

Article

Decision-Sufficient Feature Sets, Decision Cores, and Decision Complexes: A Topological Approach to Classification

Sidney A. Morris ^{1,2} 

¹ School of Engineering, IT and Physical Sciences, Federation University Australia, P.O. Box 663, Ballarat, VI 3353, Australia; sidney.morris@latrobe.edu.au

² Department of Mathematical and Physical Sciences, La Trobe University, Melbourne, VI 3086, Australia

Abstract

Statistical classification is traditionally studied using probability, decision theory, information theory, and optimisation. We show that classification problems also possess a natural combinatorial and topological structure. Given a binary classification problem, we introduce the notion of a decision-sufficient feature set, namely a collection of features that preserves the optimal Bayes decision rule. Rather than focusing on a single optimal feature subset, we study the entire family of decision-sufficient feature sets. We prove that this family forms an order filter in the Boolean lattice of feature subsets and that its complement forms an abstract simplicial complex. Consequently, every classification problem canonically determines a topological object, called the decision complex, the faces of which correspond to feature subsets that fail to preserve optimal decisions. This viewpoint leads naturally to three invariants: the decision core, the decision dimension, and the decision complex. General structural results are established, and a complete characterisation is obtained for Gaussian-discriminant models. In that setting, the decision core coincides with the support of the discriminant vector, yielding explicit formulae for the decision core and decision dimension. These results establish a new connection between statistical decision theory and combinatorial topology.

Keywords: decision sufficiency; feature selection; Bayes classification; decision theory; simplicial complex; decision core; decision dimension; combinatorial topology.

MSC: 62C99, 62H30, 68T05, 05E45, 55U10

1. Introduction

Statistical classification is one of the central problems of modern statistics, machine learning, and artificial intelligence. Its mathematical foundations draw upon probability theory, statistical decision theory, information theory, optimisation, and statistical learning [1–4]. The objective is to construct decision rules that assign observations to one of several classes while minimising the probability or cost of incorrect decisions.

A recurring theme in classification theory is the search for smaller collections of features that preserve predictive performance. This problem appears in feature selection, sparse learning, graphical models, rough-set theory, explainable artificial intelligence, and interpretable machine learning. Reducing the number of variables can improve computational efficiency, reduce measurement costs, enhance interpretability, and sometimes improve generalisation performance [5–8].



Academic Editor: Emil Saucan

Received: 17 June 2026

Revised: 17 July 2026

Accepted: 17 July 2026

Published: 21 July 2026

Copyright: © 2026 by the author.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Most existing approaches formulate feature selection as an optimisation problem. Filter methods rank variables according to statistical criteria such as mutual information or correlation; wrapper methods repeatedly evaluate predictive performance; embedded methods incorporate feature selection within the learning algorithm itself; and rough-set methods seek minimal reducts that preserve classification capability. Although these approaches differ substantially, they all attempt to identify feature subsets that retain essential predictive information.

The present paper adopts a different viewpoint.

Rather than searching for a single optimal feature subset, we investigate the entire family of feature subsets that preserve the optimal Bayes decision rule. This shifts the emphasis from optimisation to mathematical structure. Instead of asking which subset should be selected, we ask how all decision-preserving subsets are organised.

Assume

$$X = (X_1, \dots, X_n)$$

is a collection of observed features and

$$D \in \{0, 1\}$$

is the class label. Under zero–one loss, the optimal Bayes classifier is completely determined by the posterior probability $P(D = 1 \mid X)$. We call a feature subset *decision-sufficient* if it preserves this optimal decision rule exactly, even though it need not preserve the complete posterior distribution.

The principal contribution of this paper is to show that the family of all decision-sufficient feature sets possesses a remarkably rich combinatorial structure. We prove that this family forms an order filter in the Boolean lattice of feature subsets. Its complement therefore forms an abstract simplicial complex, called the *decision complex*. Consequently every binary classification problem canonically determines a finite topological object whose faces correspond precisely to feature subsets that fail to preserve the optimal Bayes decision.

This viewpoint naturally leads to three structural invariants.

- (i) The *decision core*, consisting of features common to every decision-sufficient feature set;
- (ii) The *decision dimension*, measuring the minimum number of features required for exact decision preservation;
- (iii) The *decision complex*, recording the combinatorial topology of all non-decision-sufficient feature subsets.

Although these notions are new, they are closely related to several familiar concepts. Decision-sufficient feature sets are related to Markov blankets in probabilistic graphical models, but require only the preservation of the optimal decision rather than the full conditional distribution. Likewise, the decision core resembles a rough-set reduct, although it is defined from Bayes-optimal decision theory rather than indiscernibility relations. Thus, the proposed framework provides a unified mathematical setting in which ideas from statistical decision theory, graphical models, feature selection, rough-set theory, and combinatorial topology can be studied together.

The theory developed here is primarily structural rather than algorithmic. Nevertheless, the resulting invariants have potential applications to sparse classification, explainable artificial intelligence, interpretable machine learning, and computational feature selection. In particular, the decision complex provides a new topological representation of feature dependence that may prove useful in understanding the intrinsic complexity of classification problems.

