*+1}^{M} \zeta_{p,p_2}$, $1 \leq p \leq M_1'$ and $M_2' = M - p^*$ for $\Gamma_p,\ p^* < p \leq M$. For the former subproblem (48) holds because it does so in the original problem $\forall \mathbf{v} \in X^M$ with $\mathbf{v}_p = 0\ \forall p > p^*$. Similarly, (48) holds for the latter subproblem in consequence of (48) $\forall \mathbf{v} \in X^M$ with $\mathbf{v}_p = 1 \ \ \forall p \leq p^*$ and because (48) is an equality for $\mathbf{v}^*$.▷

Application of Lemma 5 to the last sum in (44) establishes a necessary condition for recognizability of a classification by Core-2:

Theorem 7. If a classification $O = C_1 \sqcup C_2$ is recognizable with some learning material by Core-2, C_1 is comprised of $\mathbf{v} \in O$ satisfying (32), where $\{\epsilon_{p_1,p_2}\}$ is an arbitrary set of integers satisfying (42) and

$$\left|\epsilon_p + \tfrac{1}{2}\left(\sum_{p_2=1}^{p-1} \epsilon_{p_2,p} + \sum_{p_2=p+1}^{M} \epsilon_{p,p_2}\right)\right|$$

$$\leq 2 + \Gamma_p + \tfrac{1}{2}\left(\sum_{p_2=1}^{p-1} \min(4 - |\epsilon_{p_2,p}|, |\epsilon_{p_2,p}|) + \sum_{p_2=p+1}^{M} \min(4 - |\epsilon_{p,p_2}|, |\epsilon_{p,p_2}|)\right).$$

Here $\{\Gamma_p\}$ is any set of non-negative integers satisfying (48), where the ζ_{p_1,p_2} are provided by (46). The same holds with respect to C_2.▷

Remark 15. Due to (33), $|\epsilon_p| \leq 2M$ in any set of acceptable coefficients of (32).

5. Low-dimensional object spaces.

To address the questions formulated in the Introduction an exhaustive computer analysis of all possible classification problems with the surjective coding of objects was performed for $3 \leq M \leq 5$.

5.1. Generation of isotypical classifications.

It is not practical to generate all 2^{2^M} partitions $X^M = C_1 \sqcup C_2$ given that the following natural equivalences of classifications allow one to significantly reduce the number of variants:

1. The classes C_1 and C_2 are interchangeable.
2. The values 0 and 1 of any particular descriptor can be simultaneously interchanged in all vectors (an affirmative answer to a particular question may equally be coded by 1 or by 0).
3. The order of descriptors (i.e. of questions in the questionnaire for coding) is irrelevant.
4. The order of vectors within each C_1 and C_2 is irrelevant.

In other words, equivalence classes are orbits of action of the group of symmetries of an M-dimensional cube on the set of partitions of the cube's vertices into two sets.

Example. There exist 13 equivalence classes of classifications $X^M = C_1 \sqcup C_2$ for $M = 3$ (see Figure 1).

For $i = 1, 2$ denote by V_i an $M \times |C_i|$ matrix comprised of row vectors $\mathbf{v}^j \in C_i$.

Isotypical representatives of each class of equivalence of classifications satisfying the following requirements were generated:

1. $|C_1| \leq |C_2|$.
2. The vector $\mathbf{v}$, where $\mathbf{v}_p = 1 \ \forall p, \ 1 \leq p \leq M$, belongs to C_1.
3. Columns in $V = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}$ and rows in V_1 and V_2 lexicographically decrease.

Condition 1 resolves equivalence 1, unless $|C_1| = |C_2|$. Condition 2 partially resolves equivalence 2 of the objects' coding: it may be satisfied by changing the encoding of the appropriate descriptors so that a given object in C_1 becomes coded by a string of ones. If encoding of descriptors is set, for a given order of vectors ordering of columns resolves equivalence 3, while for a given order of columns ordering of rows resolves equivalence 4. It is legitimate to demand that both orders are enforced simultaneously:

Lemma 6. A set of partitions mutually equivalent by equivalences 3 and 4 includes a representative where rows in V_1 and V_2 and columns in V simultaneously decrease lexicographically.

