HyperVector and SuperHyperVector Spaces with Applications in Machine Learning: Feature, Support, and Relevance Vectors

Takaaki Fujita [1] *

[1] Independent Researcher, Shinjuku, Shinjuku-ku, Tokyo, Japan.

Abstract

This paper introduces the concept of a *SuperHyperVector Space*, which extends classical vector spaces via the SuperHyperstructure framework built on the n^{th} iterated powerset. We first review how Hyperstructures and SuperHyperstructures arise by applying the powerset and iterated powerset operations to a base set. We then recall that a vector space consists of a set equipped with addition and scalar multiplication satisfying linearity axioms, and that a hypervector space generalizes this structure by using a scalar hyperoperation that assigns to each scalar–vector pair a nonempty subset of vectors while preserving distributivity and associativity. Building on these ideas, we define SuperHyperVector Spaces by introducing a SuperHyperOperation on the iterated powerset of the underlying group and briefly examine their fundamental properties and hierarchical modeling potential. Furthermore, in the context of Machine Learning, we investigate extensions of the HyperVector concept—including *Feature Vector*, *Support Vector*, and *Relevance Vector*—through the use of HyperVector and SuperHyperVector representations.

Keywords: Hyperstructure, Superhyperstructure, Vector, HyperVector, SuperHyperVector, Feature Vector, Support Vector, Relevance Vector

Contents

The format of this paper is described below.

1 Preliminaries

In this paper, we collect the basic notions and notation used throughout. Unless stated otherwise, all sets and structures considered are finite.

1.1 Classical Structures, Hyperstructures, and n-Superhyperstructures

We now review three increasingly general frameworks: classical structures, hyperstructures, and n-superhyperstructures. Intuitively, hyperstructures arise by applying the powerset operation once [1–4], while n-superhyperstructures employ the n^{th} iterated powerset [5–8].

Definition 1.1 (Base Set)**.** [9, 10] A *base set* S is the underlying collection from which all higher-order constructions are built:

$$S = \{\, x \mid x \text{ is an element of the domain}\}.$$

Every element of $\mathcal{P}(S)$ or $\mathcal{P}_n(S)$ originates from S.

Definition 1.2 (Powerset)**.** [11, 12] For any set S, its *powerset* $\mathcal{P}(S)$ is the family of all subsets of S, including the empty set:

$$\mathcal{P}(S) \;=\; \{\, A \mid A \subseteq S\}.$$

Definition 1.3 (n-th Powerset)**.** [13–17] Let H be a nonempty set. The *n-th powerset* of H, denoted $\mathcal{P}_n(H)$, is defined by

$$\mathcal{P}_1(H) = \mathcal{P}(H), \quad \mathcal{P}_{k+1}(H) = \mathcal{P}\big(\mathcal{P}_k(H)\big) \quad (k \geq 1).$$

Similarly, the *n-th nonempty powerset* $\mathcal{P}^*_n(H)$ is given by

$$\mathcal{P}^*_1(H) = \mathcal{P}^*(H), \quad \mathcal{P}^*_{k+1}(H) = \mathcal{P}^*\big(\mathcal{P}^*_k(H)\big),$$

where $\mathcal{P}^*(H) = \mathcal{P}(H) \setminus \{\emptyset\}$ excludes the empty set.

Example 1.4 (n^{th} Powerset in Electronic Circuit Design)**.** Consider the set of basic circuit components

$$H = \{R,\ C,\ L\},$$

where R is a resistor, C a capacitor, and L an inductor.

$\mathcal{P}_1(H) = \mathcal{P}(H)$ consists of all possible sub-circuits:

$$\mathcal{P}_1(H) = \big\{\,\emptyset,\ \{R\},\ \{C\},\ \{L\},\ \{R,C\},\ \{R,L\},\ \{C,L\},\ \{R,C,L\}\big\}.$$

The second iterated powerset, $\mathcal{P}_2(H) = \mathcal{P}(\mathcal{P}_1(H))$, is the collection of all sets of sub-circuits. For example:

$$X = \big\{\{R\},\ \{C,L\}\big\} \;\in\; \mathcal{P}_2(H),$$

represents the design choice of either the single-resistor circuit $\{R\}$ or the series LC circuit $\{C, L\}$.

More generally, higher iterates $\mathcal{P}_n(H)$ with $n \geq 3$ would model "choices of choices" of circuit configurations, enabling a hierarchical organization of design alternatives.

To establish a formal foundation for the concepts of Hyperstructures and Superhyperstructures, we present the following definitions and propositions.

Definition 1.5 (Classical Structure)**.** (cf. [8, 18]) A *Classical Structure* is a mathematical framework defined on a non-empty set H, equipped with one or more *Classical Operations* that satisfy specified *Classical Axioms*. Specifically:

A *Classical Operation* is a function of the form:

$$\#_0 : H^m \to H,$$

where $m \geq 1$ is a positive integer, and H^m denotes the m-fold Cartesian product of H. Common examples include addition and multiplication in algebraic structures such as groups, rings, and fields.

Definition 1.6 (Hyperoperation)**.** (cf. [19–22]) A *hyperoperation* is a generalization of a binary operation where the result of combining two elements is a set, not a single element. Formally, for a set S, a hyperoperation $\circ$ is defined as:

$$\circ : S \times S \to \mathcal{P}(S),$$

where $\mathcal{P}(S)$ is the powerset of S.

Definition 1.7 (Hyperstructure)**.** (cf. [4,8,23,24]) A *Hyperstructure* extends the notion of a Classical Structure by operating on the powerset of a base set. Formally, it is defined as:

$$\mathcal{H} = (\mathcal{P}(S), \circ),$$

where S is the base set, $\mathcal{P}(S)$ is the powerset of S, and $\circ$ is an operation defined on subsets of $\mathcal{P}(S)$. Hyperstructures allow for generalized operations that can apply to collections of elements rather than single elements.

Example 1.8 (Power Hyperstructure of an Abelian Group)**.** Let $(G, +)$ be an abelian group and denote by $\mathcal{P}(G)$ its powerset. Define a hyperoperation

$$\oplus : \mathcal{P}(G) \times \mathcal{P}(G) \longrightarrow \mathcal{P}(G) \quad \text{by} \quad A \oplus B = \{ a + b \mid a \in A,\ b \in B \}.$$

Then $\mathcal{H} = \big(\mathcal{P}(G), \oplus\big)$ is a hyperstructure. In particular:

$$(A \oplus B) \oplus C = \{ (a + b) + c \mid a \in A,\ b \in B,\ c \in C \} = A \oplus (B \oplus C),$$

so $\oplus$ is associative, and the singleton $\{0\}$ is the identity: $\{0\} \oplus A = A \oplus \{0\} = A$ for all $A \subseteq G$.

Example 1.9 (Interval Hyperring on $\mathbb{R}$)**.** Fix a tolerance $\delta > 0$. For real numbers $x, y \in \mathbb{R}$, set

$$x \boxplus y = [\, x + y - \delta,\ x + y + \delta \,], \qquad x \boxtimes y = [\, x\,y - \delta,\ x\,y + \delta \,].$$

Extend these to nonempty subsets $A, B \subseteq \mathbb{R}$ by

$$A \boxplus B = \bigcup_{x \in A,\ y \in B} (x \boxplus y), \qquad A \boxtimes B = \bigcup_{x \in A,\ y \in B} (x \boxtimes y).$$

Then $\mathcal{H} = (\mathcal{P}^*(\mathbb{R}), \boxplus, \boxtimes)$ is a hyperstructure with two hyperoperations. It satisfies the analogues of ring axioms in the sense that $\boxplus$ is associative and commutative with identity $\{0\}$, $\boxtimes$ is associative with identity $\{1\}$, and $\boxtimes$ distributes over $\boxplus$ in the hyper-set sense.

Definition 1.10 (SuperHyperOperations)**.** (cf. [8]) Let H be a non-empty set, and let $\mathcal{P}(H)$ denote the powerset of H. The n-th powerset $\mathcal{P}^n(H)$ is defined recursively as follows:

$$\mathcal{P}^0(H) = H, \quad \mathcal{P}^{k+1}(H) = \mathcal{P}(\mathcal{P}^k(H)), \quad \text{for } k \geq 0.$$

A *SuperHyperOperation* of order (m, n) is an m-ary operation:

$$\circ^{(m,n)} : H^m \to \mathcal{P}^n_*(H),$$

where $\mathcal{P}^n_*(H)$ represents the n-th powerset of H, either excluding or including the empty set, depending on the type of operation:

- If the codomain is $\mathcal{P}^n_*(H)$ excluding the empty set, it is called a *classical-type* (m, n)*-SuperHyperOperation*.
- If the codomain is $\mathcal{P}^n(H)$ including the empty set, it is called a *Neutrosophic* (m, n)*-SuperHyperOperation*.

These SuperHyperOperations are higher-order generalizations of hyperoperations, capturing multi-level complexity through the construction of n-th powersets.

Definition 1.11 (n-Superhyperstructure)**.** (cf. [8, 25–27]) An *n-Superhyperstructure* further generalizes a Hyperstructure by incorporating the n-th powerset of a base set. It is formally described as:

$$\mathcal{SH}_n = (\mathcal{P}_n(S), \circ),$$

where S is the base set, $\mathcal{P}_n(S)$ is the n-th powerset of S, and $\circ$ represents an operation defined on elements of $\mathcal{P}_n(S)$. This iterative framework allows for increasingly hierarchical and complex representations of relationships within the base set.

Example 1.12 (Union Superhyperstructure on a Two-Element Set)**.** Let $S = \{a, b\}$. Then

$$\mathcal{P}^1(S) = \mathcal{P}(S) = \{\emptyset, \{a\}, \{b\}, \{a, b\}\}, \quad \mathcal{P}^2(S) = \mathcal{P}(\mathcal{P}^1(S)),$$

which has $2^4 = 16$ elements (all subsets of $\mathcal{P}^1(S)$). We now define a binary SuperHyperOperation

$$\star : \mathcal{P}^2(S) \times \mathcal{P}^2(S) \longrightarrow \mathcal{P}^2(S)$$

by

$$X \star Y = \{U \cup V \mid U \in X,\ V \in Y\},$$

where each $U, V \subseteq S$. Concretely, if

$$X = \{\{a\}, \{a, b\}\}, \quad Y = \{\{b\}, \{a, b\}\},$$

then

$$X \star Y = \{\{a\} \cup \{b\},\ \{a\} \cup \{a, b\},\ \{a, b\} \cup \{b\},\ \{a, b\} \cup \{a, b\}\} = \{\{a, b\}\}.$$

One checks readily that:

- $\star$ is *closed* in $\mathcal{P}^2(S)$.
- $\star$ is *associative*, since union in $\mathcal{P}(S)$ is associative:

$$(X \star Y) \star Z = \{ (U \cup V) \cup W \mid U \in X,\ V \in Y,\ W \in Z\} = X \star (Y \star Z).$$

- The singleton $\{\emptyset\} \subseteq \mathcal{P}^1(S)$ viewed in $\mathcal{P}^2(S)$ acts as a *neutral element*: $\{\emptyset\} \star X = X \star \{\emptyset\} = X$.

Therefore $(\mathcal{P}^2(S), \star)$ is a concrete 2-Superhyperstructure (i.e. $(m, n) = (2, 2)$).

Note that related concepts to SuperHyperStructure include SuperHyperGraph [28–32], Chemical SuperHypersturcture [5], and SuperHyperAlgebra [33–35].

1.2 (m, n)-Superhyperstructure

An (m, n)-superhyperstructure generalizes hyperstructures by defining an m-ary operation on the hierarchical n-th iterated powerset of a finite base set (cf. [36, 37]).

Definition 1.13 ((m, n)-Superhyperstructure)**.** Let S be a nonempty set and let $\mathcal{P}_n(S)$ be its n^{th} iterated powerset (Definition 1.3). Fix a positive integer m. An (m, n)*-superhyperstructure* is a pair

$$\mathcal{SH}_{m,n} = (\mathcal{P}_n(S),\ \circ^{(m,n)}),$$

where

$$\circ^{(m,n)} : (\mathcal{P}_n(S))^m \longrightarrow \mathcal{P}_n(S)$$

is an m-ary *SuperHyperOperation* of order (m, n), meaning that for any $\mathbf{X} = (X_1, \ldots, X_m) \in (\mathcal{P}_n(S))^m$, $\circ^{(m,n)}(\mathbf{X})$ is a subset of $\mathcal{P}_n(S)$ satisfying whatever axioms (e.g. associativity, distributivity) are imposed for the given structure.

Example 1.14 (Weekly meal planning as a real–life $(2, 2)$-superhyperstructure)**.** Let the base set S be the catalogue of single–day dishes

$$S = \{\text{Chicken, Tofu, Rice, Salad, Pasta}\}.$$

Elements of $\mathcal{P}_1(S) = \mathcal{P}(S)$ are *daily menus* (sets of dishes). Elements of $\mathcal{P}_2(S) = \mathcal{P}(\mathcal{P}(S))$ are *collections of daily menus* (e.g., choices provided by different meal kits).

Fix a daily calorie budget $C > 0$ and a cost function $\mathrm{cal} : \mathcal{P}(S) \to \mathbb{R}_{\geq 0}$ given by the additive per–dish calories:

$$\mathrm{cal}(\emptyset) = 0, \quad \mathrm{cal}(U) = \sum_{d \in U} \mathrm{cal}(d).$$

Define the binary SuperHyperOperation

$$\diamond : \mathcal{P}_2(S) \times \mathcal{P}_2(S) \longrightarrow \mathcal{P}_2(S), \qquad X \diamond Y \;:=\; \big\{ U \cup V \;\big|\; U \in X,\ V \in Y,\ \mathrm{cal}(U \cup V) \leq C \big\}.$$

Then $(\mathcal{P}_2(S), \diamond)$ is a concrete $(2, 2)$-superhyperstructure: it is closed in $\mathcal{P}_2(S)$, and associativity holds because $(U \cup V) \cup W = U \cup (V \cup W)$ and the predicate $\mathrm{cal}(\cdot) \leq C$ depends only on the final union, not on grouping:

$$(X \diamond Y) \diamond Z = \big\{ (U \cup V) \cup W \;\big|\; U \in X,\ V \in Y,\ W \in Z,\ \mathrm{cal}((U \cup V) \cup W) \leq C \big\} = X \diamond (Y \diamond Z).$$

Numerical instance. Let

$$\mathrm{cal}(\text{Chicken}) = 400, \quad \mathrm{cal}(\text{Tofu}) = 250, \quad \mathrm{cal}(\text{Rice}) = 200,$$

$$\mathrm{cal}(\text{Salad}) = 100, \quad \mathrm{cal}(\text{Pasta}) = 500, \quad C = 800.$$

Take two collections of menus

$$X = \big\{\{\text{Chicken, Rice}\},\ \{\text{Tofu, Salad}\}\big\}, \qquad Y = \big\{\{\text{Salad}\},\ \{\text{Pasta}\}\big\}.$$

All pairwise unions and their calories are

$$\begin{array}{lcl}
\{\text{Chicken, Rice}\} \cup \{\text{Salad}\} & = & \{\text{Chicken, Rice, Salad}\} \quad (400 + 200 + 100 = 700 \leq 800) \checkmark \\
\{\text{Chicken, Rice}\} \cup \{\text{Pasta}\} & = & \{\text{Chicken, Rice, Pasta}\} \quad (400 + 200 + 500 = 1100 > 800) \times \\
\{\text{Tofu, Salad}\} \cup \{\text{Salad}\} & = & \{\text{Tofu, Salad}\} \quad (250 + 100 = 350 \leq 800) \checkmark \\
\{\text{Tofu, Salad}\} \cup \{\text{Pasta}\} & = & \{\text{Tofu, Salad, Pasta}\} \quad (250 + 100 + 500 = 850 > 800) \times
\end{array}$$

Hence

$$X \diamond Y = \big\{\{\text{Chicken, Rice, Salad}\},\ \{\text{Tofu, Salad}\}\big\} \;\in\; \mathcal{P}_2(S).$$

Operationally, $X \diamond Y$ lists all feasible daily menus that arise by combining one proposal from each provider (e.g., two meal–kit catalogs) while enforcing the real–life calorie budget. This models "choice of choices" under a practical constraint and thus constitutes a tangible $(2, 2)$-superhyperstructure.