The principal contributions of this paper are as follows.

- (i) We introduce decision-sufficient feature sets and establish their basic properties.
- (ii) We prove that the family of decision-sufficient feature sets forms an order filter in the Boolean lattice.
- (iii) We introduce the decision core and decision dimension as structural invariants associated with classification problems.
- (iv) We prove that the family of non-decision-sufficient feature sets forms an abstract simplicial complex.
- (v) We obtain complete characterisations of these invariants for Gaussian-discriminant models.
- (vi) We discuss relationships with Markov blankets, rough-set reducts, and modern feature-selection methodologies.

The remainder of the paper is organised as follows: Section 2 introduces decision-sufficient feature sets and establishes their basic properties. Section 3 examines their relationship with conditional independence and Markov blankets. Sections 4 and 5 develop the theories of decision cores and decision dimension. Section 6 presents explicit results for Gaussian-discriminant models. Section 7 introduces decision complexes and their associated topological structures. The final sections present examples, computational perspectives, future research directions, and concluding remarks.

Table 1. Principal notation used throughout the paper.

Symbol	Meaning
D	Binary class label
$X = (X_1, \dots, X_n)$	Feature vector
$S \subseteq [n]$	Feature subset
f^*	Bayes classifier
$\mathcal{D}$	Family of decision-sufficient feature sets
$\mathcal{N}$	Decision complex
C	Decision core
d_{DS}	Decision dimension
$\text{supp}(w)$	Support of the discriminant vector

2. Decision-Sufficient Feature Sets

In this section, we introduce the fundamental notion of a *decision-sufficient feature set* and establish its basic structural properties. The concept captures precisely those collections of features from which the optimal Bayes decision rule can be recovered exactly.

Throughout the paper, let

$$D \in \{0, 1\}$$

denote the binary outcome variable, and let

$$X = (X_1, \dots, X_n)$$

be the observed feature vector.

For a subset

$$S \subseteq [n] := \{1, \dots, n\},$$

we write

$$X_S = (X_i)_{i \in S}$$

for the corresponding subvector of observed features.

Let

$$\eta(X) = \text{Prob}(D = 1 \mid X)$$

denote the posterior probability of the positive class.

Under zero–one loss, the Bayes classifier is

$$f^*(X) = \mathbf{1}_{\{\eta(X) \geq \frac{1}{2}\}}.$$

More generally, if a fixed decision threshold

$$\tau \in (0, 1)$$

is prescribed, then

$$f_\tau^*(X) = \mathbf{1}_{\{\eta(X) \geq \tau\}}.$$

All structural results established below remain valid for an arbitrary fixed threshold, τ . Consequently, the order-theoretic and topological structure developed in this paper depends only upon the resulting decision rule and not upon the particular choice of threshold.

The central question is therefore:

Which collections of features contain sufficient information to recover the optimal Bayes decision?

This motivates the following definition.

Definition 1. A subset

$$S \subseteq [n]$$

is called decision-sufficient if the Bayes classifier $f^*(X)$ is measurable with respect to the σ -algebra $\sigma(X_S)$ generated by the retained features.

Equivalently,

$$f^*(X) \in \sigma(X_S).$$

Thus, a feature set is decision-sufficient precisely when the optimal decision can be recovered from the retained variables alone.

Unlike classical statistical sufficiency, no requirement is imposed on the entire posterior distribution. Only the optimal decision itself must be recoverable.

Theorem 1 (Decision Characterisation Theorem). *Let*

$$S \subseteq [n].$$

The following statements are equivalent.

1. *S is decision-sufficient.*
2. *There exists a measurable function*

$$g : \mathbb{R}^{|S|} \longrightarrow \{0, 1\}$$

such that

$$f^*(X) = g(X_S)$$

almost surely.

3. *There exists a measurable set*

$$A \subseteq \mathbb{R}^{|S|}$$

such that

$$f^*(X) = \mathbf{1}_{\{X_S \in A\}}$$

almost surely.

Proof. Assume first that S is decision-sufficient.

Since $f^*(X)$ is measurable with respect to $\sigma(X_S)$, the Doob–Dynkin lemma [9] implies the existence of a measurable function

$$g : \mathbb{R}^{|S|} \longrightarrow \{0, 1\}$$

such that

$$f^*(X) = g(X_S)$$

almost surely.

Thus, (1) implies (2).

If (2) holds, let

$$A = g^{-1}(\{1\}).$$

Since A is measurable,

$$g(X_S) = \mathbf{1}_{\{X_S \in A\}},$$

showing that (2) implies (3).

Finally, if

$$f^*(X) = \mathbf{1}_{\{X_S \in A\}},$$

then the right-hand side is measurable with respect to $\sigma(X_S)$.

Hence, $f^*(X)$ is measurable with respect to $\sigma(X_S)$, proving that (3) implies (1). $\square$

The preceding theorem shows that decision sufficiency is fundamentally a measurability property. The optimal Bayes decision factors through the reduced feature vector, X_S , regardless of how the posterior probability itself is represented.