Proof. Consider the binary number $\mathcal{V}(V) = \mathbf{v}_1^1 ... \mathbf{v}_M^1 \mathbf{v}_1^2 ... \mathbf{v}_M^2 ... \mathbf{v}_1^{2^M} ... \mathbf{v}_M^{2^M}$, composed by inlining the rows of V. Each of two operations: ordering of unordered rows in V_1 or

V_2, and ordering of unordered columns in V causes $\mathcal{V}(V)$ to strictly increase. However, $\mathcal{V}(V) < 2^{M2^M}$, and hence after a finite number of alternating applications of these operations neither further ordering of rows, nor of columns can alter $\mathcal{V}(V)$. Therefore, in the corresponding matrix V rows and columns are ordered as desired.$\triangleright$

Remark 16. A representative of an equivalence class constructed in Lemma 6 is unique for a fixed coding of individual descriptors and numbering of classes.

Furthermore, ordering can be performed on vectors of V_1 alone:

Lemma 7. Suppose in the above notation

a. coding of objects is surjective,

b. rows decrease lexicographically in both V_1 and V_2,

c. columns decrease lexicographically in V_1.

Then columns of the entire matrix V also decrease lexicographically.

Proof. Suppose the P-th column lexicographically precedes the $P+1$-th column. Hence $\exists J,\ 1 \leq J \leq 2^M$, such that

$$\mathbf{v}_P^j = \mathbf{v}_{P+1}^j \quad \forall j,\ 1 \leq j < J, \tag{49}$$

and

$$\mathbf{v}_P^J > \mathbf{v}_{P+1}^J. \tag{50}$$

V is comprised of binary vectors; thus (50) implies $\mathbf{v}_P^J = 1$ and $\mathbf{v}_{P+1}^J = 0$. However, since $\mathbf{v}^J \in V_2$ (by c), a vector with $\mathbf{v}_p^{J'} = \mathbf{v}_p^J\ \ \forall p,\ 1 \leq p < P,\ \mathbf{v}_P^{J'} = 0$ and $\mathbf{v}_{P+1}^{J'} = 1$ (present by a) must be (by b) some row above $\mathbf{v}^J$ in $V:\ J' < J$. This furnishes a contradiction with (49).$\triangleright$

Finally, selection of a representative with the maximal value of $\mathcal{V}(V)$ entirely resolves equivalence 2 (an ambiguity in the choice of the vector in C_1 coded by a string of ones), as well as equivalence 1 when $|C_1| = |C_2|$. The isotypical classification with maximal $\mathcal{V}(V)$ yields a maximum over the class of equivalent classifications of the *index* of the class C_1:

Definition 8. Denote by $\pi(\mathbf{v})$ the ordinal position of a vector $\mathbf{v}$ in the list of all 2^M binary vectors in lexicographically decreasing order. The index of a class C_1 is the binary number $\mathcal{I} = i_1 i_2 ... i_{2^M}$, where $i_{\pi(\mathbf{v})} = 1$ if $\mathbf{v} \in C_1$ and $i_{\pi(\mathbf{v})} = 0$ otherwise.

5.2. Results of computation.

Types of recognizability of classifications found in computations are represented in Tables 1-3 by strings of codes of length $M-1$, where the n-th character describes the type of recognizability by Core-n. (Theorem 1 guarantees that any classification is self-recognizable by Core-M (Remark 8); thus the M-th character in the string of codes would be redundant). The following codes based on Theorems 1-7 are used for $M = 3, 4$ (referring to the n-th character in the string):

`h` - at least one of the classes is a union of hyperplanes of codimension n, and hence by Theorem 2 the classification is self-recognizable by Core-n' $\forall n' \geq n$;

`s` - none of the classes is a union of hyperplanes of codimension n, but the classification is self-recognizable by Core-n;

`c` - for the current value of n the classification has the "chessboard" structure satisfying the conditions of Theorem 4, and thus cannot be recognized by Core-n' $\forall n' \leq n$ with any learning material;

`n` - for $L_1 = C_1,\ L_2 = C_2$, the classification satisfies the conditions of Theorem 5 and thus is not self-recognizable by Core-n' $\forall n' \geq n$, but does not satisfy the conditions of Theorem 4;

`r` - the classification is not self-recognizable by Core-n, but is recognizable by it with some learning material;

`i` - an additional code for $n = 2$: each class C_i is comprised of vectors satisfying inequality (32) with an acceptable set of coefficients constructed in Theorem 7, but the classification is not recognizable by Core-2 with any learning material;

`(blank)` - none of the above: the classification is not recognizable by Core-n with any learning material but does not satisfy the conditions of Theorem 4 for current n, and for $n = 2$ neither class can be described by (32) with acceptable coefficients.