Theorem 1.15. *Every n-superhyperstructure $\mathcal{SH}_n = (\mathcal{P}_n(S),\ \circ)$ with a single binary operation $\circ$ can be regarded as a $(2, n)$-superhyperstructure.*

Proof. Let $\mathcal{SH}_n = (\mathcal{P}_n(S),\ \circ)$ be an n-superhyperstructure as in the Definition. Define

$$\circ^{(2,n)} \;:\; \mathcal{P}_n(S) \times \mathcal{P}_n(S) \;\longrightarrow\; \mathcal{P}_n(S) \quad \text{by} \quad \circ^{(2,n)}(X, Y) = X \;\circ\; Y.$$

Since $\circ$ already maps two inputs in $\mathcal{P}_n(S)$ to an element of $\mathcal{P}_n(S)$, $\circ^{(2,n)}$ is a valid binary SuperHyperOperation of order $(2, n)$. All structural axioms (e.g. associativity or distributivity) that held for $\circ$ continue to hold for $\circ^{(2,n)}$. Hence $(\mathcal{P}_n(S), \circ^{(2,n)})$ satisfies the definition of a $(2, n)$-superhyperstructure. □

1.3 HyperVector Space

A *vector space* consists of a set V of vectors equipped with an addition $+$ and a scalar multiplication over a field K, satisfying the standard linearity axioms [38–41]. Well-known extensions include fuzzy vector spaces [42–45] and neutrosophic vector spaces [46–49]. A *hypervector space* further generalizes this structure by replacing scalar multiplication with a scalar hyperoperation

$$\circ : K \times V \longrightarrow \mathcal{P}^*(V),$$

which assigns to each $(a, x) \in K \times V$ a nonempty subset $a \circ x \subseteq V$ while preserving distributivity and associativity (cf. [50, 51]).

Definition 1.16 (Hypervector Space)**.** (cf. [50, 51]) Let K be a field and $(V, +)$ an abelian group. Denote by $\mathcal{P}^*(V)$ the set of all nonempty subsets of V. A *hypervector space* over K is a quadruple $(V, +, \circ, K)$, where

$$\circ : K \times V \ \longrightarrow\ \mathcal{P}^*(V)$$

is an external hyperoperation satisfying, for all $a, b \in K$ and $x, y \in V$, the following axioms:

(H1) $a \circ (x + y) \ \subseteq\ a \circ x \ +\ a \circ y,$ (right distributivity)

(H2) $(a + b) \circ x \ \subseteq\ a \circ x \ +\ b \circ x,$ (left distributivity)

(H3) $a \circ \bigl(b \circ x\bigr) \ =\ (ab) \circ x,$ (associativity)

(H4) $a \circ (-x) \ =\ (-a) \circ x \ =\ -\bigl(a \circ x\bigr),$

(H5) $x \ \in\ 1 \circ x.$

Here, for any subsets $A, B \subseteq V$, we write

$$A + B \ =\ \{u + v \mid u \in A,\ v \in B\},$$

and interpret $a \circ \bigl(b \circ x\bigr) = \bigcup_{y \in b \circ x} \bigl(a \circ y\bigr).$

Example 1.17 (Sign-Symmetric Hypervector Space)**.** Let K be any field and $(V, +)$ a (classical) vector space over K. Define a scalar-hyperoperation

$$\circ : K \times V \ \longrightarrow\ \mathcal{P}^*(V) \quad \text{by} \quad a \circ x \ =\ \{a\,x,\ -a\,x\},$$

for all $a \in K$ and $x \in V$. Then $\bigl(V, +, \circ, K\bigr)$ is a hypervector space. Indeed:

(H1) Right distributivity:

$$a \circ (x + y) = \{\,a(x + y),\ -a(x + y)\} \subseteq \{\,ax + ay,\ -ax - ay\} \subseteq a \circ x \ +\ a \circ y.$$

(H2) Left distributivity:

$$(a + b) \circ x = \{(a + b)x,\ -(a + b)x\} \subseteq a \circ x \ +\ b \circ x.$$

(H3) Associativity of scalars:

$$a \circ \bigl(b \circ x\bigr) = \bigcup_{y \in \{bx, -bx\}} \{\,a\,y,\ -a\,y\} = \{abx,\ -abx\} = (ab) \circ x.$$

(H4) Compatibility with negatives:

$$a \circ (-x) = \{a(-x),\ -a(-x)\} = \{-ax,\ ax\} = (-a) \circ x = -\bigl(a \circ x\bigr).$$

(H5) Unit scalar:

$$1 \circ x = \{1 \cdot x,\ -1 \cdot x\} = \{x,\ -x\} \ni x.$$

Here, for subsets $A, B \subseteq V$, $A + B = \{u + v \mid u \in A,\ v \in B\}$, and $a \circ (b \circ x) = \bigcup_{y \in b \circ x} (a \circ y)$.

2 Result: (m, n)-SuperhyperVector Space

An (m, n)-SuperhyperVector Space generalizes hypervector spaces by introducing an m-ary scalar SuperHyperOperation on the n-th iterated powerset of the underlying vector group.

Definition 2.1 ((m, n)-SuperhyperVector Space)**.** Let K be a field, $(V, +)$ an abelian group, and $m, n \geq 1$ integers. Denote by $\mathcal{P}_n^*(V)$ the set of all nonempty elements of the n-th iterated powerset $\mathcal{P}_n(V)$. An (m, n)-*SuperhyperVector Space* over K is a quadruple

$$\bigl(V,\ +\ ,\ \circ^{(m,n)},\ K\bigr),$$

where

$$\circ^{(m,n)} : K^m \times V \longrightarrow \mathcal{P}_n^*(V)$$

is an m-ary SuperHyperOperation satisfying, for all $\mathbf{a} = (a_1, \ldots, a_m)$, $\mathbf{b} = (b_1, \ldots, b_m) \in K^m$ and all $x, y \in V$:

(SV1) $\circ^{(m,n)}(\mathbf{a},\ x + y) \ \subseteq\ \circ^{(m,n)}(\mathbf{a},\ x)\ +\ \circ^{(m,n)}(\mathbf{a},\ y),$

(SV2) $\circ^{(m,n)}(\mathbf{a} + \mathbf{b},\ x) \ \subseteq\ \circ^{(m,n)}(\mathbf{a},\ x)\ +\ \circ^{(m,n)}(\mathbf{b},\ x),$

(SV3) $\circ^{(m,n)}(\mathbf{a},\ \circ^{(m,n)}(\mathbf{b}, x)) \ =\ \circ^{(m,n)}(\mathbf{a} \cdot \mathbf{b},\ x),$

(SV4) $\circ^{(m,n)}(-\mathbf{a},\ x) \ =\ -\circ^{(m,n)}\ (\mathbf{a},\ x),$

(SV5) $x \ \in\ \circ^{(m,n)}((1, \ldots, 1),\ x).$

Here:

$$\mathbf{a} + \mathbf{b} = (a_1 + b_1, \ldots, a_m + b_m), \quad \mathbf{a} \cdot \mathbf{b} = (a_1 b_1, \ldots, a_m b_m),$$

and for subsets $A, B \subseteq V$, $A + B = \{u + v \mid u \in A,\ v \in B\}$.

Example 2.2 ($(2, 2)$-SuperhyperVector Space on the Real Line)**.** Let $K = \mathbb{R}$ (the field of real numbers) and $V = \mathbb{R}$ viewed as an additive group. Then

$$\mathcal{P}_1(V) = \mathcal{P}(\mathbb{R}), \quad \mathcal{P}_2(V) = \mathcal{P}\bigl(\mathcal{P}(\mathbb{R})\bigr).$$

We define

$$\circ^{(2,2)} : K^2 \times V \longrightarrow \mathcal{P}_2^*(V) \quad \text{by} \quad \circ^{(2,2)}\bigl((a, b),\ x\bigr) = \bigl\{\{a\,x\},\ \{b\,x\}\bigr\}.$$

Thus each pair $(a, b) \in \mathbb{R}^2$ and $x \in \mathbb{R}$ is sent to the nonempty set of two singleton subsets of $\mathbb{R}$.

We verify the axioms (SV1)–(SV5):

(SV1) For $x, y \in \mathbb{R}$,

$$\circ^{(2,2)}\bigl((a, b),\ x{+}y\bigr) = \{\{a(x{+}y)\}, \{b(x{+}y)\}\} \subseteq \{\{ax{+}ay\}, \{bx{+}by\}\} \subseteq \circ^{(2,2)}((a, b), x)\ +\ \circ^{(2,2)}((a, b), y).$$

(SV2) For $(a, b), (c, d) \in \mathbb{R}^2$,

$$\circ^{(2,2)}\bigl((a{+}c, b{+}d),\ x\bigr) = \{\{(a{+}c)x\}, \{(b{+}d)x\}\} \subseteq \{\{ax\}, \{bx\}\}{+}\{\{cx\}, \{dx\}\} = \circ^{(2,2)}((a, b), x){+}\circ^{(2,2)}((c, d), x).$$

(SV3) For $(a, b), (c, d) \in \mathbb{R}^2$,

$$\circ^{(2,2)}\bigl((a, b),\ \circ^{(2,2)}((c, d), x)\bigr) = \bigl\{\{a \cdot (cx)\}, \{b \cdot (dx)\}\bigr\} = \circ^{(2,2)}\bigl((ac, bd),\ x\bigr).$$

(SV4) Negation is compatible:

$$\circ^{(2,2)}\bigl((-a, -b),\ x\bigr) = \{\{-ax\}, \{-bx\}\} = -\{\{ax\}, \{bx\}\} = -\circ^{(2,2)}\ ((a, b), x).$$

(SV5) The unit pair $(1,1)$ satisfies

$$\circ^{(2,2)}\big((1,1),\,x\big) = \{\{x\},\{x\}\} = \{\{x\}\} \ni \{x\},$$

so in particular $x \in \{x\}$.

Therefore $\big(\mathbb{R},\ +\ ,\ \circ^{(2,2)},\ \mathbb{R}\big)$ is a bona fide $(2,2)$-SuperhyperVector Space, exhibiting "choices of choices" of scalar-multiples at two hierarchical levels.

Example 2.3 ((1,3)-SuperhyperVector Space on $\mathbb{R}$)**.** Let $K = \mathbb{R}$ and $V = \mathbb{R}$ as an additive group. We have

$$\mathcal{P}_1(V) = \mathcal{P}(\mathbb{R}), \quad \mathcal{P}_2(V) = \mathcal{P}\big(\mathcal{P}(\mathbb{R})\big), \quad \mathcal{P}_3(V) = \mathcal{P}\big(\mathcal{P}_2(V)\big).$$

Define the scalar SuperHyperOperation

$$\circ^{(1,3)} : K \times V \longrightarrow \mathcal{P}_3^*(V) \quad \text{by} \quad \circ^{(1,3)}\,(a,x) = \big\{\ \{\{\,a\,x\}\}\big\}.$$

Thus each $\circ^{(1,3)}(a,x)$ is the singleton subset of $\mathcal{P}_2(V)$ containing the singleton $\{ax\} \subseteq \mathbb{R}$.

We verify the axioms (SV1)–(SV5) for all $a,b \in \mathbb{R}$ and $x,y \in \mathbb{R}$:

(SV1) Right distributivity:

$$\circ^{(1,3)}(a,\,x+y) = \{\{\{a(x+y)\}\}\} = \{\{\{ax+ay\}\}\} \subseteq \{\{\{ax\}\}\}\ +\ \{\{\{ay\}\}\} = \circ^{(1,3)}(a,x)\ +\ \circ^{(1,3)}(a,y).$$

(SV2) Left distributivity:

$$\circ^{(1,3)}(a+b,\,x) = \{\{\{(a+b)x\}\}\} = \{\{\{ax+bx\}\}\} \subseteq \{\{\{ax\}\}\}\ +\ \{\{\{bx\}\}\} = \circ^{(1,3)}(a,x)\ +\ \circ^{(1,3)}(b,x).$$

(SV3) Scalar associativity:

$$\circ^{(1,3)}\big(a,\ \circ^{(1,3)}(b,x)\big) = \bigcup_{Z \in \{\{\{bx\}\}\}} \circ^{(1,3)}(a,Z) = \circ^{(1,3)}(a,bx) = \{\{\{a(bx)\}\}\} = \circ^{(1,3)}(ab,x).$$

(SV4) Negation compatibility:

$$\circ^{(1,3)}(-a,\,x) = \{\{\{-a\,x\}\}\} = -\{\{\{a\,x\}\}\} = -\circ^{(1,3)}\,(a,x).$$

(SV5) Unit scalar:

$$\circ^{(1,3)}(1,\,x) = \{\{\{1 \cdot x\}\}\} = \{\{\{x\}\}\} \ni \{\{x\}\},$$

so in particular $x \in \{\{x\}\}$.

Hence $\big(\mathbb{R},\ +, \circ^{(1,3)}, \mathbb{R}\big)$ is a valid $(1,3)$-SuperhyperVector Space, exhibiting three-level hierarchical "choices" of scalar multiplication.

Theorem 2.4. *Every (m,n)-SuperhyperVector Space $\big(V,+,\circ^{(m,n)},K\big)$ is an (m,n)-superhyperstructure on the base set V.*

Proof. By definition an (m,n)-superhyperstructure consists of a base set S (here $S = V$) together with an m-ary operation $\circ^{(m,n)} : S^m \to \mathcal{P}_n(S)$. In our case $\circ^{(m,n)}$ indeed maps $K^m \times V \to \mathcal{P}_n^*(V) \subseteq \mathcal{P}_n(V)$. The axioms (SV1)–(SV5) impose the usual closure, associativity, distributivity and identity conditions required for a superhyperstructure. Hence $\big(V,\circ^{(m,n)}\big)$ is an (m,n)-superhyperstructure, with the additional vector-space-like properties built into (SV1)–(SV5). □

Theorem 2.5. *When $m = n = 1$, an (m,n)-SuperhyperVector Space reduces to the usual hypervector space.*

Proof. Setting $m = n = 1$ gives $\circ^{(1,1)} : K\times V \to \mathcal{P}_1^*(V) = \mathcal{P}^*(V)$, and the axioms (SV1)–(SV5) become exactly the hypervector space axioms (H1)–(H5) in Definition of hypervector space. Thus any $(1,1)$-SuperhyperVector Space is precisely a hypervector space. □

Theorem 2.6 (Intersection of Subspaces)**.** *Let $\{W_i\}_{i\in I}$ be any family of superhypervector subspaces of V. Then*

$$\bigcap_{i\in I} W_i$$

is also a superhypervector subspace of V.

Proof. Set $W = \bigcap_{i\in I} W_i$. Since each W_i is nonempty and a subgroup, W is nonempty and closed under addition and inverses, so W is a subgroup of $(V,+)$. Next, let $\mathbf{a} \in K^m$ and $x \in W$. Then $x \in W_i$ for all i, so

$$\circ^{(m,n)}(\mathbf{a}, x) \subseteq W_i \quad \text{for every } i,$$

hence $\circ^{(m,n)}(\mathbf{a}, x) \subseteq \bigcap_i W_i = W$. Therefore W satisfies (S1)–(S2) and is a superhypervector subspace. □

Theorem 2.7 (Sum of Two Subspaces)**.** *If W_1 and W_2 are superhypervector subspaces of V, then their sum*

$$W_1 + W_2 \;=\; \{\, u + v \mid u \in W_1,\; v \in W_2 \}$$

is also a superhypervector subspace of V.

Proof. First, $W_1 + W_2$ is a subgroup of $(V,+)$ because the sum of two subgroups is again a subgroup. Next, take any $\mathbf{a} \in K^m$ and $w \in W_1 + W_2$. By definition $w = u + v$ with $u \in W_1$, $v \in W_2$. Then by axiom (SV1) of the ambient space,

$$\circ^{(m,n)}(\mathbf{a}, w) = \circ^{(m,n)}\big(\mathbf{a},\; u + v\big) \subseteq \circ^{(m,n)}(\mathbf{a}, u) \;+\; \circ^{(m,n)}(\mathbf{a}, v).$$

Since each W_i is a subspace, $\circ^{(m,n)}(\mathbf{a}, u) \subseteq W_1$ and $\circ^{(m,n)}(\mathbf{a}, v) \subseteq W_2$. Hence $\circ^{(m,n)}(\mathbf{a}, w) \subseteq W_1 + W_2$, showing closure under the hyperoperation. Therefore $W_1 + W_2$ is a superhypervector subspace. □

3 Additional Result: Some Applications

This section discusses several applications of the HyperVector and SuperHyperVector concepts developed in this paper. In particular, we extend the notions of *feature vector*, *support vector*, and *Relevance vector* by employing the HyperVector and SuperHyperVector frameworks, and we examine concrete examples of their use in machine learning.

3.1 Feature Vector, HyperVector, and SuperHyperVector

3.1.1 Feature Vector

A feature vector is an ordered list of numerical descriptors representing an object's characteristics for use in computational models or algorithms [52–58].

Definition 3.1 (Feature space and feature map)**.** (cf. [59–61]) Let $\mathcal{X}$ be a (nonempty) set of objects (inputs). A *feature space* is a finite-dimensional real vector space $\mathbb{R}^n$ (for some $n \in \mathbb{N}$). A *feature map* is a function

$$\varphi :\; \mathcal{X} \longrightarrow \mathbb{R}^n, \qquad x \longmapsto \varphi(x) = (\varphi_1(x), \ldots, \varphi_n(x))^\top,$$

whose coordinates $\varphi_j : \mathcal{X} \to \mathbb{R}$ are called *features*. When raw attributes are non-numeric (e.g., categorical), a fixed encoding e (such as one-hot or an embedding) is understood to be composed into φ, so that $\varphi(x) \in \mathbb{R}^n$ is always numerical.