Proposition 1 (Monotonicity). *Assume*

$$S \subseteq T \subseteq [n].$$

If S is decision-sufficient, then T is decision-sufficient.

Proof. Since

$$\sigma(X_S) \subseteq \sigma(X_T),$$

every function measurable with respect to $\sigma(X_S)$ is also measurable with respect to $\sigma(X_T)$.

Hence, if $f^*(X)$ is measurable with respect to $\sigma(X_S)$, it is also measurable with respect to $\sigma(X_T)$.

Therefore, T is decision-sufficient. $\square$

Definition 2. *The family of all decision-sufficient feature sets is*

$$\mathcal{D} = \{S \subseteq [n] : S \text{ is decision-sufficient}\}.$$

Corollary 1. *The family $\mathcal{D}$ is an order filter in the Boolean lattice $2^{[n]}$.*

Proof. Let

$$S \in \mathcal{D}$$

and assume

$$S \subseteq T.$$

By Proposition 1, $T \in \mathcal{D}$.

Hence, $\mathcal{D}$ is upward closed and therefore forms an order filter. $\square$

The order filter property is the starting point for the remainder of the paper. It provides the combinatorial foundation upon which the decision core, decision dimension, and decision complex are constructed.

3. Decision Sufficiency and Conditional Independence

The notion of decision sufficiency is closely related to several established concepts in statistics and machine learning, including conditional independence, Markov blankets, minimal sufficient feature subsets, and rough-set reducts. The purpose of this section is to clarify these relationships and to position the present framework within the existing literature.

3.1. Conditional Independence as a Sufficient Condition

Let

$$S \subseteq [n].$$

Assume that the outcome variable D is conditionally independent of the discarded features once the retained features are known. Thus,

$$D \perp X_{S^c} \mid X_S.$$

In this situation, the retained features contain all probabilistic information about the outcome variable.

The following theorem shows that such feature sets are necessarily decision-sufficient.

Theorem 2. Let

$$S \subseteq [n].$$

If

$$D \perp X_{S^c} \mid X_S,$$

then S is decision-sufficient.

Proof. Conditional independence implies

$$\text{Prob}(D = 1 \mid X) = \text{Prob}(D = 1 \mid X_S)$$

almost surely [10] (Chapter 3).

Hence,

$$\eta(X) = \eta_S(X_S),$$

where

$$\eta_S(X_S) = \text{Prob}(D = 1 \mid X_S).$$

Therefore,

$$f^*(X) = \mathbf{1}_{\{\eta(X) \geq 1/2\}} = \mathbf{1}_{\{\eta_S(X_S) \geq 1/2\}},$$

which depends only upon X_S .

Thus, $f^*(X)$ is measurable with respect to $\sigma(X_S)$, proving that S is decision-sufficient. $\square$

3.2. Decision Sufficiency Is Strictly Weaker

The converse of Theorem 2 is generally false.

Conditional independence preserves the complete posterior distribution, whereas decision sufficiency preserves only the optimal Bayes decision. Consequently, two different posterior distributions may induce exactly the same classifier while differing substantially as probability models.

Decision sufficiency should therefore be viewed as a decision-theoretic property rather than an inferential one. It identifies precisely the information required for optimal decisions and no more.

3.3. Relationship with Markov Blankets

Markov blankets play a fundamental role in probabilistic graphical models [7,8]. A Markov blanket is a set of variables that renders the target variable conditionally independent of all remaining variables.

Theorem 2 immediately implies that every Markov blanket is decision-sufficient.

The converse need not hold.

A decision-sufficient feature set may preserve the optimal Bayes decision without preserving the entire conditional distribution. Consequently, the family of decision-sufficient feature sets is generally larger than the family of Markov blankets.

3.4. Relationship with Rough-Set Reducts

The decision core also bears a close conceptual relationship to reducts in rough-set theory.

Both notions seek collections of features that preserve classification behaviour while eliminating unnecessary variables. However, the underlying mathematical frameworks are quite different.

Classical rough-set reducts are defined through indiscernibility relations and set approximations, whereas the decision core introduced here is defined through preservation of the Bayes-optimal decision rule.

Thus, the decision core may be viewed as a decision-theoretic analogue of a rough-set reduct rather than a reformulation of the rough-set concept itself.

3.5. Relationship with Feature Relevance

Many modern feature-selection methods distinguish between relevant, irrelevant, and redundant variables.

The present framework approaches this question from a different perspective.

Rather than assigning relevance scores to individual variables, we study the entire family of feature subsets that preserve optimal decisions. This global view naturally reveals the order-theoretic structure of decision-preserving feature sets and leads directly to the notions of decision cores, decision dimension, and decision complexes.

3.6. Hierarchy of Information Preservation

The preceding discussion yields the hierarchy

Conditional Independence $\implies$ Posterior Preservation $\implies$ Decision Sufficiency.

The first implication follows directly from the definition of conditional independence. The second follows from Theorem 2.

In general neither implication can be reversed.