Example 1. The sixth line of Table 2 states that for $M = 4$ there exist 78 distinct isotypical classifications with surjective coding, which for $L_1 = C_1$, $L_2 = C_2$ and $n = 1$ satisfy the conditions of Theorem 5 and thereby are not self-recognizable by Core-1; which are not self-recognizable by Core-2, but are recognizable by Core-2 with some learning material; at least one of the classes of which is a union of hyperplanes of codimension 3, and hence by Theorem 2 the classification is self-recognizable by Core-3.

Example 2. For $O = X^M$ one class is a union of hyperfaces of codimension 1 if and only if the second class is a hyperface (such classifications are considered in Theorem 1). Two classifications of this type are equivalent if and only if the hyperfaces representing the second classes have equal codimensions. Therefore, recognizability of the M such isotypical classifications is coded by "$\underbrace{\text{h...h}}_{M-1}$" (first lines in the Tables 1-3).

Table 1. Types of recognizability of classifications for $M = 3$.

Recognizability type code	Number of occurrences	Isotypical representatives on Figure 1
"hh"	3	a,h,m
"sh"	2	b,e
"nh"	4	c,g,k,l
"ni"	1	d
" i"	1	j
"ch"	1	f
"cc"	1	i

Table 2. Types of recognizability of classifications for $M = 4$.

Recognizability type code	Number of occurrences
"hhh"	4
"shh"	6
"ssh"	1
"nhh"	25
"nsh"	17
"nrh"	78
"nrr"	2
"nnh"	13
"nnr"	11
"nn "	2
"nih"	14
"nir"	4
"n h"	6
"n r"	10
"n "	1
" hh"	4
" sh"	1
" rh"	11
" ir"	4
" h"	1
" r"	3
"chh"	1
"cch"	1
"ccc"	1

Table 3. Types of self-recognizability of classifications for $M = 5$.

Self-recognizability type code	Number of occurrences	Changes in the self-recognizability type caused by introduction of a redundant descriptor
"hhhh"	5	
"shhh"	11	
"sshh"	6	
"nhhh"	144	
"nshh"	307	"nshh"(265), "nShh"(6), "n hh"(36)
"nssh"	53	"nssh"(42), "nSsh"(1), "n sh"(9), "n h"(1)
"nShh"	42	"nShh"(24), "n hh"(18)
"nSsh"	1	"n sh"(1)
"nnhh"	1267	
"nnsh"	533	"nnsh"(218), "nnSh"(30), "nn h"(285)
"nnSh"	4	"nnSh"(2), "nn h"(2)
"nnnh"	147	
"nnn "	459	
"nn h"	61179	
"nn s"	10	"nn S"(10)
"nn "	12981	
"n hh"	33608	
"n sh"	8371	"n sh"(2552), "n Sh"(481), "n h"(5338)
"n Sh"	102	"n sh"(2), "n Sh"(66), "n h"(34)
"n h"	428487	"n sh"(5), "n Sh"(1), "n h"(428481)
"n s"	50	"n s"(1), "n S"(49)
"n "	65175	
" hhh"	30	
" shh"	61	" shh"(57), " Shh"(1), " hh"(3)
" ssh"	3	
" Shh"	6	" Shh"(4), " hh"(2)
" Ssh"	1	" sh"(1)
" hh"	498	
" sh"	74	" sh"(36), " Sh"(4), " h"(34)
" Sh"	7	" Sh"(4), " h"(3)
" h"	1876	
" s"	8	" s"(1), " S"(7)
" "	615	
"chhh"	1	
"cchh"	1	
"ccch"	1	
"cccc"	1	

Example 3. Two classifications considered in Theorem 4 are equivalent if and only if the sets $\mathcal{P}$ defining their "chessboard" structure have the same cardinality. In any such classification each class is a union of hyperfaces of codimension $|\mathcal{P}|$. Therefore $\forall k$, $0 < k < M$, recognizability of a unique isotypical classification is coded by "$\underbrace{\mathrm{c}\ldots\mathrm{c}}_{k}\ \underbrace{\mathrm{h}\ldots\mathrm{h}}_{M-1-k}$ " (last $M-1$ lines of the Tables).

Theorem 8. For surjective coding of objects onto X^3 there exist 13 distinct equivalence classes[2] of classifications (see Figure 1) which exhibit 7 distinct types of recognizability (see Table 1). For surjective coding onto X^4 there exist 221 distinct equivalence classes[2] of classifications, exhibiting 24 distinct types of recognizability (see Table 2). The Tables display the number of isotypical classifications of each type.

