Théorie de la Concordance Discrète : Déterminant SAT de l'AUROC et Limites de la Résolvabilité

Charles Dana

Décembre 2025

Abstract

Nous définissons le plafond théorique de l'Area Under the Receiver Operating Curve (AUROC) en reformulant l'apprentissage supervisé comme un problème de satisfiabilité (SAT) discret. En nous appuyant sur le Théorème de Dana, qui établit une correspondance entre datasets finis et représentations SAT en temps $O(mn^2)$, nous dérivons une fonction de score $s(x)$ fondée sur le ratio de *lookalikes* logiques. Nous démontrons que l'AUROC converge vers un déterminant de concordance discrète, offrant une explication structurelle aux plafonds de performance observés dans les modèles d'ensemble (RF/GB).

1 Introduction

Contrairement aux approches statistiques traitant l'AUROC comme une métrique empirique de classement, nous soutenons que la classification binaire est intrinsèquement un problème discret. Une fois les caractéristiques fixées et le dataset fini, l'apprentissage devient la construction d'une fonction de décision booléenne consistante avec les étiquettes.

2 Le Théorème de Dana et la Représentabilité

Théorème 1 (Dana, 2024)**.** *Pour toute matrice binaire $A \in \{0,1\}^{n\times m}$ à lignes distinctes et vecteur d'étiquettes $X \in \{0,1\}^n$, il existe une formule en forme normale conjonctive (CNF) ϕ telle que :*

1. *$\phi(A_{i,*}) = X_i$ pour tout $i \leq n$.*
2. *Le nombre de clauses $|F| \leq n$.*
3. *La construction s'effectue en temps $O(mn^2)$.*

Ce théorème garantit l'existence d'un classificateur parfait pour tout dataset fini, déplaçant l'enjeu de l'apprentissage de la *représentabilité* vers la *généralisation logiquement contrainte*.

3 Mécanique de la Concordance Discrète

Proposition 1 (Déterminant de l'AUROC)**.** *L'AUROC d'un classificateur logique construit par résolution de conflits par paires converge vers le différentiel des espérances de satisfaction :*

$$AUROC = \Pr\left[\mathbb{E}(T_L|X_T)\mathbb{E}(F_L|X_F) - \mathbb{E}(F_L|X_T)\mathbb{E}(T_L|X_F) > 0\right] \tag{1}$$

Proof. Soit un ensemble d'entraînement $\mathcal{D} = \{(X_i, y_i)\}_{i=1}^n$ avec $X_i \in \mathbb{R}^m$ et $y_i \in \{0, 1\}$. Nous définissons l'espace des littéraux admissibles $\mathcal{L}$ par l'ensemble des hyperplans médiateurs séparant les classes opposées :

$$\mathcal{L} = \left\{ (k, \tau, \text{pol}) \mid k \in [1, m], \tau = \frac{X_{ik} + X_{jk}}{2}, y_i \neq y_j \right\} \tag{2}$$

Le cardinal de $\mathcal{L}$ est fini et borné par $O(mn^2)$. Par extension, l'espace des clauses C (conjonctions de littéraux) et l'univers des modèles SAT $\varphi \in \Phi$ sont finis.

1. Construction de la Clause Minimale Pour un échantillon cible F de la classe 0, une clause C est générée par sélection aléatoire de littéraux dans $\mathcal{L}$ jusqu'à ce que C discrimine F de l'ensemble des points T de la classe 1. Le processus de *pruning* garantit que C est minimale au sens de l'inclusion des littéraux. La répétition de ce processus produit un modèle $\varphi(\omega)$ satisfaisant le Théorème de Dana : $\forall i, \varphi(X_i) = y_i$.