Consequently, decision sufficiency occupies a distinct position among existing feature-reduction concepts. It preserves exactly the information required for optimal decisions while allowing substantially greater flexibility than approaches based upon complete probabilistic preservation.

Remark 1. *In many practical applications—including medical diagnosis, credit approval, fraud detection, industrial monitoring, and autonomous decision making—the optimal action is of primary importance.*

In such situations, preserving the complete posterior distribution may be unnecessary. Decision sufficiency identifies precisely the information required to reproduce the optimal decision rule and therefore provides a natural theoretical framework for interpretable feature selection.

The comparisons developed in this section position decision sufficiency within the broader literature on graphical models, feature selection, and rough-set theory. They also motivate the structural invariants introduced in the following sections, beginning with the decision core.

4. Decision Cores

The family of decision-sufficient feature sets introduced in the preceding sections forms an order filter in the Boolean lattice $2^{[n]}$. It is therefore natural to ask whether there exists a collection of features that is common to every decision-sufficient feature set.

This leads to the first structural invariant associated with a classification problem.

4.1. Definition

Recall that

$$\mathcal{D} = \{ S \subseteq [n] : S \text{ is decision-sufficient} \}.$$

Since the full feature set $[n]$ is always decision-sufficient, $\mathcal{D}$ is non-empty.

Definition 3. *The decision core is the set*

$$C = \bigcap_{S \in \mathcal{D}} S.$$

Thus, the decision core consists precisely of those features that occur in every decision-sufficient feature set.

It identifies the universally indispensable variables for preserving the optimal Bayes decision.

Proposition 2. *For every decision-sufficient feature set*

$$S \in \mathcal{D},$$

one has

$$C \subseteq S.$$

Proof. Since

$$C = \bigcap_{T \in \mathcal{D}} T,$$

the result follows immediately. $\square$

It is important to emphasise that the decision core need not itself be decision-sufficient.

Indeed, intersections of members of an order filter need not themselves belong to the filter.

Consequently, the decision core records universally necessary features, but may not by itself determine the optimal Bayes classifier.

4.2. When Is the Core Decision-Sufficient?

The next theorem characterises precisely those classification problems for which the decision core itself preserves the optimal decision rule. In this case, the entire family of decision-sufficient feature sets is generated by the core alone.

Theorem 3 (Core Criterion). *Let*

$$C = \bigcap_{S \in \mathcal{D}} S.$$

The following are equivalent.

1. *The decision core C is decision-sufficient.*
2. *The decision core is the unique minimal element of $\mathcal{D}$.*
3. $\mathcal{D} = \{ S \subseteq [n] : C \subseteq S \}.$

That is, $\mathcal{D}$ is the principal order filter generated by C .

Proof. Assume first that C is decision-sufficient.

Since Proposition 2 shows that C is contained in every member of $\mathcal{D}$, it follows that no smaller decision-sufficient subset exists. Hence, C is the unique minimal element of $\mathcal{D}$.

Conversely, assume C is the unique minimal element of $\mathcal{D}$.

Because $\mathcal{D}$ is an order filter, every superset of C belongs to $\mathcal{D}$.

Since every member of $\mathcal{D}$ contains C , we obtain

$$\mathcal{D} = \{ S \subseteq [n] : C \subseteq S \}.$$

Finally, if

$$\mathcal{D} = \{ S \subseteq [n] : C \subseteq S \},$$

then $C \subseteq C$, so $C \in \mathcal{D}$.

Hence, C is decision-sufficient. $\square$

The theorem shows that the decision core plays two distinct roles.

In every classification problem it records the universally indispensable features.

In the important special case where it is itself decision-sufficient, it becomes the unique minimal decision-preserving feature set and completely determines the order filter $\mathcal{D}$.

Following standard terminology in order theory, we shall refer to this situation by saying that $\mathcal{D}$ is a principal filter generated by the decision core.

Definition 4. *A classification problem is said to possess a principal decision core if the equivalent conditions of Theorem 3 hold.*

5. Decision Dimension

The decision core identifies those features that are indispensable for preserving the optimal Bayes decision. A complementary numerical invariant is obtained by asking a different question:

What is the smallest number of features required to recover the optimal Bayes decision?

The answer leads naturally to the notion of decision dimension.

5.1. Definition

Recall that

$$\mathcal{D} = \{ S \subseteq [n] : S \text{ is decision-sufficient} \}.$$

Since the full feature set $[n]$ is always decision-sufficient, the family $\mathcal{D}$ is non-empty.

Definition 5. The decision dimension of a classification problem is

$$d_{DS} = \min\{|S| : S \in \mathcal{D}\}.$$

Thus, the decision dimension measures the minimum number of observed features required to reproduce the optimal Bayes decision exactly.

Unlike the ambient dimension n , which merely counts the available features, the decision dimension measures the intrinsic complexity of the classification problem.

5.2. Basic Properties

The definition immediately yields the following elementary bounds.

Proposition 3. For every non-trivial classification problem,

$$1 \leq d_{DS} \leq n.$$

Proof. Since the classification problem is assumed to be non-trivial, the Bayes classifier is not constant.