For $M = 3, 4$ the following statements are true:

[2]For both $M = 3$ and 4, degenerate classifications involving an empty class are omitted.

i. If a classification is self-recognizable by Core-n, it can be self-recognized by Core-n with $A_i \geq \min(2^{M-n-1},\ |C_i|)\ \ \forall i = 1, 2$, and thus by Theorem 3 it is also self-recognizable by Core-n' $\ \forall n' \geq n$.

ii. If for some n and for $L_1 = C_1,\ L_2 = C_2$, a classification satisfies the conditions of Theorem 5 (in particular, if it has the "chessboard" structure satisfying for that n the conditions of Theorem 4), then it is not recognizable by Core-n with any learning material; if in addition $n = 2$, neither class of the classification can be described by an inequality (32) with any acceptable set of coefficients of Theorem 7.

Proof. The Theorem was verified by direct computation.▷

Because of limited computational resources, examination of recognizability of classifications with an arbitrary learning material and generation of all classifications satisfying (32) for any acceptable coefficients were not performed for $M = 5$. Consequently, a modified system of codes is used in Table 3:

`h` - same as for $M = 3, 4$;

`s` - none of the classes is a union of hyperplanes of codimension n, but the classification is self-recognizable by Core-n with A_1 and A_2 satisfying (8) and hence by Theorem 3 it is self-recognizable by Core-n' $\ \forall n' \geq n$;

`S` - the classification is not of the above two types, but it is still self-recognizable by Core-n;

`c, n` - same as for $M = 3, 4$;

`(blank)` - none of the above: the classification is not self-recognizable by Core-n but does not satisfy the conditions of Theorem 5 for current n.

Theorem 9. For surjective coding of objects onto X^5 there exist 616125 distinct equivalence classes[3] of classifications. They exhibit 37 distinct types of recognizability listed in Table 3 with the number of isotypical classifications of each type.

Except for two isotypical classifications (codes "`nSsh`" and " `Ssh`"; see Figure 2), if a classification is self-recognizable by Core-n then it is self-recognizable by Core-n with A_1 and A_2 satisfying (8) and/or at least one of the classes is a union of hyperplanes of codimension $n+1$, and hence by Theorems 3 and 2, respectively, the classification is also self-recognizable by Core-n' $\ \forall n' \geq n$.

Proof. The Theorem was verified by direct computation.▷

Theorems 1-3, 6, 8 and 9 suggest the following

Conjecture. Let coding be surjective. If a classification is recognizable by Core-n with the classification itself used as a learning material, then it is also self-recognizable by Core-n' $\forall n' > n$.▷

According to the Tables, the numbers of self-recognizable and of recognizable classifications decrease rapidly with n; hence the assumption of self-recognizability of a classification with low n is a strong restriction on its structure.

6. Impact of low-informative descriptors.

Though simple statistical tools (see Gelfand et al., 1976) can be used to discriminate some of the descriptors as non-informative (i.e. to exclude some questions from the questionnaire at the stage of coding, reducing the dimension of the hypercube X^M), in research problems vectors representing objects commonly have redundant or low-informative descriptors. A numerical experiment showing their impact on results of pattern recognition is briefly discussed in this Section.

Let $p : X^{M+1} \to X^M$ be the projection: $p\mathbf{v} = (\mathbf{v}_1, ..., \mathbf{v}_M)$ for $\mathbf{v} = (\mathbf{v}_1, ..., \mathbf{v}_M, \mathbf{v}_{M+1})$.

Definition 9. Recognition of a classification $O = C_1 \sqcup C_2 \subset X^M$ with learning material $L_1 \sqcup L_2$ by Core-n is *stable under addition of a non-informative descriptor* $\mathbf{v}_{M+1}$ if the classification $p^{-1}C_1 \sqcup p^{-1}C_2 \subset X^{M+1}$ can be recognized by Core-n with the learning material $p^{-1}L_1 \sqcup p^{-1}L_2$.

[3] The degenerate classification involving an empty class is omitted.

By Theorems 1,2 and 6 self-recognition of a classification $C_1 \sqcup C_2$ would necessarily be stable under addition of a non-informative descriptor if at least one of the classes C_i is any hyperplane or a union of hyperplanes of codimension n, or if the original classification $C_1 \sqcup C_2$ is recognizable by Core-1.