Definition 3.2 (Feature vector and design matrix)**.** For $x \in \mathcal{X}$, the vector $\varphi(x) \in \mathbb{R}^n$ is the *feature vector* of x. Given a dataset $D = (x_1, \ldots, x_N) \in \mathcal{X}^N$, the *design matrix* (feature matrix) is

$$\Phi(D) \;:=\; \begin{bmatrix} \varphi(x_1)^\top \\ \vdots \\ \varphi(x_N)^\top \end{bmatrix} \in \mathbb{R}^{N \times n}.$$

For a weight vector $w \in \mathbb{R}^n$, a linear score (predictor) is $s_w(x) := \langle w, \varphi(x) \rangle = w^\top \varphi(x)$, used in regression/classification; a bias can be included by augmenting φ with a leading 1.

Example 3.3 (Two simple feature maps)**.**

- **Bag of words.** Let $\mathcal{X}$ be documents and $V = \{w_1, \ldots, w_n\}$ a vocabulary. Define $\varphi(x) = (\mathrm{tf}(w_1, x), \ldots, \mathrm{tf}(w_n, x))^\top$, where tf is the term frequency; then $\varphi(x) \in \mathbb{R}^n$.
- **Tabular data with a categorical field.** Suppose $x = (\texttt{age}, \texttt{color})$ with $\texttt{age} \in \mathbb{R}$ and $\texttt{color} \in \{\mathrm{red}, \mathrm{blue}, \mathrm{green}\}$. Let $e(\mathrm{red}) = (1,0,0)$, $e(\mathrm{blue}) = (0,1,0)$, $e(\mathrm{green}) = (0,0,1)$. Set $\varphi(x) = (\texttt{age}, e(\texttt{color})) \in \mathbb{R}^4$.

3.1.2 Feature HyperVector

A feature hypervector generalizes feature vectors to hypervector spaces, encoding complex, multi-valued, or set-valued attributes for richer representation.

Definition 3.4 (Feature hypermap / Feature HyperVector)**.** Let $\mathcal{X}$ be a nonempty set of objects and fix $n \in \mathbb{N}$. A *feature hypermap* is a set–valued map

$$\widetilde{\varphi} \;:\; \mathcal{X} \longrightarrow \mathcal{P}^*(\mathbb{R}^n), \qquad x \longmapsto \widetilde{\varphi}(x),$$

assigning to each object x a nonempty set $\widetilde{\varphi}(x) \subseteq \mathbb{R}^n$ of admissible feature *realizations*. Any $z \in \widetilde{\varphi}(x)$ is a *feature realization* of x, and the set $\widetilde{\varphi}(x)$ is the *Feature HyperVector* of x.

Definition 3.5 (Design hypermatrix and linear score set)**.** For a dataset $D = (x_1, \ldots, x_N)$, the *design hypermatrix* is the tuple

$$\widetilde{\Phi}(D) \;:=\; \big(\widetilde{\varphi}(x_i)\big)_{i=1}^N \;\in\; \big(\mathcal{P}^*(\mathbb{R}^n)\big)^N.$$

Given a weight vector $w \in \mathbb{R}^n$ and bias $b \in \mathbb{R}$, the *hyperlinear score set* of x is

$$S_{w,b}(x) \;:=\; \big\{\, \langle w, z \rangle - b \;\big|\; z \in \widetilde{\varphi}(x) \,\big\} \;\subseteq \mathbb{R}.$$

Let $B : \mathcal{P}_*(\mathbb{R}) \to \mathbb{R}$ be a base aggregator with $B(\{a\}) = a$ and monotonicity. The *B–crisp score* of x is

$$s^B_{w,b}(x) \;:=\; B\big(S_{w,b}(x)\big) \;\in\; \mathbb{R}.$$

Typical choices are $B_{\min}(S) = \inf S$ (robust/worst–case), $B_{\max}(S) = \sup S$ (optimistic/best–case), or an average/quantile.

Example 3.6 (Two concrete Feature HyperVectors)**.**

- **Intervalized/tabular sensors.** For $x \in \mathcal{X}$ let a nominal feature $a \in \mathbb{R}^n$ be known with componentwise tolerances $\varepsilon \in \mathbb{R}^n_{\geq 0}$. Set

  $$\widetilde{\varphi}(x) \;=\; \big\{\, z \in \mathbb{R}^n \;\big|\; |z_j - a_j| \leq \varepsilon_j \;(j = 1, \ldots, n) \,\big\}.$$

  Then, for any (w, b),

  $$S_{w,b}(x) = [\, \langle w, a \rangle - b - \|w \odot \varepsilon\|_1, \;\; \langle w, a \rangle - b + \|w \odot \varepsilon\|_1 \,],$$

  so $s^{\inf}_{w,b}(x) = \langle w, a \rangle - b - \|w \odot \varepsilon\|_1$ and $s^{\sup}_{w,b}(x) = \langle w, a \rangle - b + \|w \odot \varepsilon\|_1$.

- **Uncertain one–hot.** A categorical attribute `color` $\in$ {red, blue, green} is missing but narrowed to {red, blue}. With one–hot $e(\cdot) \in \mathbb{R}^3$ and numeric covariates $u \in \mathbb{R}^d$, define

$$\widetilde{\varphi}(x) \;=\; \big\{\, (u,\; e(\mathrm{red}))\,,\; (u,\; e(\mathrm{blue}))\,\big\} \;\subset \mathbb{R}^{d+3}.$$

  The score set is the two–point set $\{\langle w, (u, e(\mathrm{red}))\rangle - b,\; \langle w, (u, e(\mathrm{blue}))\rangle - b\}$.

Theorem 3.7 (Reduction to ordinary features)**.** *If $\widetilde{\varphi}(x) = \{\varphi(x)\}$ is a singleton for all $x \in \mathcal{X}$, then $S_{w,b}(x) = \{\langle w, \varphi(x)\rangle - b\}$ and $s^B_{w,b}(x) = \langle w, \varphi(x)\rangle - b$ for any admissible B. In particular, Feature HyperVectors reduce to ordinary feature vectors.*

Proof. Fix (w, b) and $x \in \mathcal{X}$. By the singleton hypothesis $\widetilde{\varphi}(x) = \{\varphi(x)\}$, the score set (Definition 3.5) is

$$S_{w,b}(x) \;=\; \{\langle w, z\rangle - b \,:\, z \in \widetilde{\varphi}(x)\} \;=\; \{\langle w, \varphi(x)\rangle - b\}.$$

Let $a := \langle w, \varphi(x)\rangle - b \in \mathbb{R}$. Then by the normalization axiom of the base aggregator B,

$$s^B_{w,b}(x) \;=\; B\big(S_{w,b}(x)\big) \;=\; B(\{a\}) \;=\; a \;=\; \langle w, \varphi(x)\rangle - b.$$

Hence Feature HyperVectors coincide with ordinary feature vectors in this case. □

3.1.3 Feature SuperHyperVector

A feature superhypervector extends feature hypervectors to superhypervector spaces, modeling hierarchical, nested, and multi-layered attribute structures for advanced learning systems.

Remark 3.8. Let the feature space be the real vector space $V := \mathbb{R}^d$ with the usual addition $+$. For each input object $x \in \mathcal{X}$, let $\Phi_1(x) \subseteq V$ be a *nonempty* set of admissible level-1 feature vectors (e.g., produced by alternative tokenizers or encoders). Throughout, write $\mathcal{P}_n(\,\cdot\,)$ for the n-th iterated powerset, and $\mathcal{P}^*_n(\,\cdot\,)$ for its nonempty part.

Definition 3.9 (Coordinatewise gate monoids and operator set)**.** For each coordinate $j \in \{1, \ldots, d\}$ choose a nonempty set $R_j \subseteq \mathbb{R}_{\geq 0}$ such that

$$0 \in R_j, \qquad 1 \in R_j, \qquad r, s \in R_j \Rightarrow rs \in R_j.$$

Thus R_j is a multiplicative submonoid of $\mathbb{R}_{\geq 0}$ containing 0 and 1. Define the gate space $\mathcal{R} := R_1 \times \cdots \times R_d \subseteq \mathbb{R}^d_{\geq 0}$, with componentwise product $(r \odot s)_j = r_j s_j$. For $r \in \mathcal{R}$, write $D_r := \mathrm{diag}(r_1, \ldots, r_d) \in \mathbb{R}^{d\times d}$. Let

$$\mathfrak{G} \;:=\; \{\, D_r \mid r \in \mathcal{R}\,\}$$

be the set of admissible diagonal gate operators. Then $\mathfrak{G}$ is closed under matrix multiplication and contains the identity I and the zero operator 0.

Definition 3.10 (Level-n nesting)**.** For $n \geq 1$, define $\mathrm{nest}_n : V \to \mathcal{P}_n(V)$ by

$$\mathrm{nest}_1(u) := \{u\}, \qquad \mathrm{nest}_{k+1}(u) := \{\mathrm{nest}_k(u)\} \quad (k \geq 1).$$

Thus $\mathrm{nest}_n(u)$ is the n-fold singleton tower whose leaf is $u \in V$.

Definition 3.11 (Level-n sum and action lifting)**.** For $U, W \in \mathcal{P}_n^*(V)$, define the level-$n$ sum by

$$U + W := \left\{ \mathrm{nest}_n(u + v) \;\middle|\; \mathrm{leaf}(U) = u,\ \mathrm{leaf}(W) = v \right\}.$$

For $a \in \mathbb{R}^m$ and $U \in \mathcal{P}_n^*(V)$, lift the action by

$$\circ^{(m,n)}(a, U) := \bigcup_{\mathrm{leaf}(U)=u} \circ^{(m,n)}(a, u).$$

Definition 3.12 (Scalarization map)**.** For $a = (a_1, \ldots, a_m) \in \mathbb{R}^m$, set

$$\sigma(a) := \sum_{j=1}^{m} a_j \in \mathbb{R}.$$

Definition 3.13 ((m, n)-feature superhyperoperation)**.** Fix integers $m, n \geq 1$. Define

$$\circ^{(m,n)} : \mathbb{R}^m \times V \longrightarrow \mathcal{P}_n^*(V), \qquad \circ^{(m,n)}(a, z) := \left\{ \mathrm{nest}_n\big(G\big(\sigma(a)\, z\big)\big) \;\middle|\; G \in \mathfrak{G} \right\}.$$

Comment: Only the scalarization $\sigma(a) = \sum_{j=1}^{m} a_j$ of the m-tuple a acts on z. This additive choice ensures linearity in a (cf. Theorem3.16).

Definition 3.14 (Feature SuperHyperVector (as an n-level realization set))**.** For any $x \in \mathcal{X}$ define the n-level Feature SuperHyperVector of x by

$$\widetilde{\Phi}_n(x) := \left\{ \mathrm{nest}_n\big(G\, z\big) \;\middle|\; z \in \Phi_1(x),\ G \in \mathfrak{G} \right\} \in \mathcal{P}_n^*(V).$$

Each element of $\widetilde{\Phi}_n(x)$ is an n-nested singleton tower whose leaf is a gated feature vector Gz.

Example 3.15 (Real-world example: Wearable fall detection with sensor–reliability gates)**.** Let the feature space be $V = \mathbb{R}^5$ with coordinates

$$(\mathrm{acc\_mean},\ \mathrm{acc\_var},\ \mathrm{gyro\_mean},\ \mathrm{gyro\_var},\ \mathrm{step\_rate}).$$

Consider a subject x wearing an IMU. Two alternative level-1 encoders provide admissible features

$$\Phi_1(x) = \left\{ z^{(\mathrm{time})},\ z^{(\mathrm{freq})} \right\}, \qquad z^{(\mathrm{time})} = \begin{bmatrix} 1.20 \\ -0.80 \\ 0.50 \\ 2.10 \\ -0.30 \end{bmatrix}, \quad z^{(\mathrm{freq})} = \begin{bmatrix} 0.90 \\ -0.60 \\ 0.40 \\ 1.70 \\ 0.10 \end{bmatrix}.$$

To reflect sensor reliability (dropout or attenuation flags) on each coordinate, choose multiplicative submonoids

$$R_1 = R_3 = R_5 = \{0, 1\}, \qquad R_2 = R_4 = \{0\} \cup \{2^{-k} \mid k \in \mathbb{N}_0\} \subset \mathbb{R}_{\geq 0},$$

and set $\mathcal{R} = R_1 \times \cdots \times R_5$. For $r = (r_1, \ldots, r_5) \in \mathcal{R}$, let $D_r = \mathrm{diag}(r_1, \ldots, r_5)$ and $\mathfrak{G} = \{D_r : r \in \mathcal{R}\}$.

Fix the nesting depth $n = 3$. By Definition 3.14,

$$\widetilde{\Phi}_3(x) = \left\{ \mathrm{nest}_3(Gz) \;\middle|\; z \in \Phi_1(x),\ G \in \mathfrak{G} \right\} = \left\{ \{\{\{Gz\}\}\} \;\middle|\; z \in \Phi_1(x),\ r \in \mathcal{R},\ G = D_r \right\}.$$

Three concrete realizations in $\widetilde{\Phi}_3(x)$ (obtained from real reliability patterns) are:

$$\begin{aligned}
&\text{(i) No attenuation: } r^{(1)} = (1,1,1,1,1) \Rightarrow G^{(1)} = I, \\
&u_1 = G^{(1)} z^{(\mathrm{time})} = z^{(\mathrm{time})} = \begin{bmatrix} 1.20 \\ -0.80 \\ 0.50 \\ 2.10 \\ -0.30 \end{bmatrix}, \qquad \mathrm{nest}_3(u_1) = \{\ \{\ \{u_1 \\
&\text{(ii) Mild noise on variance channels, gyro_mean muted: } r^{(2)} = \left(1, \tfrac{1}{2}, 0, \tfrac{1}{2}, 1\right) \Rightarrow G^{(2)} = \mathrm{diag}\left(1, \tfrac{1}{2}, 0, \tfrac{1}{2}, 1\right), \\
&u_2 = G^{(2)} z^{(\mathrm{time})} = \begin{bmatrix} 1.20 \\ -0.80 \times \frac{1}{2} \\ 0.50 \times 0 \\ 2.10 \times \frac{1}{2} \\ -0.30 \times 1 \end{bmatrix} = \begin{bmatrix} 1.20 \\ -0.40 \\ 0.00 \\ 1.05 \\ -0.30 \end{bmatrix}, \qquad \mathrm{nest}_3(u_2) = \{ \\
&\text{(iii) Dropped acc_mean/gyro_var, strong attenuation on acc_var: } r^{(3)} = \left(0, \tfrac{1}{4}, 1, 0, 1\right) \Rightarrow G^{(3)} = \mathrm{diag}\left(0, \tfrac{1}{4}, 1, 0, 1\right), \\
&u_3 = G^{(3)} z^{(\mathrm{freq})} = \begin{bmatrix} 0.90 \times 0 \\ -0.60 \times \frac{1}{4} \\ 0.40 \times 1 \\ 1.70 \times 0 \\ 0.10 \times 1 \end{bmatrix} = \begin{bmatrix} 0.00 \\ -0.15 \\ 0.40 \\ 0.00 \\ 0.10 \end{bmatrix}, \qquad \mathrm{nest}_3(u_3) = \{
\end{aligned}$$

Interpretation. The diagonal gates D_r encode coordinatewise reliability: entries equal to 0 mute unusable signals (e.g., saturated sensor), entries 2^{-k} down-weight noisy channels, and entries 1 pass through trusted measurements. The set $\widetilde{\Phi}_3(x)$ thus collects all gated realizations of either admissible encoder, each represented as a threefold singleton tower whose leaf is the concrete gated feature vector.

Theorem 3.16 ((m,n)-SuperHyperVector structure)**.** *Let $V = \mathbb{R}^d$ with the usual addition $+$, and let $\circ^{(m,n)}$ be as in Definition3.13. Then for all $x, y \in V$ and $a, b \in \mathbb{R}^m$, writing $\alpha := \sigma(a)$, $\beta := \sigma(b)$,*

$$\begin{aligned}
&\text{(SV1)} \quad \circ^{(m,n)}(a, x+y) \subseteq \circ^{(m,n)}(a,x) + \circ^{(m,n)}(a,y), \\
&\text{(SV2)} \quad \circ^{(m,n)}(a+b, x) \subseteq \circ^{(m,n)}(a,x) + \circ^{(m,n)}(b,x), \\
&\text{(SV3)} \quad \circ^{(m,n)}\left(a,\ \circ^{(m,n)}(b,x)\right) = \left\{ \mathrm{nest}_n\left(K\left((\alpha\beta)x\right)\right) \;\middle|\; K \in \mathfrak{G} \right\}, \\
&\text{(SV4)} \quad \circ^{(m,n)}(-a, x) = -\circ^{(m,n)}(a,x), \\
&\text{(SV5)} \quad x \in \circ^{(m,n)}\left((1,\dots,1),\ x\right) \quad \textit{(as the leaf of a nested singleton).}
\end{aligned}$$

Proof. We repeatedly use: linearity of each $G \in \mathfrak{G}$; closure $G, H \in \mathfrak{G} \Rightarrow GH \in \mathfrak{G}$; and for all $u, v \in V$ and $n \geq 1$,

$$\mathrm{nest}_n(u) + \mathrm{nest}_n(v) = \{\mathrm{nest}_n(u+v)\} \quad \text{(singletons at each level).}$$

(SV1) For any $G \in \mathfrak{G}$,

$$\mathrm{nest}_n\left(G(\alpha(x+y))\right) = \mathrm{nest}_n\left(G(\alpha x) + G(\alpha y)\right) \in \mathrm{nest}_n(G\alpha x) + \mathrm{nest}_n(G\alpha y).$$

Taking the union over G gives the inclusion.