Consequently, the empty feature set cannot be decision-sufficient, so

$$d_{DS} \geq 1.$$

Since

$$[n] \in \mathcal{D},$$

we also have

$$d_{DS} \leq n.$$

In the degenerate case where the Bayes classifier is constant, the empty feature set is decision-sufficient and therefore $d_{DS} = 0$. $\square$

Proposition 4. Assume a feature vector

$$Y = (X, Z)$$

is obtained by adjoining additional variables Z .

Then

$$d_{DS}(D, Y) \leq d_{DS}(D, X).$$

Proof. Every decision-sufficient subset of the original feature vector remains decision-sufficient after additional variables are introduced.

Hence, enlarging the ambient feature space cannot increase the minimum cardinality of a decision-sufficient subset. $\square$

5.3. Principal Decision Cores

When a principal decision core exists, the decision dimension admits an especially simple description.

Theorem 4. Assume the classification problem possesses a principal decision core C .
Then,

$$d_{DS} = |C|.$$

Proof. By Theorem 3,

$$\mathcal{D} = \{S \subseteq [n] : C \subseteq S\}.$$

Every decision-sufficient feature set therefore contains every element of C , and consequently has cardinality at least $|C|$.

Since the core itself is decision-sufficient,

$$C \in \mathcal{D},$$

it follows immediately that $d_{DS} = |C|$. $\square$

Theorem 4 shows that, whenever a principal decision core exists, the decision dimension is simply the number of indispensable features.

Thus, the decision core and the decision dimension provide complementary descriptions of the same decision structure: one set-theoretic and the other numerical.

5.4. Decision Bases

The decision dimension is closely related to minimal decision-sufficient feature sets.

Definition 6. A decision-sufficient feature set

$$B \in \mathcal{D}$$

is called a decision basis if no proper subset of B is decision-sufficient.

Decision bases are precisely the minimal elements of the order filter $\mathcal{D}$.

Proposition 5. Every decision basis has cardinality at least d_{DS} .

Moreover, there exists a decision basis having cardinality exactly d_{DS} .

Proof. The first statement follows directly from the definition of d_{DS} .

Choose a decision-sufficient feature set having minimum cardinality. No proper subset can be decision-sufficient, for otherwise the defining minimum would not be attained.

Hence, every smallest decision-sufficient feature set is a decision basis. $\square$

The decision dimension therefore admits two equivalent interpretations.

It is both the minimum cardinality of a decision-sufficient feature set and the cardinality of a smallest decision basis.

This numerical invariant complements the decision core by measuring the minimum amount of information required for exact decision preservation, independently of the total number of available features.

6. Gaussian-Discriminant Models

The abstract theory developed in the preceding sections applies to arbitrary binary classification problems. In this section, we consider an important class of models for which the decision-sufficient structure can be described completely.

Gaussian-discriminant models occupy a central position in statistical classification theory and include classical linear discriminant analysis as a special case [3,4,11]. They therefore provide a natural testing ground for the concepts introduced in this paper.

The principal result of this section is that, in the Gaussian setting, the decision core, decision dimension, and family of decision-sufficient feature sets admit explicit closed-form descriptions.

6.1. Model Description

Throughout this section, we assume that

$$D \in \{0, 1\},$$

and that

$$X | D = d \sim N(\mu_d, \Sigma), \quad d \in \{0, 1\},$$

where

$$\mu_0, \mu_1 \in \mathbb{R}^n$$

and

$$\Sigma$$

is a positive-definite covariance matrix common to both classes.

Let

$$\pi_d = \text{Prob}(D = d), \quad d \in \{0, 1\},$$

denote the prior probabilities.

It is well known that the log-likelihood ratio is

$$\Lambda(x) = \log \frac{f_1(x)}{f_0(x)} = w^\top x + c,$$

where

$$w = \Sigma^{-1}(\mu_1 - \mu_0)$$

and

$$c = -\frac{1}{2}(\mu_1 + \mu_0)^\top \Sigma^{-1}(\mu_1 - \mu_0) + \log \frac{\pi_1}{\pi_0}.$$

Consequently, the Bayes classifier is

$$f^*(x) = \mathbf{1}_{\{w^\top x + c \geq 0\}}.$$

Thus, optimal classification depends only upon the linear discriminant $w^\top x$.

6.2. Discriminant Support

The coordinates that genuinely influence the Bayes decision are precisely those corresponding to the non-zero coefficients of the discriminant vector.

Definition 7. The discriminant support is

$$\text{supp}(w) = \{i \in [n] : w_i \neq 0\}.$$

Theorem 5 (Gaussian Characterisation). A feature subset

$$S \subseteq [n]$$

is decision-sufficient if and only if

$$\text{supp}(w) \subseteq S.$$

Proof. Assume first that

$$\text{supp}(w) \subseteq S.$$

Then, every coefficient outside S is zero, so

$$w^\top x = \sum_{i \in S} w_i x_i.$$

Hence,

$$f^*(x) = \mathbf{1}_{\{\sum_{i \in S} w_i x_i + c \geq 0\}},$$

which depends only upon the retained coordinates.