Computations revealed that for $M = 3$ recognition of classifications is always stable under addition of a redundant descriptor. For $M = 4$ only two isotypical classifications (of the type "**nsh**"; see Figure 3) are not self-recognizable stably in this sense: they cease to be self-recognizable by Core-2 when a redundant descriptor is added. (In particular, no cases of the **S**-type of self-recognizability arise.) However, for $M = 5$, changes in the type of self-recognizability of classifications become significant. Codes of new types abridged to 4 leading characters are listed in Table 3 (numbers of isotypical classifications of new types are shown in brackets), unless the experiment produces no change for a particular code. (It was not examined in this experiment how addition of a redundant descriptor alters recognizability of classifications with an arbitrary learning material.)

These results demonstrate that even for a problem with a moderate number of descriptors the presence of low-informative descriptors can significantly alter results of recognition by Core-n, and their influence may grow with the dimension of the problem. It is therefore advantageous to remove these descriptors during data preparation, if possible.

Examination of the stability of recognition under addition of redundant descriptors can be recommended to verify that the results obtained for some classification problem are not spurious due to the presence of low-informative descriptors in the coding of the objects.

Acknowledgments.

We are grateful to Karl Wehrhahn for fruitful discussions. SS was supported by a Summer Vacation Scholarship from the School of Mathematics & Statistics at the University of Sydney and by the Southern Cross Foundation, Australia.

References.

Bongard M.M. (1967). *Problema uznavaniya*, Moscow, Nauka (in Russian). English tr.: *Pattern recognition*, New York, Spartan Books, 1970.

Caputo M. & Kolari J. (1990). Pattern Recognition of the Financial Condition of Banks. *Accademia delle Scienze di Ferrara Atti*, vol. **660** anno accademico 1660, 1989/90, 1-13.

Gelfand I.M., Guberman Sh.A., Izvekova M.A. et al. (1972). On criteria of high seismicity. *Doklady AN SSSR*, **202**, # 6, 1317-1320.

Gelfand I.M., Guberman Sh.A., Keilis-Borok V.I., Knopoff L., Press F., Ranzman E.Ya., Rotwain I.M. & Sadovsky. A.M. (1976). Pattern recognition applied to earthquake epicenters in California. *Phys. Earth and Planetary Interior*, **11**, 227-283.

Gvishiani A.D. & Gurvich V.A. (1984). Dynamical problems of pattern recognition. 1. Stability conditions for the forecast of places of strong earthquakes. In *Mathematical modeling and interpretation of geophysical data. Computational seismology*, iss. **16**, 70-88. Moscow, Nauka.

Lichtman, A.J. & DeCell K. (1990). *The thirteen keys to the presidency*, Lanham, Madison Books, 449 pp.

Figure 1. The maximal index isotypical representatives of each equivalence class of partitions $C_1 \sqcup C_2 = X^3$. C_1 - solid vertices, C_2 - hollow vertices. (a) $\mathcal{I} = 11110000$, (b) $\mathcal{I} = 11101000$, (c) $\mathcal{I} = 11100100$, (d) $\mathcal{I} = 11100001$, (e) $\mathcal{I} = 11100000$, (f) $\mathcal{I} = 11000011$, (g) $\mathcal{I} = 11000010$, (h) $\mathcal{I} = 11000000$, (i) $\mathcal{I} = 10010110$, (j) $\mathcal{I} = 10010100$, (k) $\mathcal{I} = 10010000$, (l) $\mathcal{I} = 10000001$, (m) $\mathcal{I} = 10000000$.

Figure 2. Isotypical classifications in X^5, self-recognizable by Core-n $\forall n \geq 2$, where none of the classes is a union of hyperplanes of codimension 3, and which are not self-recognizable by Core-2 with any A_1 and A_2 satisfying (8) for $n = 2$ (see Theorem 9). C_1 - solid vertices, C_2 - hollow vertices. (a) code "**nSsh**", (b) code " **Ssh**". (A standard projection of a 4-dimensional hypercube on the plane is used to represent the hypercubes $x_5 = 0$ (bold lines) and $x_5 = 1$ (normal lines).)

Figure 3. Isotypical classifications in X^4 which are not stably self-recognizable.
(a) $\mathcal{I} = 1110010000100100$, (b) $\mathcal{I} = 1110010000000011$.
(The cubes $x_4 = 0$ and $x_4 = 1$ are shown in bold and normal lines, respectively.)