(SV2) Since $\sigma(a+b) = \sigma(a) + \sigma(b) = \alpha + \beta$,

$$\mathrm{nest}_n\big(G\big((\alpha+\beta)x\big)\big) = \mathrm{nest}_n\big(G(\alpha x) + G(\beta x)\big) \in \mathrm{nest}_n(G\alpha x) + \mathrm{nest}_n(G\beta x),$$

and union over G yields the claim.

(SV3) Unfolding,

$$\circ^{(m,n)}(b, x) = \{\mathrm{nest}_n(H(\beta x)) \mid H \in \mathfrak{G}\}.$$

Applying $\circ^{(m,n)}(a, \cdot)$ to each leaf $H(\beta x)$ gives

$$\mathrm{nest}_n\big(G(\alpha\, H(\beta x))\big) = \mathrm{nest}_n\big((GH)((\alpha\beta)x)\big),$$

and the set of all such leaves is $\{\mathrm{nest}_n(K((\alpha\beta)x)) \mid K \in \mathfrak{G}\}$ by closure.

(SV4) For any $G \in \mathfrak{G}$,

$$\mathrm{nest}_n\big(G(\sigma(-a)\,x)\big) = \mathrm{nest}_n\big(G(-\alpha x)\big) = \mathrm{nest}_n\big(-\,G(\alpha x)\big) = -\,\mathrm{nest}_n\big(G(\alpha x)\big).$$

(SV5) Since $I \in \mathfrak{G}$ and $\sigma(1, \dots, 1) = m$, we have $\mathrm{nest}_n(I \cdot mx) \in \circ^{(m,n)}((1, \dots, 1), x)$, hence x appears at the leaf after rescaling by m^{-1} inside V if desired (the leaf is linear in x). □

3.2 Support Vector, HyperVector, and SuperHyperVector

3.2.1 Support Vector

A support vector is a data point lying closest to a decision boundary in Support Vector Machines, defining the position of the margin [62–67]. Related concepts include the fuzzy support vector [68–74] and the neutrosophic support vector [75, 76].

Definition 3.17 (Training data and linear decision function)**.** Let $\{(x_i, y_i)\}_{i=1}^{n}$ be a labeled sample with $x_i \in \mathbb{R}^p$ and $y_i \in \{-1, +1\}$. A linear decision function is $f(x) := w^\top x - b$ with $w \in \mathbb{R}^p$ and $b \in \mathbb{R}$ (cf. [77, 78]).

Definition 3.18 (Hard–margin SVM and support vectors)**.** In the linearly separable case, the (hard–margin) SVM (cf. [79–81]) solves

$$\min_{w,b} \ \tfrac{1}{2}\|w\|^2 \quad \text{s.t.} \quad y_i\,(w^\top x_i - b) \ \geq \ 1 \quad (i = 1, \dots, n).$$

Let $(w^\star, b^\star)$ be any optimal solution. A training point x_i is a *support vector* iff it lies on the margin:

$$y_i\,(w^{\star\top} x_i - b^\star) \ = \ 1.$$

Equivalently, in the dual, if $\alpha_i > 0$ for the optimal multipliers $\alpha = (\alpha_i)_{i=1}^{n}$.

Definition 3.19 (Soft–margin SVM and support vectors)**.** For general (not necessarily separable) data, the (soft–margin) SVM [82–84] solves

$$\min_{w,b,\xi \geq 0} \ \tfrac{1}{2}\|w\|^2 + C \sum_{i=1}^{n} \xi_i \quad \text{s.t.} \quad y_i\,(w^\top x_i - b) \ \geq \ 1 - \xi_i \quad (i = 1, \dots, n),$$

with $C > 0$. Let $(w^\star, b^\star, \xi^\star)$ be optimal and let $\alpha^\star = (\alpha_i^\star)_{i=1}^{n}$ be optimal dual variables. A training point x_i is a *support vector* iff its dual weight is positive:

$$\alpha_i^\star \ > \ 0.$$

Equivalently (primal view), every support vector satisfies

$$y_i\, f^\star(x_i) \ \leq \ 1, \qquad f^\star(x) := w^{\star\top} x - b^\star,$$

i.e., it lies *on* the margin ($= 1$) or *inside* it (< 1).

3.2.2 Support HyperVector

A support hypervector extends support vectors to hypervector spaces, representing multidimensional boundary-defining elements in hypergraph-based or high-dimensional learning models.

Definition 3.20 (Setting and realizations of a hypervector)**.** Let $(V, +, \circ, K)$ be a hypervector space over a field K (in the sense given earlier), and assume $K = \mathbb{R}$ so that V is a real vector space endowed with an inner product $\langle \cdot, \cdot \rangle$. For $x \in V$, its *realization set* is

$$\mathcal{R}(x) \;:=\; 1 \circ x \;\subseteq V,$$

which is nonempty by the axiom $x \in 1 \circ x$. We additionally use the shorthand $\mathcal{R}(A) := \bigcup_{x \in A} \mathcal{R}(x)$ for $A \subseteq V$.

Example 3.21 (Setting and realizations of a hypervector: sensor noise as a box)**.** Let $V = \mathbb{R}^2$ and fix a tolerance vector $\varepsilon = (0.2,\ 0.1)^\top$. Define a hypervector space $(V, +, \circ, \mathbb{R})$ by the scalar hyperoperation

$$a \circ x \;:=\; \{\, a\,x + \delta \;\mid\; \|\delta\|_\infty \le |a|\,\varepsilon \,\} \quad \text{(componentwise: } |\delta_j| \le |a|\,\varepsilon_j).$$

Then, for any $x \in V$, the realization set is

$$\mathcal{R}(x) \;=\; 1 \circ x \;= \{\, x + \delta \mid \|\delta\|_\infty \le \varepsilon \,\}.$$

Concrete instance. Take $x = (2,\ -1)^\top$. Then

$$\mathcal{R}(x) \;=\; \big\{ (u_1, u_2)^\top \in \mathbb{R}^2 \;\big|\; |u_1 - 2| \le 0.2,\ |u_2 + 1| \le 0.1 \big\}.$$

For example, with $\delta = (0.1,\ -0.08)^\top$ (which satisfies $|\delta_1| \le 0.2$, $|\delta_2| \le 0.1$), we have one realization $u = x + \delta = (2.1,\ -1.08)^\top \in \mathcal{R}(x)$. Thus $\mathcal{R}(x)$ models all admissible sensor–noise perturbations of x within an ℓ_∞ box.

Definition 3.22 (Hyperlinear score set and aggregated margin)**.** Fix a pair $(w, b) \in V \times \mathbb{R}$ (a linear functional $u \mapsto \langle w, u \rangle - b$). For a labeled example (x, y) with $x \in V$ and $y \in \{-1, +1\}$, the *hyperlinear score set* is

$$S_{w,b}(x) \;:=\; \big\{\, \langle w, u \rangle - b \;\big|\; u \in \mathcal{R}(x) \,\big\} \;\subseteq \mathbb{R},$$

and its *signed margin set* is $M_{w,b}(x, y) := \{\, y\,s \mid s \in S_{w,b}(x) \,\} \subseteq \mathbb{R}$. Let $B : \mathcal{P}_*(\mathbb{R}) \to \mathbb{R}$ be any monotone *aggregator* with $B(\{a\}) = a$ (e.g. $B = \inf$ (robust), $B = \sup$ (optimistic), or an average). The *B–aggregated margin* is

$$\mathrm{margin}_B(w, b \mid x, y) \;:=\; B\big(M_{w,b}(x, y)\big).$$

Example 3.23 (Hyperlinear score set and aggregated margin: closed forms)**.** Let $(w, b) \in V \times \mathbb{R}$ and labeled (x, y) with $y \in \{-1, +1\}$. For the box realization $\mathcal{R}(x) = \{x + \delta : \ |\delta_j| \le \varepsilon_j\}$, the hyperlinear score set and signed margin set are intervals:

$$S_{w,b}(x) = \big[\ \langle w, x \rangle - b - \|\,|w| \odot \varepsilon\,\|_1,\ \ \langle w, x \rangle - b + \|\,|w| \odot \varepsilon\,\|_1\ \big],$$

$$M_{w,b}(x, y) = \big[\ y(\langle w, x \rangle - b) - \|\,|w| \odot \varepsilon\,\|_1,\ \ y(\langle w, x \rangle - b) + \|\,|w| \odot \varepsilon\,\|_1\ \big].$$

Hence the $B_{\min}/B_{\max}$ aggregated margins are

$$\mathrm{margin}_{B_{\min}}(w, b \mid x, y) = y(\langle w, x \rangle - b) - \|\,|w| \odot \varepsilon\,\|_1, \qquad \mathrm{margin}_{B_{\max}}(w, b \mid x, y) = y(\langle w, x \rangle - b) + \|\,|w| \odot \varepsilon\,\|_1.$$

Numbers. Let $x = (2,\ -1)^\top$, $\varepsilon = (0.2,\ 0.1)^\top$, $w = (1,\ -3)^\top$, $b = 0.5$, $y = +1$. Compute

$$\langle w, x \rangle = 1 \cdot 2 + (-3) \cdot (-1) = 5, \qquad \|\,|w| \odot \varepsilon\,\|_1 = |1| \cdot 0.2 + |-3| \cdot 0.1 = 0.5,$$

so

$$S_{w,b}(x) = [\,5 - 0.5 - 0.5,\ \ 5 - 0.5 + 0.5\,] = [\,4.0,\ 5.0\,],$$

$$\mathrm{margin}_{B_{\min}}(w, b \mid x, y) = 4.0, \qquad \mathrm{margin}_{B_{\max}}(w, b \mid x, y) = 5.0.$$

Thus $B_{\min}$ gives the robust (worst–case) margin and $B_{\max}$ the optimistic (best–case) margin.

Definition 3.24 (Support HyperVector (hard and soft variants))**.** Given a training set $\{(x_i, y_i)\}_{i=1}^n \subseteq V \times \{-1, +1\}$ and an aggregator B as in Definition 3.22:

(a) **Hard–margin B–SVM on hypervectors** is the convex program

$$\min_{w,b} \tfrac{1}{2}\|w\|^2 \quad \text{s.t.} \quad \mathrm{margin}_B(w, b \mid x_i, y_i) \;\geq\; 1 \quad (i = 1, \ldots, n).$$

Any optimal $(w^\star, b^\star)$ defines the set of *B–support hypervectors*

$$\mathsf{SV}_B^{\text{hard}} \;:=\; \big\{\, i : \; \mathrm{margin}_B(w^\star, b^\star \mid x_i, y_i) = 1 \,\big\}.$$

(b) **Soft–margin B–SVM on hypervectors** (with $C > 0$) is

$$\min_{w,b,\ \xi\geq 0} \tfrac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \xi_i \quad \text{s.t.} \quad \mathrm{margin}_B(w, b \mid x_i, y_i) \;\geq\; 1 - \xi_i \quad (i = 1, \ldots, n).$$

For any optimal $(w^\star, b^\star, \xi^\star)$ we define the *B–support hypervectors*

$$\mathsf{SV}_B^{\text{soft}} \;:=\; \big\{\, i : \; \mathrm{margin}_B(w^\star, b^\star \mid x_i, y_i) \;\leq\; 1 \,\big\}.$$

Two canonical choices are:

$$B_{\min}(S) = \inf S \quad (\textit{robust} \text{ / worst–case}), \qquad B_{\max}(S) = \sup S \quad (\textit{optimistic} \text{ / best–case}).$$

In words, a *Support HyperVector* is a training hypervector whose aggregated signed margin is tight (equals the margin threshold) or violates it (in the soft case), generalizing the classical notion of support vectors to hypervector realizations.

Example 3.25 (Support HyperVector (hard/soft): tiny robust SVM by hand)**.** We use $B_{\min}$ (robust aggregation) and the box realization $\mathcal{R}(x) = \{x + \delta : \ |\delta_j| \leq \varepsilon_j\}$ with $\varepsilon = (0.1,\ 0.1)^\top$.

(i) Hard margin (feasible and tight). Training set with two points:

$$(x_1, y_1) = ((1,\ 1)^\top,\ +1), \qquad (x_2, y_2) = ((-1,\ -1)^\top,\ -1).$$

Consider $(w, b) = ((1,\ 1)^\top,\ 0.8)$. Since $\|\,|w| \odot \varepsilon\,\|_1 = 0.1 + 0.1 = 0.2$, the robust margins are

$$\mathrm{margin}_{B_{\min}}(w, b \mid x_1, y_1) = +1 \cdot (\langle w, x_1\rangle - b) - 0.2 = (2 - 0.8) - 0.2 = 1.0,$$

$$\mathrm{margin}_{B_{\min}}(w, b \mid x_2, y_2) = -1 \cdot (\langle w, x_2\rangle - b) - 0.2 = -1 \cdot (-2 - 0.8) - 0.2 = 2.6.$$

Both constraints ≥ 1 hold, and x_1 is *exactly* on the robust margin; hence

$$\mathsf{SV}_{B_{\min}}^{\text{hard}} = \{\, 1 \,\}.$$

(ii) Soft margin (violator becomes support). Add a third point

$$(x_3, y_3) = ((0,\ 0.2)^\top,\ +1).$$

With the same (w, b),

$$\mathrm{margin}_{B_{\min}}(w, b \mid x_3, y_3) = +1 \cdot (\langle w, x_3\rangle - b) - 0.2 = (0.2 - 0.8) - 0.2 = -0.8.$$

In the soft–margin problem $\min_{w,b,\xi\geq 0} \frac{1}{2}\|w\|^2 + C\sum_i \xi_i$ s.t. $\mathrm{margin}_{B_{\min}}(w, b \mid x_i, y_i) \geq 1 - \xi_i$, the optimal slack for $i = 3$ must satisfy $\xi_3^\star \geq 1 - (-0.8) = 1.8$. By the soft support criterion (tight or violated), the support set is

$$\mathsf{SV}_{B_{\min}}^{\text{soft}} = \big\{\, i : \; \mathrm{margin}_{B_{\min}}(w^\star, b^\star \mid x_i, y_i) \leq 1 \,\big\} \;\supseteq\; \{\, 1,\ 3 \,\},$$

where $i = 1$ is tight (equals 1) and $i = 3$ is a violator (needs slack).

Remark 3.26 (Dual and KKT characterization (robust choice))**.** For $B = \inf$ and compact convex realization sets $\mathcal{R}(x_i)$, the soft–margin problem is a convex semi–infinite program that admits Lagrange multipliers $\alpha_i \in [0, C]$. At optimality one obtains the stationarity $w^\star = \sum_{i=1}^{n} \alpha_i^\star y_i\, u_i^\star$ for some *active realizations* $u_i^\star \in \mathcal{R}(x_i)$ attaining the infimum, and the complementarity $\alpha_i^\star\big(\inf_{u\in\mathcal{R}(x_i)} y_i(\langle w^\star, u\rangle - b^\star) - 1 + \xi_i^\star\big) = 0$, $(C - \alpha_i^\star)\xi_i^\star = 0$. Consequently, $\alpha_i^\star > 0 \;\Rightarrow\; i \in \mathsf{SV}_{\inf}^{\text{soft}}$.

Theorem 3.27 (Reduction to classical support vectors)**.** *If every realization set is a singleton, $\mathcal{R}(x) = \{x\}$ for all x, then the hard/soft B–SVMs above reduce to the standard SVM, and $\mathsf{SV}_B^{\text{hard/soft}}$ coincides with the usual set of support vectors (independently of B).*

Proof. If $\mathcal{R}(x) = \{x\}$ then $S_{w,b}(x) = \{\langle w, x\rangle - b\}$ and $M_{w,b}(x, y) = \{y(\langle w, x\rangle - b)\}$. Any admissible B satisfies $B(\{a\}) = a$, so the constraints become the standard SVM constraints and tightness conditions become identical. □

3.2.3 Support SuperHyperVector

A support superhypervector generalizes support hypervectors to superhypervector spaces, capturing hierarchical and nested boundary-defining structures in multi-layered learning frameworks. We first define an m-ary scalar superhyperoperation with values in the n-th iterated powerset of the data space; the support concept is then induced by a margin functional built on this structure.