Therefore, S is decision-sufficient.

Conversely, assume that

$$\text{supp}(w) \not\subseteq S.$$

Choose

$$j \in \text{supp}(w) \setminus S.$$

Then, $w_j \neq 0$.

Fix every coordinate except x_j .

Since the discriminant function depends linearly upon x_j , varying this coordinate changes the value of

$$w^\top x + c.$$

Hence, there exist observations having identical values of X_S but different Bayes decisions.

Therefore, the classifier cannot be determined solely from X_S , so S is not decision-sufficient. $\square$

6.3. The Decision Core

The Gaussian characterisation immediately identifies the decision core.

Theorem 6. *The Gaussian-discriminant model possesses a principal decision core*

$$C = \text{supp}(w).$$

Proof. By Theorem 5,

$$\mathcal{D} = \{S \subseteq [n] : \text{supp}(w) \subseteq S\}.$$

Hence, $\text{supp}(w)$ is the unique minimal decision-sufficient feature set.

The Core Criterion therefore implies $C = \text{supp}(w)$. $\square$

6.4. Decision Dimension

The decision dimension follows immediately.

Theorem 7. *For Gaussian-discriminant models,*

$$d_{DS} = |\text{supp}(w)|.$$

Proof. The smallest decision-sufficient feature set is $\text{supp}(w)$.

The result therefore follows directly from the definition of decision dimension. $\square$

6.5. Interpretation

The Gaussian theory provides an important consistency check for the general framework developed in this paper.

In this setting, the decision core coincides exactly with the support of the classical discriminant vector, while the decision dimension is simply the number of non-zero discriminant coefficients.

Consequently, the abstract invariants introduced earlier reduce to familiar statistical quantities in one of the most widely used classification models.

This illustrates that the proposed theory extends, rather than replaces, classical feature-selection ideas. In more general nonlinear or non-parametric classification problems, analogous explicit formulae may not exist, making the decision core and decision complex useful structural invariants for studying decision-preserving feature sets.

7. Decision Complexes and Topological Structure

The preceding sections established that the family

$$\mathcal{D} = \{ S \subseteq [n] : S \text{ is decision-sufficient} \}$$

forms an order filter in the Boolean lattice $2^{[n]}$.

The purpose of this section is to investigate the complementary family

$$\mathcal{N} = 2^{[n]} \setminus \mathcal{D},$$

consisting of all feature subsets that fail to preserve the optimal Bayes decision.

The principal observation of this paper is that $\mathcal{N}$ naturally carries the structure of an abstract simplicial complex. Consequently every binary classification problem determines not only a probabilistic decision structure but also a canonical combinatorial-topological object.

Unlike the simplicial complexes encountered in topological data analysis, which are typically constructed from geometric or metric information, the decision complex introduced here arises directly from the logical structure of optimal decision making. It is therefore an intrinsic invariant of the classification problem itself.

7.1. The Decision Complex

We begin with the fundamental structural result.

Theorem 8 (Decision Complex Theorem). *The family*

$$\mathcal{N} = 2^{[n]} \setminus \mathcal{D}$$

is an abstract simplicial complex.

Proof. Let

$$T \in \mathcal{N}.$$

Thus, T is not decision-sufficient.

Assume

$$S \subseteq T.$$

If S were decision-sufficient, then Proposition 1 would imply that T is also decision-sufficient, contradicting the choice of T .

Hence

$$S \in \mathcal{N}.$$

Therefore, every subset of every member of $\mathcal{N}$ again belongs to $\mathcal{N}$, which is precisely the definition of an abstract simplicial complex. $\square$

Definition 8. *The simplicial complex*

$$\mathcal{N} = 2^{[n]} \setminus \mathcal{D}$$

is called the decision complex associated with the classification problem.

The decision complex records precisely those feature subsets that fail to preserve the optimal Bayes decision. It therefore provides a global combinatorial description of feature insufficiency, complementing the order filter formed by the decision-sufficient subsets.

7.2. Facets

The maximal simplices of the decision complex have an immediate decision-theoretic interpretation.

Definition 9. *A facet of the decision complex is a maximal element of $\mathcal{N}$ under inclusion.*

Equivalently, a facet is a largest feature subset that still fails to preserve the optimal Bayes decision.

Proposition 6. *A subset*

$$F \subseteq [n]$$

is a facet of the decision complex if and only if

$$F \notin \mathcal{D}$$

and

$$F \cup \{i\} \in \mathcal{D}$$

for every

$$i \notin F.$$

Proof. This is simply the definition of maximality in the partially ordered set $(2^{[n]}, \subseteq)$. $\square$

7.3. Decision Bases and Facets

Recall that a decision basis is a minimal decision-sufficient feature set.

Proposition 7. *A subset*

$$B \subseteq [n]$$

is a decision basis if and only if

$$B \in \mathcal{D}$$

and

$$B \setminus \{i\} \in \mathcal{N}$$

for every

$$i \in B.$$

Proof. The result follows immediately from the definition of minimality. $\square$

Thus, decision bases and facets occupy dual positions relative to the order filter and the simplicial complex. Minimal decision-preserving feature sets correspond to maximal failures of decision preservation.