2. Condition de Lookalike par Élimination Pour une réalisation ω, nous définissons la relation de proximité logique (*lookalike*) entre un point de test x et un point d'entraînement X_i de même classe cible. Initialement, x est considéré lookalike de tout X_i. Pour chaque clause $C \in \varphi(\omega)$, si x satisfait la condition de discrimination ($C(x) = 1$) alors que X_i ne la satisfait pas ($C(X_i) = 0$), l'association est rompue. Ainsi :

$$x \text{ lookalike de } X_i \iff \forall C \in \varphi(\omega), [C(X_i) = 0 \Rightarrow C(x) = 0] \tag{3}$$

3. Convergence du Score Le score $s(x)$ est défini comme le ratio des densités de lookalikes observées sur K couches :

$$s(x) = \frac{\sum_{k=1}^K \#\{i : y_i = 1, i \in \text{lookalike}(x, \omega_k)\}}{\sum_{k=1}^K \#\{j : j \in \text{lookalike}(x, \omega_k)\}} \tag{4}$$

Par la Loi Forte des Grands Nombres, lorsque $K \to \infty$, ce ratio converge vers le rapport des espérances de satisfaction sur l'espace discret fini Ω :

$$s(x) \xrightarrow{a.s.} \frac{\mathbb{E}_\omega[T_L|x]}{\mathbb{E}_\omega[T_L|x] + \mathbb{E}_\omega[F_L|x]} \tag{5}$$

Où T_L (resp. F_L) est la variable aléatoire dénombrant les lookalikes de classe 1 (resp. 0).

4. Dérivation de l'AUROC L'AUROC est défini par $\Pr(s(X_T) > s(X_F))$ pour un couple (X_T, X_F) de vrais positifs et vrais négatifs. En substituant l'expression limite de $s(x)$ et par linéarité de l'espérance, la condition $s(X_T) > s(X_F)$ est algébriquement équivalente à :

$$\frac{\mathbb{E}(T_L|X_T)}{\mathbb{E}(T_L|X_T) + \mathbb{E}(F_L|X_T)} > \frac{\mathbb{E}(T_L|X_F)}{\mathbb{E}(T_L|X_F) + \mathbb{E}(F_L|X_F)} \tag{6}$$

Le produit en croix de cette inégalité, après simplification des termes communs, donne précisément le déterminant de concordance :

$$\mathbb{E}(T_L|X_T)\mathbb{E}(F_L|X_F) - \mathbb{E}(F_L|X_T)\mathbb{E}(T_L|X_F) > 0 \tag{7}$$

La finitude de Ω et la construction polynomiale $O(mn^2)$ assurent que cet équilibre de concordance est une propriété structurelle stable du dataset. □

4 Conclusion

En cadrant la classification binaire comme un problème de satisfiabilité (SAT) discret, ces travaux apportent une explication de principe aux plafonds de performance de l'AUROC observés dans les données réelles. Cette théorie démontre que le plafond de performance est intrinsèquement dicté par la structure logique des *features* plutôt que par la sophistication de l'algorithme d'apprentissage.

Toutefois, deux nuances fondamentales doivent être soulignées :

1. **Performance relative :** Bien qu'il serve de base théorique, l'un des constats empiriques est que l'`AlgorithmeClassifier` ne performe pas nécessairement sur tous les jeux de données aussi bien que les méthodes d'ensemble de type *Random Forest* (RF) ou *Gradient Boosting* (GB).

2. **Agnosticisme et Optimalité :** L'AUROC peut être artificiellement amélioré par une connaissance préalable de la base de test si le modèle n'est pas strictement agnostique aux données. Par conséquent, nous définissons l'**AUROC Optimal Théorique** comme la limite supérieure pour un modèle n'ayant accès qu'aux seules informations de l'entraînement (X_{train}, y_{train}).

Dans ce cadre, l'erreur résiduelle (souvent proche de $0,001$) représente des instances « logiquement dures » dont la résolution exigerait une recherche exponentielle, marquant la frontière pratique de la complexité NP-difficile dans le scoring de décision.

Une implémentation de l'`AlgorithmeClassifier` (basée sur l'algorithme « Algorithme Snake ») est disponible à la racine de ce dépôt GitHub et démontre empiriquement un AUROC Optimal théorique dans une construction $O(mn^2)$ pour l'entraînement et $O(mnn_{test})$ pour l'inférence.