Remark 3.28 (Setup and notation)**.** Let $V = \mathbb{R}^p$ with the usual vector addition +. For $n \geq 1$, let $\mathcal{P}_n(V)$ denote the n-th iterated powerset and $\mathcal{P}_n^*(V)$ its nonempty part. For $u \in V$, define the n-level singleton tower

$$\mathrm{nest}_1(u) := \{u\}, \qquad \mathrm{nest}_{k+1}(u) := \{\mathrm{nest}_k(u)\} \quad (k \geq 1).$$

Whenever we add two n-level singleton towers, we use the lifted Minkowski rule

$$\mathrm{nest}_n(u) \ + \ \mathrm{nest}_n(v) \ := \ \big\{\mathrm{nest}_n(u+v)\big\}.$$

(All sets produced below consist of such nested singletons, so this definition suffices.)

Definition 3.29 (Admissible linear transformation monoid)**.** Let $\mathfrak{T} \subseteq \mathrm{End}_{\mathbb{R}}(V)$ be a nonempty set of linear operators satisfying:

$$I \in \mathfrak{T}, \qquad T_1, T_2 \in \mathfrak{T} \ \Rightarrow \ T_1 T_2 \in \mathfrak{T}.$$

Thus $\mathfrak{T}$ is a multiplicative monoid of linear endomorphisms (e.g. coordinate gates, scalings, or other linear pre-processors).

Definition 3.30 ((m, n)-superhyperoperation for support)**.** Fix integers $m, n \geq 1$. Define

$$\circ^{(m,n)} : \ \mathbb{R}^m \times V \longrightarrow \mathcal{P}_n^*(V)$$

by

$$\circ^{(m,n)}\big((a_1, \ldots, a_m), \, x\big) \ := \ \left\{ \mathrm{nest}_n\Big(T\big((\prod_{j=1}^{m} a_j)\, x\big)\Big) \ \Big| \ T \in \mathfrak{T} \right\}.$$

Definition 3.31 (SuperHyperrealization and support score)**.** For any $x \in V$, its n-level *SuperHyperrealization set* is

$$\widetilde{\mathcal{R}}_n(x) \ := \ \circ^{(m,n)}\big((1, \ldots, 1), \, x\big) \ \in \ \mathcal{P}_n^*(V).$$

Given $(w, b) \in V \times \mathbb{R}$ and a base set-aggregator B on nonempty subsets of $\mathbb{R}$ with $B(\{a\}) = a$ (e.g. $B_{\min} = \inf$, $B_{\max} = \sup$), define the n-level score family and its levelwise flattening:

$$\widetilde{S}_n(w, b \mid x) := \left\{ \{\cdots\{\langle w, u\rangle - b : \ u \in U\}\cdots\} \ \Big| \ U \in \widetilde{\mathcal{R}}_n(x) \right\} \subseteq \mathcal{P}_n(\mathbb{R}),$$

and the *SuperHyper–B crisp score* $s_{w,b}^{B,(n)}(x) := \mathrm{crisp}_n^B\big(\widetilde{S}_n(w, b \mid x)\big) \in \mathbb{R}$ (“inner level first, then outward” aggregation). For a labeled point (x, y), $y \in \{-1, +1\}$, the *SuperHyper–B aggregated margin* is

$$\mathrm{margin}_B^{(n)}(w, b \mid x, y) \ := \ y \cdot s_{w,b}^{B,(n)}(x).$$

Definition 3.32 (Support SuperHyperVector)**.** Given training data $\{(x_i, y_i)\}_{i=1}^N \subseteq V \times \{-1, +1\}$ and a choice of aggregator B, consider the large-margin program

$$\min_{w,b} \ \tfrac{1}{2}\|w\|^2 \quad \text{s.t.} \quad \mathrm{margin}_B^{(n)}(w, b \mid x_i, y_i) \ \geq \ 1 \quad (i = 1, \ldots, N).$$

At any optimum $(w^\star, b^\star)$, the indices with equality,

$$\mathsf{SV}_B^{(n)} := \{ i : \operatorname{margin}_B^{(n)}(w^\star, b^\star \mid x_i, y_i) = 1 \},$$

are the *Support SuperHyperVectors*. (The soft-margin variant adds slacks $\xi_i \geq 0$ and uses the tightness criterion ≤ 1.)

Example 3.33 (A concrete Support SuperHyperVector under two–level robustness)**.** Let $V = \mathbb{R}^2$, choose the robust base aggregator $B_{\min}(S) = \inf S$, and fix the label $y = +1$. Consider a single training point $x = (1, 1)^\top$ with an $n = 2$ SuperHyperrealization consisting of two level–1 alternatives:

$$Z_1 = \left\{ u \in \mathbb{R}^2 \;\middle|\; \|u - x\|_\infty \leq r_1 \right\}, \qquad Z_2 = \left\{ \alpha x \;\middle|\; \alpha \in [\alpha_{\min}, \alpha_{\max}] \right\},$$

and the level–2 family $\widetilde{\mathcal{R}}_2(x) = \{ Z_1, Z_2 \}$. Take

$$w = \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \qquad b = -0.6, \qquad r_1 = 0.2, \qquad \alpha_{\min} = 0.8,\ \alpha_{\max} = 1.0.$$

We compute the inner (level–1) robust margins first, then the outer (level–2) minimum.

Inner robust margin on Z_1 (axis–aligned jitter). By standard ℓ_∞ worst–case linearization,

$$\inf_{u \in Z_1} y\,(\langle w, u\rangle - b) = y\,(\langle w, x\rangle - b) - \|w\|_1\, r_1.$$

Since $\langle w, x\rangle = 2 \cdot 1 + (-1) \cdot 1 = 1$ and $\|w\|_1 = |2| + |-1| = 3$,

$$\inf_{u \in Z_1} y\,(\langle w, u\rangle - b) = (1 - (-0.6)) - 3 \cdot 0.2 = 1.6 - 0.6 = 1.0.$$

Inner robust margin on Z_2 (global rescale). For $\alpha \in [\alpha_{\min}, \alpha_{\max}]$,

$$\inf_{u \in Z_2} y\,(\langle w, u\rangle - b) = \inf_{\alpha \in [\alpha_{\min}, \alpha_{\max}]} \big(\alpha\, y\, \langle w, x\rangle - b\big) = \alpha_{\min}\, \langle w, x\rangle - b = 0.8 \cdot 1 - (-0.6) = 1.4.$$

Outer (level–2) aggregation. With $B_{\min}$ at the outer level,

$$\operatorname{margin}_{B_{\min}}^{(2)}(w, b \mid x, y) = \min\left\{ \inf_{u \in Z_1} y(\langle w, u\rangle - b),\ \inf_{u \in Z_2} y(\langle w, u\rangle - b) \right\} = \min\{ 1.0,\ 1.4 \} = 1.0.$$

Under the robust ($B_{\min}$) two–level SuperHyperrealization $\widetilde{\mathcal{R}}_2(x)$, the aggregated margin of (x, y) equals the margin threshold 1. Hence, at any hard–margin optimum $(w^\star, b^\star) = (w, b)$ for which these constraints are active, (x, y) is a *Support SuperHyperVector* in the sense of Definition 3.32.

Theorem 3.34 (Support SuperHyperVector as an (m, n)-SuperHyperVector)**.** *With $V = \mathbb{R}^p$, the usual addition $+$, base field $\mathbb{R}$, and the superhyperoperation $\circ^{(m,n)}$ from Definition 3.30, the quadruple*

$$\big(V,\ +,\ \circ^{(m,n)},\ \mathbb{R}\big)$$

is an (m, n)-SuperHyperVector space. Writing $\Pi\mathbf{a} := \prod_{j=1}^m a_j$ for $\mathbf{a} = (a_1, \ldots, a_m)$, the following axioms hold for all $x, y \in V$ and $\mathbf{a}, \mathbf{b} \in \mathbb{R}^m$:

$$\begin{aligned}
&\text{(SV1)}\ \circ^{(m,n)}(\mathbf{a}, x + y) \subseteq \circ^{(m,n)}(\mathbf{a}, x) + \circ^{(m,n)}(\mathbf{a}, y),\\
&\text{(SV2)}\ \circ^{(m,n)}(\mathbf{a} + \mathbf{b}, x) \subseteq \circ^{(m,n)}(\mathbf{a}, x) + \circ^{(m,n)}(\mathbf{b}, x),\\
&\text{(SV3)}\ \circ^{(m,n)}\big(\mathbf{a}, \circ^{(m,n)}(\mathbf{b}, x)\big) = \circ^{(m,n)}(\mathbf{a}{\cdot}\mathbf{b}, x),\\
&\text{(SV4)}\ \circ^{(m,n)}(-\mathbf{a}, x) = -\circ^{(m,n)}(\mathbf{a}, x),\\
&\text{(SV5)}\ x \in \circ^{(m,n)}\big((1, \ldots, 1), x\big) \quad \textit{(as the leaf of a nested singleton).}
\end{aligned}$$

Proof. Fix $x, y \in V$, $\mathbf{a}, \mathbf{b} \in \mathbb{R}^m$. Let $T \in \mathfrak{T}$ be arbitrary. Since T is linear and $\mathfrak{T}$ is closed under composition:

(SV1) $\mathrm{nest}_n\Big(T\big(\Pi\mathbf{a}\,(x+y)\big)\Big) = \mathrm{nest}_n\Big(T(\Pi\mathbf{a}\,x) + T(\Pi\mathbf{a}\,y)\Big) \in \mathrm{nest}_n\big(T(\Pi\mathbf{a}\,x)\big) + \mathrm{nest}_n\big(T(\Pi\mathbf{a}\,y)\big)$. Taking the union over T yields the inclusion.

(SV2) Using $\Pi(\mathbf{a}+\mathbf{b})\,x = (\Pi\mathbf{a})\,x + (\Pi\mathbf{b})\,x$ at the scalar level, the same argument as (SV1) applies.

(SV3) Unfold definitions: $\circ^{(m,n)}(\mathbf{b}, x) = \{\mathrm{nest}_n(H(\Pi\mathbf{b}\,x)) \mid H \in \mathfrak{T}\}$. Applying $\circ^{(m,n)}(\mathbf{a}, \cdot)$ maps each leaf $H(\Pi\mathbf{b}\,x)$ to $\mathrm{nest}_n\big(G(\Pi\mathbf{a}\,H(\Pi\mathbf{b}\,x))\big) = \mathrm{nest}_n\big((GH)(\Pi(\mathbf{a}\cdot\mathbf{b})\,x)\big)$. Since $G, H \in \mathfrak{T} \Rightarrow GH \in \mathfrak{T}$ and every element of $\mathfrak{T}$ arises as such a product, the image equals $\{\mathrm{nest}_n(K(\Pi(\mathbf{a}\cdot\mathbf{b})\,x)) \mid K \in \mathfrak{T}\} = \circ^{(m,n)}(\mathbf{a}\cdot\mathbf{b}, x)$.

(SV4) Linearity gives $\mathrm{nest}_n\big(T(\Pi(-\mathbf{a})\,x)\big) = \mathrm{nest}_n\big(-T(\Pi\mathbf{a}\,x)\big) = -\mathrm{nest}_n\big(T(\Pi\mathbf{a}\,x)\big)$, hence $\circ^{(m,n)}(-\mathbf{a}, x) = -\circ^{(m,n)}(\mathbf{a}, x)$.

(SV5) Since $I \in \mathfrak{T}$, $\mathrm{nest}_n(I \cdot x) = \mathrm{nest}_n(x) \in \circ^{(m,n)}((1, \ldots, 1), x)$, so x appears as the leaf. □

3.3 Relevance Vector, HyperVector, and SuperHyperVector

3.3.1 Relevance Vector

A relevance vector is a sparse subset of training samples with nonzero weights in a kernel model, selected through automatic relevance determination Bayesian inference [85–89]. Like the support vector, the relevance vector plays a highly significant role in the field of AI, and related concepts such as the fuzzy relevance vector [90–92] are also well known.

Definition 3.35 (Relevance vector)**.** (cf. [93–95]) Let $\{(x_j, t_j)\}_{j=1}^N$ be training data with $x_j \in \mathbb{R}^d$ and targets t_j. Fix basis functions $\varphi_i(x) := k(x, x_i)$ built from a positive–definite kernel k. Consider the linear model

$$f(x) = \sum_{i=1}^{N} w_i\, \varphi_i(x),$$

with (independent) automatic–relevance–determination (ARD) Gaussian prior

$$p(w \mid \alpha) = \mathcal{N}\big(0, A^{-1}\big), \qquad A = \mathrm{diag}(\alpha_1, \ldots, \alpha_N), \;\; \alpha_i > 0,$$

and a standard likelihood (e.g. Gaussian for regression, logistic/probit for classification). Let $\alpha^\star = (\alpha_1^\star, \ldots, \alpha_N^\star)$ denote any maximizer of the marginal likelihood (Type–II ML). Then many precisions diverge, $\alpha_i^\star \to \infty$, which effectively forces $w_i \to 0$.

The *relevance index set* is

$$\mathcal{R} := \big\{ i \in \{1, \ldots, N\} \;:\; \alpha_i^\star < \infty \big\},$$

equivalently (in regression) those with nonzero posterior mean weight $\mu_i^\star \neq 0$. The corresponding training inputs

$$\{ x_i \;:\; i \in \mathcal{R} \}$$

are called the *relevance vectors*. The predictive function uses only them:

$$f(x) = \sum_{i \in \mathcal{R}} \mu_i^\star\, k(x, x_i),$$

and, for probabilistic classification with a sigmoid link σ, one has $p(t = 1 \mid x) \approx \sigma\big(f(x)\big)$.

3.3.2 Relevance HyperVector

We now lift the classical *Relevance Vector* to a set–valued (hyper) representation that (i) encodes on/off (or shrunk) basis selection in a mathematically explicit way, (ii) reduces to the standard RVM when the gates are fixed, and (iii) induces a bona fide *hypervector space* structure.

Definition 3.36 (Relevance HyperVector)**.** Let $\{(x_j, t_j)\}_{j=1}^{N}$ be training data, k a positive–definite kernel, and $\varphi_i(x) := k(x, x_i)$ the i-th basis function. Let $\mu = (\mu_1, \ldots, \mu_N)^\top$ be a fixed weight vector (e.g. the posterior mean from sparse Bayesian learning/RVM).

For each index i, choose a nonempty *relevance gate set* $R_i \subseteq \mathbb{R}_{\geq 0}$ such that

$$0 \in R_i, \qquad 1 \in R_i, \qquad \text{and} \qquad r, s \in R_i \;\Rightarrow\; rs \in R_i,$$

(i.e. R_i is a multiplicative submonoid of $\mathbb{R}_{\geq 0}$ containing 0 and 1). Define the product gate set

$$\mathcal{R} \;:=\; R_1 \times \cdots \times R_N$$

with componentwise multiplication, and for $r = (r_1, \ldots, r_N) \in \mathcal{R}$ write $D_r := \operatorname{diag}(r_1, \ldots, r_N)$. Let $V := \mathbb{R}^N$ with the usual addition $+$.

(A) Hyperoperation. Define the scalar hyperoperation $\circ : \mathbb{R} \times V \to \mathcal{P}^*(V)$ by

$$a \circ z \;:=\; \big\{ D_r\,(az) \;\big|\; r \in \mathcal{R} \big\}, \qquad a \in \mathbb{R},\ z \in V,$$

where $\mathcal{P}^*(V)$ denotes the collection of nonempty subsets of V.

(B) Relevance HyperVector at input x. Let $\varphi(x) := \big(\varphi_1(x), \ldots, \varphi_N(x)\big)^\top \in \mathbb{R}^N$ and let $\odot$ denote the Hadamard (componentwise) product. Define the *base contribution vector*

$$v(x) \;:=\; \mu \odot \varphi(x) \;\in\; \mathbb{R}^N.$$

The *Relevance HyperVector* of x is the realization set

$$\widetilde{v}(x) \;:=\; 1 \circ v(x) \;=\; \big\{ D_r\, v(x) \;\big|\; r \in \mathcal{R} \big\} \;\subseteq \mathbb{R}^N.$$

Each $u \in \widetilde{v}(x)$ collects the (possibly gated/shrunk) per–basis contributions $u_i = r_i\, \mu_i\, \varphi_i(x)$. The induced *score set* is

$$S(x) \;:=\; \big\{ \langle \mathbf{1}, u \rangle \;\big|\; u \in \widetilde{v}(x) \big\} \;=\; \Big\{ \sum_{i=1}^{N} r_i\, \mu_i\, \varphi_i(x) \;\Big|\; r \in \mathcal{R} \Big\},$$

where $\mathbf{1} = (1, \ldots, 1)^\top$.