7.4. Gaussian Decision Complex

The Gaussian theory developed in the previous section allows the decision complex to be described explicitly.

Let

$$C = \text{supp}(w)$$

denote the Gaussian decision core.

Then

$$\mathcal{N} = \{S \subseteq [n] : C \not\subseteq S\}.$$

Hence, a feature subset belongs to the decision complex precisely when it omits at least one indispensable feature.

Theorem 9. *The facets of the Gaussian decision complex are precisely*

$$[n] \setminus \{i\}, \quad i \in C.$$

Proof. A subset belongs to the decision complex exactly when it fails to contain the entire decision core.

A maximal such subset is obtained by omitting exactly one element of the core and retaining every other feature. $\square$

Consequently the number of facets equals

$$|C| = d_{DS}.$$

7.5. Topological Invariants

Once the decision complex has been constructed, one may study its classical topological invariants.

These include

1. The facets;
2. The f -vector;
3. The Euler characteristic;
4. Homology groups;
5. Betti numbers;
6. Homotopy type.

These invariants provide additional information about the global organisation of feature subsets that fail to preserve optimal decisions.

7.6. Discussion

The decision complex is the principal topological object introduced in this paper.

Together, the order filter $\mathcal{D}$, the decision core, the decision dimension, and the decision complex provide complementary descriptions of the same classification problem.

Although the present work is primarily theoretical, the proposed framework also provides a natural setting in which ideas from statistical decision theory, Markov blankets, rough-set reducts, and modern feature-selection methods may be studied within a common mathematical framework. Recent developments in rough-set theory have demonstrated the continuing importance of reduct-based methods in areas including economic decision

making, medical diagnosis, and neighbourhood-based information systems [12–14]. The resulting order-theoretic and topological viewpoint complements these optimisation-based approaches by describing the complete organisation of all decision-preserving feature subsets rather than focusing solely on individual minimal reducts.

The next section illustrates the theory through several explicit examples.

8. Examples, Computation, and Applications

The purpose of this section is to illustrate the concepts introduced in the preceding sections and to demonstrate how the decision core, decision dimension, and decision complex may be computed in representative classification problems.

The examples are intentionally simple. Their purpose is not to introduce new mathematical results but to demonstrate how the proposed invariants arise naturally and how they may assist in understanding feature relevance.

8.1. Example 1: A Linear Classifier

Assume

$$f^*(X) = \mathbf{1}_{\{X_1 + X_2 > 0\}}.$$

Assume that neither variable alone determines the decision.

Then

$$C = \{1, 2\},$$

and

$$d_{DS} = 2.$$

The family of decision-sufficient feature sets is

$$\mathcal{D} = \{S \subseteq [n] : \{1, 2\} \subseteq S\},$$

while the decision complex consists of every subset omitting at least one of the two indispensable variables.

8.2. Example 2: A Single Relevant Feature

Assume the Bayes classifier depends only upon

$$X_3.$$

Then

$$C = \{3\}, \quad d_{DS} = 1.$$

Every decision-sufficient feature set contains the third feature, while all remaining variables are irrelevant for optimal decision making.

8.3. Example 3: Majority Voting

Consider three binary features

$$X_1, X_2, X_3,$$

and assume the Bayes classifier predicts the majority vote.

Each variable is indispensable.

Consequently,

$$C = \{1, 2, 3\}, \quad d_{DS} = 3.$$

The associated decision complex is the boundary of a 2-simplex and is therefore homotopy equivalent to the circle

$$S^1.$$

This example illustrates that familiar topological spaces may arise naturally from decision-preserving feature structures.

8.4. Computational Considerations

For a classification problem involving n observed features, there are

$$2^n$$

possible feature subsets.

A direct search for all decision-sufficient subsets therefore becomes computationally infeasible when n is large.

The theoretical framework developed in this paper should therefore be viewed primarily as a structural description rather than as an exhaustive computational procedure.

Nevertheless, several observations suggest that practical algorithms may be developed.

First, once a decision-sufficient subset has been identified, every superset is automatically decision-sufficient by monotonicity.

Second, minimal decision-sufficient subsets determine the entire order filter, making it unnecessary to examine every subset individually.

Third, in important statistical models, such as Gaussian discriminant analysis, the decision core and decision dimension admit explicit closed-form expressions.

These observations suggest that efficient algorithms may be possible for important classes of classifiers.

8.5. Relationship with Existing Feature Selection

Most modern feature-selection algorithms attempt to identify a single high-quality subset of variables.

The present framework instead studies the entire family of decision-preserving subsets.

Consequently, it complements rather than replaces existing approaches.

Filter methods, wrapper methods, embedded methods, sparse optimisation, and rough-set reducts may all be interpreted within the broader structural framework introduced here.

Rather than producing one selected subset, the theory reveals the complete organisation of all subsets preserving optimal decisions.

8.6. Potential Applications

Although the present paper is theoretical, the proposed invariants suggest several possible applications.