Example 3.37 (A concrete Relevance HyperVector)**.** Let $d = 1$, $N = 3$. Use the Gaussian kernel with unit lengthscale

$$k(x, y) \;=\; \exp\!\Big(-\tfrac{(x-y)^2}{2}\Big).$$

Fix training inputs $x_1 = 0,\ x_2 = 1,\ x_3 = 2$, and a weight vector

$$\mu = \big(1.0,\ -0.5,\ 0.8\big)^\top.$$

For the query input $x = 1.5$, the basis values are

$$\varphi_1(x) = e^{-1.125} \approx 0.3246524674, \quad \varphi_2(x) = e^{-0.125} \approx 0.8824969026, \quad \varphi_3(x) = e^{-0.125} \approx 0.8824969026.$$

Hence the base contribution vector $v(x) = \mu \odot \varphi(x) \in \mathbb{R}^3$ is

$$v(x) = \big(\, 1.0 \cdot 0.3246524674,\;\; (-0.5) \cdot 0.8824969026,\;\; 0.8 \cdot 0.8824969026 \,\big)$$

$$= \big(\, 0.3246524674,\; -0.4412484513,\; 0.7059975221 \,\big).$$

Choose relevance gate sets (each a multiplicative submonoid containing 0 and 1):

$$R_1 = \{0, 1\}, \qquad R_2 = [0, 1], \qquad R_3 = \{0, 1\},$$

and define $\mathcal{R} = R_1 \times R_2 \times R_3$. For $r = (r_1, r_2, r_3) \in \mathcal{R}$, let $D_r = \mathrm{diag}(r_1, r_2, r_3)$.

By Definition 3.36, the Relevance HyperVector at x is

$$\widetilde{v}(x) = \{\, D_r\, v(x) \mid r \in \mathcal{R} \,\} = \Big\{ \big(r_1{\cdot}0.3246524674,\;\; r_2{\cdot}(-0.4412484513),\;\; r_3{\cdot}0.7059975221\big) \;\Big|\; r_1, r_3 \in \{0, 1\},\, r_2 \in [0, 1] \Big\}.$$

Three explicit realizations $u = D_r v(x) \in \widetilde{v}(x)$ are:

$$\begin{aligned}
r = (1,\ 1,\ 1): &\quad u = \big(0.3246524674,\ -0.4412484513,\ 0.7059975221\big),\\
r = (1,\ 0.30,\ 0): &\quad u = \big(0.3246524674,\ -0.1323745354,\ 0\big) \quad (\text{since } -0.4412484513 \times 0.30 = -0.1323745354),\\
r = (0,\ 0.75,\ 1): &\quad u = \big(0,\ -0.3309363385,\ 0.7059975221\big) \quad (\text{since } -0.4412484513 \times 0.75 = -0.3309363385).
\end{aligned}$$

The induced score set

$$S(x) = \Big\{ \langle \mathbf{1}, u\rangle \;\Big|\; u \in \widetilde{v}(x) \Big\} = \Big\{ r_1{\cdot}0.3246524674 + r_2{\cdot}(-0.4412484513) + r_3{\cdot}0.7059975221 \;\Big|\; r_1, r_3 \in \{0, 1\},\, r_2 \in [0, 1] \Big\}$$

is an interval. Indeed, for fixed (r_1, r_3) the map in r_2 is affine with negative slope, so

$$\big[r_1 \cdot 0.3246524674 + r_3 \cdot 0.7059975221 - 0.4412484513,\;\; r_1 \cdot 0.3246524674 + r_3 \cdot 0.7059975221\big].$$

Taking the union over $r_1, r_3 \in \{0, 1\}$ yields

$$S(x) = \big[-0.4412484513,\;\; 1.0306499894\big],$$

with the minimum attained at $(r_1, r_2, r_3) = (0, 1, 0)$ and the maximum at $(1, 0, 1)$. This example instantiates the Relevance HyperVector as a set of gated per–basis contributions and computes explicit realizations and the exact score range.

Remark 3.38 (Choice of gate sets)**.** The simplest and most faithful RVM-style gates are binary: $R_i = \{0, 1\}$, encoding "excluded" vs. "active". One may also allow controlled shrinkage, e.g. $R_i = \{0\} \cup [\rho_i^-, \rho_i^+]$ with $0 \le \rho_i^- \le 1 \le \rho_i^+$, which still satisfies the multiplicative-closure and contains $\{0, 1\}$.

Theorem 3.39 (Relevance HyperVector forms a hypervector space)**.** *With $V = \mathbb{R}^N$, $+$ the usual addition, and $\circ$ from Definition 3.36(A), the quadruple*

$$\big(V,\ +,\ \circ,\ \mathbb{R}\big)$$

is a hypervector space *in the sense of the axioms* (H1)–(H5) *stated earlier: for all $a, b \in \mathbb{R}$ and $x, y \in V$,*

(H1) $a \circ (x + y) \ \subseteq\ a \circ x \ +\ a \circ y,$

(H2) $(a + b) \circ x \ \subseteq\ a \circ x \ +\ b \circ x,$

(H3) $a \circ \big(b \circ x\big) \ =\ (ab) \circ x,$

(H4) $a \circ (-x) \ =\ (-a) \circ x \ =\ -\big(a \circ x\big),$

(H5) $x \in 1 \circ x.$

Proof. Fix $a, b \in \mathbb{R}$, $x, y \in V$ and recall $a \circ z = \{D_r(az) : r \in \mathcal{R}\}$.

(H1) For any $r \in \mathcal{R}$,

$$D_r\big(a(x + y)\big) \ =\ D_r(ax) + D_r(ay),$$

and $D_r(ax) \in a \circ x$, $D_r(ay) \in a \circ y$. Thus $D_r(a(x + y)) \in a \circ x + a \circ y$. Taking the union over r yields $a \circ (x + y) \subseteq a \circ x + a \circ y$.

(H2) Similarly,

$$D_r\big((a + b)x\big) \ =\ D_r(ax) + D_r(bx) \in a \circ x + b \circ x,$$

and union over r gives $(a + b) \circ x \subseteq a \circ x + b \circ x$.

(H3) By definition,

$$b \circ x \ =\ \{\, D_s(bx) \mid s \in \mathcal{R} \,\}.$$

Applying $a\circ$ and using multiplicative closure of $\mathcal{R}$,

$$a\circ\big(b\circ x\big) \ =\ \bigcup_{s \in \mathcal{R}} \{\, D_r\big(a\, D_s(bx)\big) \mid r \in \mathcal{R} \,\} \ =\ \{\, D_r\, D_s\, (ab)\, x \mid r, s \in \mathcal{R} \,\} \ =\ \{\, D_t\, (ab)\, x \mid t \in \mathcal{R} \,\} \ =\ (ab)\circ x,$$

because $t := r \odot s \in \mathcal{R}$ (componentwise product) and every t arises this way.

(H4) Using linearity and $D_r(-z) = -D_r z$,

$$a \circ (-x) = \{D_r(a(-x))\} = \{-D_r(ax)\} = -(a \circ x), \qquad (-a) \circ x = \{D_r((-a)x)\} = \{-D_r(ax)\}.$$

(H5) Since $1 \in R_i$ for every i, $1_{\mathcal{R}} := (1, \ldots, 1) \in \mathcal{R}$ and $D_{1_{\mathcal{R}}} x = x$. Hence $x \in 1 \circ x$. □

Theorem 3.40 (Reduction to the classical Relevance Vector)**.** *Let $\mathcal{R}^\star \subseteq \{1, \ldots, N\}$ be a fixed relevance index set (e.g. the RVM solution), and define binary gates $r_i^\star := \mathbf{1}\{i \in \mathcal{R}^\star\}$. Suppose $R_i = \{0, 1\}$ for all i and set $r^\star = (r_1^\star, \ldots, r_N^\star) \in \mathcal{R}$. Then, for every input x,*

$$\widetilde{v}(x) \ =\ 1 \circ v(x) \ \ni\ D_{r^\star} v(x),$$

and the score set contains the classical RVM predictor:

$$\sum_{i=1}^{N} r_i^\star\, \mu_i\, \varphi_i(x) \ =\ \sum_{i \in \mathcal{R}^\star} \mu_i\, k(x, x_i).$$

Hence, picking the gate $r^\star$ collapses the Relevance HyperVector to the standard Relevance Vector *representation.*

Proof. With $R_i = \{0, 1\}$, $\mathcal{R} = \{0, 1\}^N$, so $r^\star \in \mathcal{R}$ and $D_{r^\star} v(x) = (r_1^\star \mu_1 \varphi_1(x), \ldots, r_N^\star \mu_N \varphi_N(x))^\top \in \widetilde{v}(x)$. Summing coordinates gives the stated predictor, which equals the RVM form because $r_i^\star = 1$ iff $i \in \mathcal{R}^\star$. □

3.3.3 Relevance SuperHyperVector

We introduce a definition of a *Relevance SuperHyperVector* that simultaneously generalizes the classical *Relevance Vector* and the set–valued *Relevance HyperVector*. We then prove it is an (m, n)-SuperHyperVector in the sense of a scalar m-ary superhyperoperation with codomain in the n-th iterated powerset.

Remark 3.41 (Setup)**.** Let $\{(x_j, t_j)\}_{j=1}^N$ be training data in $\mathbb{R}^d \times \mathbb{R}$ (or $\{-1, +1\}$ for classification). Fix a positive–definite kernel k and basis functions $\varphi_i(\cdot) = k(\cdot, x_i)$. Let

$$\mu = (\mu_1, \dots, \mu_N)^\top \in \mathbb{R}^N \quad \text{and} \quad \varphi(x) = \big(\varphi_1(x), \dots, \varphi_N(x)\big)^\top \in \mathbb{R}^N,$$

and define the (per–basis) contribution vector

$$v(x) \ := \ \mu \odot \varphi(x) \ \in \mathbb{R}^N,$$

with $\odot$ the Hadamard product. Put $V := \mathbb{R}^N$ with the usual addition $+$.

Definition 3.42 (Gate monoid and gate product)**.** For each index $i \in \{1, \dots, N\}$, let $R_i \subseteq \mathbb{R}_{\geq 0}$ be a nonempty set with

$$0 \in R_i, \quad 1 \in R_i, \quad r, s \in R_i \Rightarrow rs \in R_i.$$

Thus R_i is a multiplicative submonoid containing 0 and 1. Define the gate space $\mathcal{R} := R_1 \times \cdots \times R_N \subseteq \mathbb{R}_{\geq 0}^N$ with componentwise product $(r \odot s)_i = r_i s_i$. For $r \in \mathcal{R}$, write $D_r := \operatorname{diag}(r_1, \dots, r_N)$.

Definition 3.43 (Level-n nesting)**.** For $n \geq 1$ define the *nesting* map $\text{nest}_n : V \to \mathcal{P}_n(V)$ by

$$\text{nest}_1(u) := \{u\}, \qquad \text{nest}_{k+1}(u) := \{\text{nest}_k(u)\} \quad (k \geq 1).$$

Thus $\text{nest}_n(u)$ is the n-fold singleton tower over u.

Definition 3.44 (Relevance SuperHyperVector and (m, n)-operation)**.** Fix integers $m, n \geq 1$. Define the m-ary scalar superhyperoperation $\circ^{(m,n)} : \mathbb{R}^m \times V \to \mathcal{P}_n^*(V)$ by

$$\circ^{(m,n)}\big((a_1, \dots, a_m), z\big) \ := \ \Big\{ \text{nest}_n\Big(D_r\big(\big(\prod_{j=1}^m a_j\big) z\big)\Big) \ \Big| \ r \in \mathcal{R} \Big\}.$$

For any input x, the *Relevance SuperHyperVector* of x is the n-level realization set

$$\widetilde{v}^{(n)}(x) \ := \ \circ^{(m,n)}\big(\underbrace{(1, \dots, 1),}_{m \text{ ones}} \ v(x)\big) \ \in \ \mathcal{P}_n^*(V).$$

Each element of $\widetilde{v}^{(n)}(x)$ is an n-nested singleton tower whose leaf is a gated contribution vector $D_r v(x)$ for some $r \in \mathcal{R}$.

Example 3.45 (Concrete Relevance SuperHyperVector and (m, n)-operation)**.** Take $d = 1$, $N = 3$. Let the kernel be the linear kernel $k(x, y) = xy$ and choose training inputs

$$x_1 = 0, \quad x_2 = 1, \quad x_3 = 2.$$

Fix the weight vector

$$\mu = \begin{bmatrix} 0.8 \\ -1.2 \\ 0.5 \end{bmatrix} \in \mathbb{R}^3, \qquad \varphi(x) = \begin{bmatrix} \varphi_1(x) \\ \varphi_2(x) \\ \varphi_3(x) \end{bmatrix} = \begin{bmatrix} k(x, 0) \\ k(x, 1) \\ k(x, 2) \end{bmatrix} = \begin{bmatrix} 0 \\ x \\ 2x \end{bmatrix}.$$

Thus the contribution vector is

$$v(x) = \mu \odot \varphi(x) = \begin{bmatrix} 0.8 \cdot 0 \\ (-1.2) \cdot x \\ 0.5 \cdot 2x \end{bmatrix} = \begin{bmatrix} 0 \\ -1.2x \\ x \end{bmatrix}.$$

Evaluate at $x = 3$:

$$v(3) = \begin{bmatrix} 0 \\ -3.6 \\ 3 \end{bmatrix}.$$

Define gate sets (each a multiplicative submonoid containing 0 and 1):

$$R_1 = \{0, 1\}, \qquad R_2 = [0, 1], \qquad R_3 = \{0, 1\},$$

and the gate space $\mathcal{R} = R_1 \times R_2 \times R_3$. For $r = (r_1, r_2, r_3) \in \mathcal{R}$, write $D_r = \mathrm{diag}(r_1, r_2, r_3)$.

Fix $(m, n) = (2, 2)$. For scalars $(a_1, a_2) = (2, \frac{1}{2})$ (so $\prod_{j=1}^{2} a_j = 1$), the (m, n)-operation of Definition 3.44 gives

$$\circ^{(2,2)}\big((2, \tfrac{1}{2}),\ v(3)\big) = \Big\{ \mathrm{nest}_2\big(D_r(1 \cdot v(3))\big) \ \Big|\ r \in \mathcal{R} \Big\} = \Big\{ \{\ D_r v(3)\ \} \ \Big|\ r \in \mathcal{R} \Big\}.$$

Hence the Relevance SuperHyperVector at $x = 3$ is the 2-fold nested realization set

$$\widetilde{v}^{(2)}(3) = \Big\{ \{\ (r_1 \cdot 0,\ r_2 \cdot (-3.6),\ r_3 \cdot 3)\ \} \ \Big|\ r_1 \in \{0, 1\},\ r_2 \in [0, 1],\ r_3 \in \{0, 1\} \Big\}.$$

Three concrete realizations (elements of $\widetilde{v}^{(2)}(3)$) are:

$$\begin{aligned} r = (1, 1, 1) &: \ \{\ (0,\ -3.6,\ 3)\ \}, \\ r = (0, 0.30, 1) &: \ \{\ (0,\ -1.08,\ 3)\ \} \quad \text{since } -3.6 \times 0.30 = -1.08, \\ r = (1, 0, 0) &: \ \{\ (0,\ 0,\ 0)\ \}. \end{aligned}$$

If one defines the score set (sum of coordinates) at $x = 3$ by

$$S(3) := \Big\{ \langle \mathbf{1},\ D_r v(3) \rangle \ \Big|\ r \in \mathcal{R} \Big\} = \Big\{ -3.6\, r_2 + 3\, r_3 \ \Big|\ r_2 \in [0, 1],\ r_3 \in \{0, 1\} \Big\},$$

then explicitly

$$S(3) = \big[-3.6,\ 0\big] \ \cup\ \big[-0.6,\ 3\big] \ =\ \big[-3.6,\ 3\big].$$

This example exhibits an $(m, n) = (2, 2)$ SuperHyperVector: a binary scalar superhyperoperation ($m = 2$) with outputs living in the second iterated powerset ($n = 2$), and concretely shows how multiplicative gates induce shrinkage/selection of per-basis contributions.

Remark 3.46 (Simplicity of the model)**.** The only freedom is the choice of gate monoids $\{R_i\}$. Taking $R_i = \{0, 1\}$ recovers hard selection; allowing intervals (e.g. $R_i = \{0\} \cup [\rho_i^-, \rho_i^+]$) yields shrinkage while preserving multiplicative closure. No other complication is needed.

Remark 3.47 (Algebra on nested sets)**.** For nonempty $A, B \subseteq V$, define $A + B := \{u + v \mid u \in A,\ v \in B\}$ (Minkowski sum). Extend this to n-nested singletons by the identity

$$\mathrm{nest}_n(u)\ +\ \mathrm{nest}_n(v)\ =\ \{\ \mathrm{nest}_n(u + v)\ \} \qquad (\forall\, u, v \in V,\ \forall\, n \geq 1), \tag{1}$$

which follows directly from the singleton structure by a trivial induction on n.