These include

1. Interpretable machine learning;
2. Explainable artificial intelligence;
3. Sparse statistical learning;
4. Computational feature selection;
5. Probabilistic graphical models;
6. Rough-set based decision analysis;
7. Medical diagnosis;
8. Financial risk assessment;
9. Industrial process monitoring.

In these settings, the decision core identifies indispensable variables, whereas the decision complex provides a global representation of feature dependence that may complement existing feature-selection methodologies.

8.7. Discussion

The examples presented above demonstrate that the proposed invariants are straightforward to compute in simple situations and admit natural interpretations.

More importantly, they illustrate that every classification problem possesses not only probabilistic structure but also combinatorial and topological structure.

This viewpoint opens several directions for future work, including the development of efficient computational algorithms, applications to modern machine learning models, and the investigation of topological invariants such as Betti numbers and homology groups in relation to classification complexity.

9. Concluding Remarks and Future Directions

This paper has introduced a new structural framework for studying feature selection in binary classification. Rather than concentrating on individual feature subsets or optimisation procedures, we have investigated the entire family of feature subsets that preserve the optimal Bayes decision rule.

The principal observation is that the collection of all decision-sufficient feature sets forms an order filter in the Boolean lattice of feature subsets. Its complement is therefore an abstract simplicial complex, called the *decision complex*. This establishes a direct connection between statistical decision theory and combinatorial topology.

Three associated invariants naturally emerge from this framework. The *decision core* identifies universally indispensable features, the *decision dimension* measures the minimum number of features required for optimal decision making, and the *decision complex* records the global topological organisation of feature subsets that fail to preserve the Bayes decision.

The theory was developed for arbitrary binary classification problems and illustrated explicitly for Gaussian-discriminant models, where the decision core coincides with the support of the discriminant vector and the decision dimension equals the number of non-zero discriminant coefficients. Thus, the abstract concepts introduced here recover familiar statistical quantities in a classical setting while extending naturally to more general models.

The present work is primarily theoretical. Nevertheless, the proposed framework suggests several promising directions for future investigation.

These include:

1. The development of efficient algorithms for identifying decision-sufficient feature sets;
2. Computational methods for constructing decision complexes for high-dimensional classification problems;
3. The use of topological invariants, including Euler characteristics, Betti numbers, and homology groups, as descriptors of classification complexity;
4. Extensions to multiclass classification, regression, and sequential decision problems;
5. Relationships with graphical models, sparse learning, rough-set theory, interpretable machine learning, and explainable artificial intelligence, together with applications of the proposed framework to decision analysis problems in which rough-set methods have proved effective [12–14];
6. Empirical investigations comparing decision cores with existing feature-selection methodologies on benchmark datasets.

Although many of these questions remain open, the present work provides a mathematical foundation upon which such developments may be built.

More broadly, the results demonstrate that every binary classification problem possesses not only probabilistic and decision-theoretic structure, but also an intrinsic combinatorial and topological structure. We hope that this viewpoint will encourage further interaction between statistical learning theory, combinatorics, topology, and machine learning.

Funding: This research received no external funding

Data Availability Statement: No data were produced for this research.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Berger, J.O. *Statistical Decision Theory and Bayesian Analysis*, 2nd ed.; Springer: New York, NY, USA, 1985.
2. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; Wiley: Hoboken, NJ, USA, 2006.
3. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning*, 2nd ed.; Springer: New York, NY, USA, 2009.
4. Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer: New York, NY, USA, 2006.
5. Guyon, I.; Elisseeff, A. An introduction to variable and feature selection. *J. Mach. Learn. Res.* **2003**, *3*, 1157–1182.
6. Peng, H.; Long, F.; Ding, C. Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1226–1238.
7. Pearl, J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*; Morgan Kaufmann: San Mateo, CA, USA, 1988.
8. Koller, D.; Friedman, N. *Probabilistic Graphical Models: Principles and Techniques*; MIT Press: Cambridge, MA, USA, 2009.
9. Billingsley, P. *Probability and Measure*, 3rd ed.; Wiley: New York, NY, USA, 1995.
10. Lauritzen, S.L. *Graphical Models*; Oxford University Press: Oxford, UK, 1996.
11. Fisher, R.A. On the mathematical foundations of theoretical statistics. *Philos. Trans. R. Soc. A* **1922**, *222*, 309–368.
12. El-Gayar, M.A.; Madbouly, N.M.; Nawar, A.S. Enhancing Economic Decision-Making through the Application of Alpha Basic-Rough Sets. *Int. J. Fuzzy Log. Intell. Syst.* **2026**, *26*, 98–115.
13. El-Gayar, M.A.; Kandil, S.A.; El-Bably, M.K. On Nearly Basic Rough Sets and Their Medical Applications. *Malays. J. Math. Sci.* **2026**, *20*, 597.
14. El-Bably, M.K.; Hosny, R.A.; El-Gayar, M.A. Innovative rough set approaches using novel initial-neighborhood systems: Applications in medical diagnosis of COVID-19 variants. *Inf. Sci.* **2025**, *708*, 122044.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.