Theorem 3.48 ((m,n)-SuperHyperVector structure)**.** *With $V = \mathbb{R}^N$, $+$ the usual addition, and $\circ^{(m,n)}$ as in Definition 3.44, the quadruple*

$$\big(V,\ +,\ \circ^{(m,n)},\ \mathbb{R}\big)$$

is an (m,n)-SuperHyperVector space: for all $x, y \in V$ and $\mathbf{a} = (a_1, \dots, a_m), \mathbf{b} = (b_1, \dots, b_m) \in \mathbb{R}^m$, writing $\Pi\mathbf{a} := \prod_{j=1}^m a_j$, the following hold:

$$\begin{aligned}
&\text{(SV1)} \quad \circ^{(m,n)}\,(\mathbf{a},\, x + y) \;\subseteq\; \circ^{(m,n)}(\mathbf{a},\, x) \;+\; \circ^{(m,n)}(\mathbf{a},\, y),\\
&\text{(SV2)} \quad \circ^{(m,n)}\,(\mathbf{a} + \mathbf{b},\, x) \;\subseteq\; \circ^{(m,n)}(\mathbf{a},\, x) \;+\; \circ^{(m,n)}(\mathbf{b},\, x),\\
&\text{(SV3)} \quad \circ^{(m,n)}\big(\mathbf{a},\, \circ^{(m,n)}(\mathbf{b},\, x)\big) \;=\; \circ^{(m,n)}(\mathbf{a}{\cdot}\mathbf{b},\, x),\\
&\text{(SV4)} \quad \circ^{(m,n)}\,(-\mathbf{a},\, x) \;=\; -\,\circ^{(m,n)}\,(\mathbf{a},\, x),\\
&\text{(SV5)} \quad x \;\in\; \circ^{(m,n)}\big((1, \dots, 1),\, x\big) \;\textit{(as a leaf of the nested singleton).}
\end{aligned}$$

Proof. Fix $x, y \in V$ and $\mathbf{a}, \mathbf{b} \in \mathbb{R}^m$. We repeatedly use the facts D_r is linear, $\mathcal{R}$ is closed under componentwise product, and (1).

(SV1) For any $r \in \mathcal{R}$,

$$\mathrm{nest}_n\Big(D_r\big((\Pi\mathbf{a})(x+y)\big)\Big) = \mathrm{nest}_n\Big(D_r\big((\Pi\mathbf{a})x\big) + D_r\big((\Pi\mathbf{a})y\big)\Big) \in \mathrm{nest}_n\Big(D_r\big((\Pi\mathbf{a})x\big)\Big) \;+\; \mathrm{nest}_n\Big(D_r\big((\Pi\mathbf{a})y\big)\Big),$$

where the last inclusion is exactly (1) at the leaf level. Taking the union over r yields (SV1).

(SV2) By linearity and $\Pi(\mathbf{a}+\mathbf{b})\,(x) = (\Pi\mathbf{a})\,x + (\Pi\mathbf{b})\,x$ at the scalar level, the same argument as in (SV1) gives the inclusion.

(SV3) Unfold definitions:

$$\circ^{(m,n)}(\mathbf{b}, x) = \Big\{\mathrm{nest}_n\big(D_s((\Pi\mathbf{b})x)\big) \;\Big|\; s \in \mathcal{R}\Big\}.$$

Applying $\circ^{(m,n)}(\mathbf{a}, \cdot)$ to each element replaces the vector leaf by $D_r\big((\Pi\mathbf{a})\,D_s((\Pi\mathbf{b})x)\big) = D_{r\odot s}\big((\Pi\mathbf{a})(\Pi\mathbf{b})\,x\big)$. Since $\mathcal{R}$ is multiplicatively closed and every $t \in \mathcal{R}$ can be written as $t = r \odot s$, we obtain

$$\circ^{(m,n)}\big(\mathbf{a},\ \circ^{(m,n)}(\mathbf{b}, x)\big) = \Big\{\mathrm{nest}_n\big(D_t((\Pi(\mathbf{a}\cdot\mathbf{b}))\,x)\big) \;\Big|\; t \in \mathcal{R}\Big\} = \circ^{(m,n)}(\mathbf{a}{\cdot}\mathbf{b}, x).$$

(SV4) For any $r \in \mathcal{R}$,

$$\mathrm{nest}_n\Big(D_r\big((\Pi(-\mathbf{a}))\,x\big)\Big) = \mathrm{nest}_n\Big(-\,D_r\big((\Pi\mathbf{a})\,x\big)\Big) = -\,\mathrm{nest}_n\Big(D_r\big((\Pi\mathbf{a})\,x\big)\Big),$$

hence the two sets coincide.

(SV5) Since $1 \in R_i$ for all i, $r^{\mathsf{id}} := (1, \dots, 1) \in \mathcal{R}$ and

$$\mathrm{nest}_n\big(D_{r^{\mathsf{id}}}(1 \cdot x)\big) = \mathrm{nest}_n(x) \in \circ^{(m,n)}\big((1, \dots, 1), x\big),$$

so x appears at the leaf. This proves all axioms. □

Theorem 3.49 (Generalizes Relevance Vector and Relevance HyperVector)**.** *(a)* ***Reduction to Relevance HyperVector.*** *For $n = 1$, Definition 3.44 yields*

$$\circ^{(m,1)}(\mathbf{a}, z) = \big\{\, D_r\big((\Pi\mathbf{a})\,z\big) \;\big|\; r \in \mathcal{R}\big\} \subseteq V,$$

so $\widetilde{v}^{(1)}(x) = \{D_r v(x) \mid r \in \mathcal{R}\}$ is exactly the set–valued Relevance HyperVector*.*

(b) ***Reduction to the classical Relevance Vector.*** *Let $R_i = \{0,1\}$ for all i and fix a gate $r^\star \in \{0,1\}^N$ corresponding to a relevance index set $\mathcal{R}^\star = \{i : r_i^\star = 1\}$ (e.g. the RVM solution). Then, for any $m \geq 1$ and $n \geq 1$,*

$$\mathrm{nest}_n\big(D_{r^\star} v(x)\big) \;\in\; \widetilde{v}^{(n)}(x),$$

and summing the coordinates recovers the RVM predictor $\sum_{i\in\mathcal{R}^\star} \mu_i\, \varphi_i(x) = \sum_{i\in\mathcal{R}^\star} \mu_i\, k(x, x_i)$.

Proof. (a) With $n = 1$, $\mathrm{nest}_1(u) = \{u\}$ and the definition collapses to a level-1 set of gated vectors, the Relevance HyperVector. (b) With binary gates, choosing $r^\star$ selects exactly the active bases; the coordinate sum is the classical relevance–vector form. In both cases, the inclusion is immediate from Definition 3.44. □

4 Conclusion

We defined SuperHyperVector Spaces by introducing a SuperHyperOperation on the iterated powerset of the underlying group and briefly examined their fundamental properties and hierarchical modeling potential. In the future, we expect further studies on various subfields employing frameworks such as Fuzzy Sets [96, 97], Intuitionistic Fuzzy Sets [98, 99], Neutrosophic Sets [100–104], Bipolar Fuzzy Sets [105, 106], HyperFuzzy Sets [107–110], SuperHyperFuzzy Sets [111], Picture Fuzzy Sets [112–114], Spherical Fuzzy Sets [115, 116], Hesitant Fuzzy Sets [117–119], Vague Sets [120, 121], and Plithogenic Sets [122–124]. We also anticipate the advancement of research on Vector, HyperVector, and SuperHyperVector structures across a wide range of other domains.

Acknowledgments

We wish to thank everyone whose guidance, ideas, and assistance contributed to this research. Our appreciation goes to the readers for their interest and to the scholars whose publications provided the groundwork for this study. We are also indebted to the individuals and institutions that supplied the resources and infrastructure necessary for completing and disseminating this paper. Lastly, we extend our gratitude to all who offered their support in various capacities.

Data Availability

This paper is purely theoretical and does not involve any empirical data. We welcome future empirical studies that build upon and test the concepts presented here.

Ethical Approval

As this work is entirely conceptual and involves no human or animal subjects, ethical approval was not required.

Conflicts of Interest

The authors declare no conflicts of interest in connection with this study or its publication.

Funding

No external or organizational funding supported this work.

Research Integrity

The authors affirm that, to the best of their knowledge, this manuscript represents their original research. It has not been previously published in any journal, nor is it currently being considered for publication elsewhere.

Disclaimer on Computational Tools

No computer-based tools—such as symbolic computation systems, automated theorem provers, or proof assistants (e.g., Mathematica, SageMath, Coq)—were employed in the development, analysis, or verification of the results contained in this paper. All derivations and proofs were conducted manually through analytical methods by the authors.

Use of Generative AI and AI-Assisted Tools

I use generative AI and AI-assisted tools for tasks such as English grammar checking, and I do not employ them in any way that violates ethical standards.

Disclaimer on Scope and Accuracy

The theoretical models and concepts proposed in this manuscript have not yet undergone empirical testing or practical deployment. Future work may investigate their utility in applied or experimental contexts. While the authors have taken care to maintain accuracy and provide appropriate citations, inadvertent errors or omissions may remain. Readers are encouraged to consult original references for confirmation and further study.

The authors assert that all mathematical results and justifications included in this work have been carefully reviewed and are believed to be correct. Should any inaccuracies or ambiguities be discovered, the authors welcome constructive feedback and will provide clarification upon request.

The conclusions presented are valid only within the specific theoretical framework and assumptions described in the text. Generalizing these results to other mathematical contexts may require further investigation. All opinions and interpretations expressed herein are solely those of the authors and do not necessarily reflect the views of their respective institutions.

References

[1] Sunday Adesina Adebisi and Adetunji Patience Ajuebishi. The order involving the neutrosophic hyperstructures, the construction and setting up of a typical neutrosophic group. *HyperSoft Set Methods in Engineering*, 3:26–31, 2025.

[2] Piergiulio Corsini and Violeta Leoreanu. *Applications of hyperstructure theory*, volume 5. Springer Science & Business Media, 2013.

[3] Bijan Davvaz, A Dehghan Nezhad, and SM Moosavi Nejad. Algebraic hyperstructure of observable elementary particles including the higgs boson. *Proceedings of the National Academy of Sciences, India Section A: Physical Sciences*, 90(1):169–176, 2020.

[4] M Al Tahan and Bijan Davvaz. Weak chemical hyperstructures associated to electrochemical cells. *Iranian Journal of Mathematical Chemistry*, 9(1):65–75, 2018.

[5] Takaaki Fujita. Chemical hyperstructures, superhyperstructures, and shv-structures: Toward a generalized framework for hierarchical chemical modeling. *ChemRxiv*, 2025.

[6] Takaaki Fujita. Toward a unified framework for knot theory, hyperknot theory, and superhyperknot theory via superhyperstructures. *Neutrosophic Knowledge*, 6:55–71, 2025.

[7] Adel Al-Odhari. A brief comparative study on hyperstructure, super hyperstructure, and n-super superhyperstructure. *Neutrosophic Knowledge*, 6:38–49, 2025.

[8] Florentin Smarandache. Foundation of superhyperstructure & neutrosophic superhyperstructure. *Neutrosophic Sets and Systems*, 63(1):21, 2024.

[9] Takaaki Fujita. An introduction and reexamination of hyperprobability and superhyperprobability: Comprehensive overview. *Asian Journal of Probability and Statistics*, 27(5):82–109, 2025.

[10] Takaaki Fujita. Rethinking strategic perception: Foundations and advancements in hypergame theory and superhypergame theory. *Prospects for Applied Mathematics and Data Analysis*, 4(2):01–14, 2024.

[11] Melody Mae Cabigting Lunar and Renson Aguilar Robles. Characterization and structure of a power set graph. *International Journal of Advanced Research and Publications*, 3(6):1–4, 2019.

[12] MA Shalu and S Devi Yamini. Counting maximal independent sets in power set graphs. *Indian Institute of Information Technology Design & Manufacturing (IIITD&M) Kancheepuram, India*, 2014.

[13] Takaaki Fujita. Extensions of multidirected graphs: Fuzzy, neutrosophic, plithogenic, rough, soft, hypergraph, and superhypergraph variants. *International Journal of Topology*, 2(3):11, 2025.

[14] Maissam Jdid, Florentin Smarandache, and Takaaki Fujita. A linear mathematical model of the vocational training problem in a company using neutrosophic logic, hyperfunctions, and superhyperfunction. *Neutrosophic Sets and Systems*, 87:1–11, 2025.

[15] Takaaki Fujita. Superhypermagma, lie superhypergroup, quotient superhypergroups, and reduced superhypergroups. *International Journal of Topology*, 2(3):10, 2025.

[16] Ajoy Kanti Das, Rajat Das, Suman Das, Bijoy Krishna Debnath, Carlos Granados, Bimal Shil, and Rakhal Das. A comprehensive study of neutrosophic superhyper bci-semigroups and their algebraic significance. *Transactions on Fuzzy Sets and Systems*, 8(2):80, 2025.

[17] Florentin Smarandache. Superhyperstructure & neutrosophic superhyperstructure, 2024. Accessed: 2024-12-01.

[18] F. Smarandache. Introduction to superhyperalgebra and neutrosophic superhyperalgebra. *Journal of Algebraic Hyperstructures and Logical Algebras*, 2022.

[19] Souzana Vougiouklí. Helix hyperoperation in teaching research. *Science & Philosophy*, 8(2):157–163, 2020.

[20] Souzana Vougiouklí. Hyperoperations defined on sets of s -helix matrices. 2020.

[21] Akbar Rezaei, Florentin Smarandache, and S. Mirvakili. Applications of (neutro/anti)sophications to semihypergroups. *Journal of Mathematics*, 2021.

[22] Souzana Vougiouklí. Helix-hyperoperations on lie-santilli admissibility. *Algebras Groups and Geometries*, 2023.

[23] Fakhry Asad Agusfrianto, Madeleine Al Tahan, and Yudi Mahatma. An introduction to neutrohyperstructures on some chemical reactions. In *NeutroGeometry, NeutroAlgebra, and SuperHyperAlgebra in Today's World*, pages 81–96. IGI Global, 2023.

[24] P Nikolaidou. Hyperstructures on bar of v & v in pieces. *Journal of Algebraic Hyperstructures and Logical Algebras*, 1(3):73–79, 2020.

[25] Takaaki Fujita. A theoretical exploration of hyperconcepts: Hyperfunctions, hyperrandomness, hyperdecision-making, and beyond (including a survey of hyperstructures). *Advancing Uncertain Combinatorics through Graphization, Hyperization, and Uncertainization: Fuzzy, Neutrosophic, Soft, Rough, and Beyond*, 344(498):111, 2025.

[26] Takaaki Fujita. Expanding horizons of plithogenic superhyperstructures: Applications in decision-making, control, and neuro systems. Technical report, 2025.

[27] Takaaki Fujita. Antihyperstructure, neutrohyperstructure, and superhyperstructure. *Advancing Uncertain Combinatorics through Graphization, Hyperization, and Uncertainization: Fuzzy, Neutrosophic, Soft, Rough, and Beyond*, page 311, 2025.

[28] Takaaki Fujita and Florentin Smarandache. Neutrosophic soft n-super-hypergraphs with real-world applications. *European Journal of Pure and Applied Mathematics*, 18(3):6621, 2025.

[29] Takaaki Fujita. Modeling hierarchical systems in graph signal processing, electric circuits, and bond graphs via hypergraphs and superhypergraphs. *Journal of Engineering Research and Reports*, 27(5):542, 2025.

[30] Takaaki Fujita, Atiqe Ur Rahman, Arkan A. Ghaib, Talal Ali Al-Hawary, and Arif Mehmood. On the properties and illustrative examples of soft superhypergraphs and rough superhypergraphs. *Prospects for Applied Mathematics and Data Analysis*, 5(1):12–31, 2025.

[31] Florentin Smarandache. *Extension of HyperGraph to n-SuperHyperGraph and to Plithogenic n-SuperHyperGraph, and Extension of HyperAlgebra to n-ary (Classical-/Neutro-/Anti-) HyperAlgebra*. Infinite Study, 2020.

[32] Takaaki Fujita and Florentin Smarandache. A concise study of some superhypergraph classes. *Neutrosophic Sets and Systems*, 77:548–593, 2024.

[33] Abdullah Kargın and Memet Şahin. Superhyper groups and neutro–superhyper groups. *2023 Neutrosophic SuperHyperAlgebra And New Types of Topologies*, 25, 2023.

[34] Florentin Smarandache. *Introduction to SuperHyperAlgebra and Neutrosophic SuperHyperAlgebra*. Infinite Study, 2022.

[35] Takaaki Fujita and Florentin Smarandache. *Neutrosophic TwoFold SuperhyperAlgebra and Anti SuperhyperAlgebra*. Infinite Study, 2025.

[36] Takaaki Fujita. Hierarchical uncertainty modeling via (m, n)-superhyperuncertain and (h, k)-ary (m, n)-superhyperuncertain sets: Unified extensions of fuzzy, neutrosophic, and plithogenic set theories.

[37] Takaaki Fujita. Exploration of new hyperstructures and superhyperstructures: Queue, markov chains, intervals, logic, and systems.

[38] Ahmed Hatip, Necati Olgun, et al. On the concepts of two-fold fuzzy vector spaces and algebraic modules. *Journal of Neutrosophic and Fuzzy Systems*, 7(2):46–52, 2023.

[39] Memet Şahin, Abdullah Kargın, and İsmet Yıldız. Neutrosophic triplet field and neutrosophic triplet vector space based on set valued neutrosophic quadruple number. *Quadruple Neutrosophic Theory And Applications*, 1:52, 2020.

[40] Maria Infusino. Topological vector spaces. 2018.

[41] Achilles Dramalidis and Thomas Vougiouklis. Fuzzy hv-substructures in a two dimensional euclidean vector space. *Iranian Journal of Fuzzy Systems*, 6(4):1–9, 2009.

[42] K Sameena et al. Fuzzy matroids from fuzzy vector spaces. *South East Asian Journal of Mathematics and Mathematical Sciences*, 17(03):381–390, 2021.

[43] P Lubczonok. Fuzzy vector spaces. *Fuzzy sets and systems*, 38(3):329–343, 1990.

[44] Davender S Malik and John N Mordeson. Fuzzy vector spaces. *Information sciences*, 55(1-3):271–281, 1991.

[45] Rajesh Kumar. Fuzzy vector spaces and fuzzy cosets. *Fuzzy sets and systems*, 45(1):109–116, 1992.

[46] AAA Agboola and SA Akinleye. Neutrosophic vector spaces. *Neutrosophic Sets and Systems*, 4:9–18, 2014.

[47] Muritala A Ibrahim, BS Badmus, SA Akinleye, et al. On refined neutrosophic vector spaces i. *International Journal of Neutrosophic Science*, 7:97–109, 2020.

[48] Mohammad Abobala. *Ah-subspaces in neutrosophic vector spaces*. Infinite Study, 2020.

[49] Mohammad Abobala. A study of ah-substructures in n-refined neutrosophic vector spaces. *International Journal of Neutrosophic Science*, 9(2):74–85, 2020.

[50] Maedeh Motameni, Reza Ameri, and R. Sadeghi. Hypermatrix based on krasner hypervector spaces. 2013.

[51] OR Dehghan. An introduction to neutrohypervector spaces. *Neutrosophic Sets and Systems*, 58(1):21, 2023.

[52] Hashida Haidros Rahima Manzil and S Manohar Naik. Android malware category detection using a novel feature vector-based machine learning model. *Cybersecurity*, 6(1):6, 2023.

[53] Gaston Baudat and Fatiha Anouar. Feature vector selection and projection using kernels. *Neurocomputing*, 55(1-2):21–38, 2003.

[54] Jeff Heaton. An empirical analysis of feature engineering for predictive modeling. In *SoutheastCon 2016*, pages 1–6. IEEE, 2016.

[55] Jiyeon Kim and Youngok Kang. Automatic classification of photos by tourist attractions using deep learning model and image feature vector clustering. *ISPRS International Journal of Geo-Information*, 11(4):245, 2022.

[56] Shinyoung Lim, Kwanyong Lee, Okhwan Byeon, and Taiyun Kim. Efficient iris recognition through improvement of feature vector and classifier. *ETRI journal*, 23(2):61–70, 2001.

[57] Bijaya Adhikari, Yao Zhang, Naren Ramakrishnan, and B Aditya Prakash. Sub2vec: Feature learning for subgraphs. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 170–182. Springer, 2018.

[58] Natalia Larios, Hongli Deng, Wei Zhang, Matt Sarpola, Jenny Yuen, Robert Paasch, Andrew Moldenke, David A Lytle, Salvador Ruiz Correa, Eric N Mortensen, et al. Automated insect identification through concatenated histograms of local appearance features: feature vector generation and region detection for deformable objects. *Machine Vision and Applications*, 19(2):105–123, 2008.

[59] W lodzis law Duch and Geerd HF Diercksen. Feature space mapping as a universal adaptive system. *Computer Physics Communications*, 87(3):341–371, 1995.

[60] Bernhard Scholkopf, Sebastian Mika, Chris JC Burges, Philipp Knirsch, K-R Muller, Gunnar Ratsch, and Alexander J Smola. Input space versus feature space in kernel-based methods. *IEEE transactions on neural networks*, 10(5):1000–1017, 1999.

[61] Maria Schuld and Nathan Killoran. Quantum machine learning in feature hilbert spaces. *Physical review letters*, 122(4):040504, 2019.

[62] Ingo Steinwart and Andreas Christmann. *Support vector machines*. Springer Science & Business Media, 2008.

[63] Marti A. Hearst, Susan T Dumais, Edgar Osuna, John Platt, and Bernhard Scholkopf. Support vector machines. *IEEE Intelligent Systems and their applications*, 13(4):18–28, 1998.

[64] William S Noble. What is a support vector machine? *Nature biotechnology*, 24(12):1565–1567, 2006.

[65] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[66] Christopher JC Burges. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2):121–167, 1998.

[67] Bernhard Schölkopf, Alex J Smola, Robert C Williamson, and Peter L Bartlett. New support vector algorithms. *Neural computation*, 12(5):1207–1245, 2000.

[68] Chun-Fu Lin and Sheng-De Wang. Fuzzy support vector machines. *IEEE transactions on neural networks*, 13(2):464–471, 2002.

[69] Abdul Quadir, Mudasir Ahmad Ganaie, and Mohammad Tanveer. Intuitionistic fuzzy generalized eigenvalue proximal support vector machine. *ArXiv*, abs/2408.01713, 2024.

[70] Ram Nayan Verma, Rahul Deo, Rakesh Srivastava, Naidu Subbarao, and Gajendra Pratap Singh. A new fuzzy support vector machine with pinball loss. *Discover Artificial Intelligence*, 3, 2023.

[71] Mehrbakhsh Nilashi, Hossein Ahmadi, Azizah Bt Abdul Manaf, Tarik A. Rashid, Sarminah Samad, Leila Shahmoradi, Nahla Aljojo, and Elnaz Akbari. Coronary heart disease diagnosis through self-organizing map and fuzzy support vector machine with incremental updates. *International Journal of Fuzzy Systems*, 22:1376 – 1388, 2020.

[72] Zhizhen Liang and Shifei Ding. Fuzzy twin support vector machines with distribution inputs. *IEEE Transactions on Fuzzy Systems*, 32:240–254, 2024.

[73] Shuihua Wang, Yang Li, Ying Shao, Carlo Cattani, Yudong Zhang, and Sidan Du. Detection of dendritic spines using wavelet packet entropy and fuzzy support vector machine. *CNS & neurological disorders drug targets*, pages 116–121, 2017.

[74] Mariza Agustia, Teuku Rizky Noviandy, Aga Maulana, Rivansyah Suhendra, Muslem Muslem, Novi Reandy Sasmita, Ghazi Mauer Idroes, Souvia Rahimah, Razief Perucha Fauzie Afidh, Muhammad Subianto, Irvanizam Irvanizam, and Rinaldi Idroes. Application of fuzzy support vector regression to predict the kovats retention indices of flavors and fragrances. *2022 International Conference on Electrical Engineering and Informatics (ICELTICs)*, pages 13–18, 2022.

[75] Muhammed Turhan, Dönüs Sengür, Songül Karabatak, Yanhui Guo, and Florentin Smarandache. Neutrosophic weighted support vector machines for the determination of school administrators who attended an action learning course based on their conflict-handling styles. *Symmetry*, 10:176, 2018.

[76] Mohammed Alshikho, Maissam Jdid, and Said Broumi. Artificial intelligence and neutrosophic machine learning in the diagnosis and detection of covid 19. *Prospects for Applied Mathematics and Data Analysis*, 2023.

[77] Wilbur H Highleyman. Linear decision functions, with application to pattern recognition. *Proceedings of the IRE*, 50(6):1501–1514, 2007.

[78] Alexander N Fuchs, Daniel Axehill, and Manfred Morari. On the choice of the linear decision functions for point location in polytopic data sets-application to explicit mpc. In *49th IEEE Conference on Decision and Control (CDC)*, pages 5283–5288. IEEE, 2010.

[79] Chen Pan, Xiangguo Yan, and Chongxun Zheng. Hard margin svm for biomedical image segmentation. In *International Symposium on Neural Networks*, pages 754–759. Springer, 2005.

[80] Houssem Sifaou, Abla Kammoun, and Mohamed-Slim Alouini. Phase transition in the hard-margin support vector machines. In *2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages 415–419. IEEE, 2019.

[81] Ítalo Santana, Breno Serrano, Maximilian Schiffer, and Thibaut Vidal. Support vector machines with the hard-margin loss: optimal training via combinatorial benders' cuts. *Journal of Global Optimization*, 92(1):205–225, 2025.

[82] Guoqiang Deng, Min Tang, Yuxing Xi, and Mingwu Zhang. Privacy-preserving online medical prediagnosis training model based on soft-margin svm. *IEEE Transactions on Services Computing*, 16(3):2072–2084, 2022.

[83] Huajun Wang, Yuanhai Shao, Shenglong Zhou, Ce Zhang, and Naihua Xiu. Support vector machine classifier via $l\_\{0/1\}$ l 0/1 soft-margin loss. *IEEE transactions on pattern analysis and machine intelligence*, 44(10):7253–7265, 2021.

[84] Vinod J Kadam, Samir S Yadav, and Shivajirao M Jadhav. Soft-margin svm incorporating feature selection using improved elitist ga for arrhythmia classification. In *International Conference on Intelligent Systems Design and Applications*, pages 965–976. Springer, 2018.

[85] Michael Tipping. The relevance vector machine. *Advances in neural information processing systems*, 12, 1999.

[86] Christopher M Bishop and Michael Tipping. Variational relevance vector machines. *arXiv preprint arXiv:1301.3838*, 2013.

[87] Tristan Fletcher. Relevance vector machines explained. *University College London: London, UK*, 2010.

[88] Michael E Tipping. Sparse bayesian learning and the relevance vector machine. *Journal of machine learning research*, 1(Jun):211–244, 2001.

[89] Arasanathan Thayananthan, Ramanan Navaratnam, Björn Stenger, Philip HS Torr, and Roberto Cipolla. Multivariate relevance vector machines for tracking. In *European conference on computer vision*, pages 124–138. Springer, 2006.

[90] V. Jayasudha, N. Deepa, and Devi. T. Fuzzy relevance vector machine-based segmentation and adaptive network-based fuzzy inference systems for parkinson's disease detection and classification. *Discov. Comput.*, 28:58, 2025.

[91] Thanikachalam Sathiya and Balasubramaniam Sathiyabhama. Fuzzy relevance vector machine based classification of lung nodules in computed tomography images. *International Journal of Imaging Systems and Technology*, 29:360 – 373, 2019.

[92] Hong-Bo Xie, Hu Huang, Jianhua Wu, and Lei Liu. A comparative study of surface emg classification by fuzzy relevance vector machine and fuzzy support vector machine. *Physiological Measurement*, 36:191 – 206, 2015.

[93] Carl Edward Rasmussen and Joaquin Quinonero-Candela. Healing the relevance vector machine through augmentation. In *Proceedings of the 22nd international conference on Machine learning*, pages 689–696, 2005.

[94] Ioannis Psorakis, Theodoros Damoulas, and Mark A Girolami. Multiclass relevance vector machines: sparsity and accuracy. *IEEE Transactions on neural networks*, 21(10):1588–1598, 2010.

[95] Fengxia Wang, Huixia Jin, and Xiao Chang. Relevance vector ranking for information retrieval. *J. Convergence Inf. Technol.*, 5(9):118–125, 2010.

[96] Lotfi A Zadeh. Fuzzy sets. *Information and control*, 8(3):338–353, 1965.

[97] Lotfi A Zadeh. Fuzzy logic, neural networks, and soft computing. In *Fuzzy sets, fuzzy logic, and fuzzy systems: selected papers by Lotfi A Zadeh*, pages 775–782. World Scientific, 1996.

[98] Krassimir T Atanassov. *On intuitionistic fuzzy sets theory*, volume 283. Springer, 2012.

[99] Krassimir T Atanassov. Circular intuitionistic fuzzy sets. *Journal of Intelligent & Fuzzy Systems*, 39(5):5981–5986, 2020.

[100] Said Broumi, Mohamed Talea, Assia Bakali, and Florentin Smarandache. Interval valued neutrosophic graphs. *Critical Review, XII*, 2016:5–33, 2016.

[101] Said Broumi, Assia Bakali, Mohamed Talea, and Florentin Smarandache. An isolated bipolar single-valued neutrosophic graphs. In *Information Systems Design and Intelligent Applications: Proceedings of Fourth International Conference INDIA 2017*, pages 816–822. Springer, 2018.

[102] Takaaki Fujita and Florentin Smarandache. Uncertain automata and uncertain graph grammar. *Neutrosophic Sets and Systems*, 74:128–191, 2024.

[103] Florentin Smarandache. *Neutrosophic Overset, Neutrosophic Underset, and Neutrosophic Offset. Similarly for Neutrosophic Over-/Under-/Off-Logic, Probability, and Statistics*. Infinite Study, 2016.

[104] Takaaki Fujita and Florentin Smarandache. Mixed graph in fuzzy, neutrosophic, and plithogenic graphs. *Neutrosophic Sets and Systems*, 74:457–479, 2024.

[105] Muhammad Akram. Bipolar fuzzy graphs. *Information sciences*, 181(24):5548–5564, 2011.

[106] Wen-Ran Zhang. Bipolar fuzzy sets and relations: a computational framework for cognitive modeling and multiagent decision analysis. *NAFIPS/IFIS/NASA '94. Proceedings of the First International Joint Conference of The North American Fuzzy Information Processing Society Biannual Conference. The Industrial Fuzzy Control and Intellige*, pages 305–309, 1994.

[107] Jayanta Ghosh and Tapas Kumar Samanta. Hyperfuzzy sets and hyperfuzzy group. *Int. J. Adv. Sci. Technol*, 41:27–37, 2012.

[108] Yong Lin Liu, Hee Sik Kim, and J. Neggers. Hyperfuzzy subsets and subgroupoids. *J. Intell. Fuzzy Syst.*, 33:1553–1562, 2017.

[109] Young Bae Jun, Seok-Zun Song, and Seon Jeong Kim. Length-fuzzy subalgebras in bck/bci-algebras. *Mathematics*, 6(1):11, 2018.

[110] Young Bae Jun, Seok-Zun Song, and Seon Jeong Kim. Distances between hyper structures and length fuzzy ideals of bck/bci-algebras based on hyper structures. *Journal of Intelligent & Fuzzy Systems*, 35(2):2257–2268, 2018.

[111] Takaaki Fujita. *Advancing Uncertain Combinatorics through Graphization, Hyperization, and Uncertainization: Fuzzy, Neutrosophic, Soft, Rough, and Beyond*. Biblio Publishing, 2025.

[112] Bui Cong Cuong and Vladik Kreinovich. Picture fuzzy sets-a new concept for computational intelligence problems. In *2013 third world congress on information and communication technologies (WICT 2013)*, pages 1–6. IEEE, 2013.

[113] Sankar Das, Ganesh Ghorai, and Madhumangal Pal. Picture fuzzy tolerance graphs with application. *Complex & Intelligent Systems*, 8(1):541–554, 2022.

[114] Waheed Ahmad Khan, Waqar Arif, Quoc Hung NGUYEN, Thanh Trung Le, and Hai Van Pham. Picture fuzzy directed hypergraphs with applications towards decision-making and managing hazardous chemicals. *IEEE Access*, 2024.

[115] Muhammad Akram, Danish Saleem, and Talal Al-Hawary. Spherical fuzzy graphs with application to decision-making. *Mathematical and Computational Applications*, 25(1):8, 2020.

[116] Fatma Kutlu Gündoğdu and Cengiz Kahraman. Spherical fuzzy sets and spherical fuzzy topsis method. *Journal of intelligent & fuzzy systems*, 36(1):337–352, 2019.

[117] Vicenç Torra and Yasuo Narukawa. On hesitant fuzzy sets and decision. In *2009 IEEE international conference on fuzzy systems*, pages 1378–1382. IEEE, 2009.

[118] Zeshui Xu. *Hesitant fuzzy sets theory*, volume 314. Springer, 2014.

[119] Vicenç Torra. Hesitant fuzzy sets. *International journal of intelligent systems*, 25(6):529–539, 2010.

[120] W-L Gau and Daniel J Buehrer. Vague sets. *IEEE transactions on systems, man, and cybernetics*, 23(2):610–614, 1993.

[121] Humberto Bustince and P Burillo. Vague sets are intuitionistic fuzzy sets. *Fuzzy sets and systems*, 79(3):403–405, 1996.

[122] Fazeelat Sultana, Muhammad Gulistan, Mumtaz Ali, Naveed Yaqoob, Muhammad Khan, Tabasam Rashid, and Tauseef Ahmed. A study of plithogenic graphs: applications in spreading coronavirus disease (covid-19) globally. *Journal of ambient intelligence and humanized computing*, 14(10):13139–13159, 2023.

[123] Shawkat Alkhazaleh. *Plithogenic soft set*. Infinite Study, 2020.

[124] Florentin Smarandache. *Plithogenic set, an extension of crisp, fuzzy, intuitionistic fuzzy, and neutrosophic sets-revisited*. Infinite study, 2018.