



Compte rendu Projet SOD322 - Operational Research for machine learning

Optimal research Tree

Réalisé par :
HELA MHIRI
MAHDI ATTIA

Filière :
SCIENCE DE L'OPTIMISATION ET DES DONNÉES

Encadré par :
M. ALES ZACHARIE

Année universitaire : 2023-2024

Table des matières

1	Formulations F	5
1.1	Formulation F sans inégalités valides	6
1.2	Formulation F avec inégalités valides	7
2	Formulation F_U	9
2.1	Formulation F_U avec <i>simpleMerge</i>	9
2.1.1	Formulation F_U avec <i>simpleMerge</i> sans inégalité valide . . .	10
2.1.2	Formulation F_U avec <i>simpleMerge</i> avec inégalité valide . . .	15
2.2	Formulation F_U avec <i>Spectral Clustering</i>	19
2.2.1	Formulation F_U avec <i>Spectral Clustering</i> sans inégalités valides	20
2.2.2	Formulation F_U avec <i>Spectral Clustering</i> avec inégalités valides	26

Remerciement

Nous tenons à remercier notre enseignant M. Ales ZACHARIE pour toutes les connaissances qu'ils nous a communiquées tout au long de ce cours.

Introduction

Dans le domaine en constante évolution de la science des données et de l'apprentissage automatique, la construction d'arbres de décision optimaux représente un défi crucial pour la classification et l'analyse prédictive. Notre projet, encadré par Zacharie Ales pour l'année universitaire 2022-2023, se concentre sur l'application de la modélisation F à divers jeux de données pour générer des arbres de décision optimaux. Cette approche est explorée à travers l'utilisation de séparations univariées et multivariées ainsi que des méthodes de regroupement des données, à la fois naïves et exactes, présentées au cours.

Le projet exploite trois jeux de données initiaux — Iris, Wine, et Seeds — comme plateforme de test pour évaluer l'efficacité de notre modélisation. En outre, il se propose d'étendre cette exploration à deux autres ensembles de données de notre choix, afin de valider la robustesse et l'adaptabilité de nos méthodes. Nous avons choisi de travailler avec ces deux jeux de données :

- Mushroom : Ce jeu de données comprend des descriptions d'échantillons hypothétiques correspondant à 23 espèces de champignons à lamelles de la famille Agaricus et Lepiota. Chaque espèce est identifiée comme définitivement comestible, définitivement toxique, ou d'édibilité inconnue et non recommandée. Cette dernière classe a été combinée avec la classe toxique.
- Dry_bean_seeds : Ce jeu de données correspond à la classification de sept types d'haricots à partir de plusieurs caractéristiques y compris 12 dimensions et 4 formes.

Le cœur de notre investigation réside dans l'application des fonctions `main()` et `main-merge()`, conçues pour traiter respectivement les données sans et avec regroupement préalable, en fonction de différents paramètres de profondeur d'arbre ($D \in \{2, 3, 4\}$).

À travers ce rapport, nous ambitionnons de présenter une analyse complète des résultats obtenus en mettant l'accent sur les temps de calcul et la performance des classifieurs en fonction des jeux de données, de la profondeur des arbres, du type de séparations et de la stratégie de regroupement utilisée. Nous abordons également une question d'ouverture, choisie parmi plusieurs proposées, dans le but d'améliorer soit les temps de calcul, soit la précision des prédictions, via des

méthodes innovantes de regroupement, l'utilisation d'inégalités valides ou d'autres approches permettant d'affiner la formulation F.

Cette exploration se veut non seulement un exercice académique pour approfondir notre compréhension des arbres de décision optimaux mais aussi une contribution modeste à la vaste quête d'efficacité dans le domaine de l'apprentissage automatique.

1

Formulations F

Dans cette partie on va travailler avec la formulation F sans les regroupements. On va tout d'abord essayer cette formulation sans ajouter des contraintes valides puis on ajoutera deux contraintes valides qui sont :

- Première inégalité validée :

$$\sum_{k \in K} c_{k,t} + u_{i,t,w} + u_{i,t,l}(t) + u_{i,t,r}(t) \leq 1 \quad \forall k \in K, \forall i \in I \setminus I_k \quad (1.0.1)$$

- Si le sommet t prédit la classe k ($c_{k,t} = 1$), alors la donnée i qui n'est pas de classe k , ne peut pas atteindre le sommet t ($u_{i,t,w} + u_{i,t,l}(t) + u_{i,t,r}(t) = 0$).
- Si le sommet t ne prédit pas la classe k ($c_{k,t} = 0$), si i atteint t , elle va soit à droite, soit à gauche, soit au puits ($u_{i,t,w} + u_{i,t,l}(t) + u_{i,t,r}(t) \leq 1$) et si i n'atteint pas t , $u_{i,t,w} + u_{i,t,l}(t) + u_{i,t,r}(t) = 0$.
- Deuxième inégalité valide :

$$\sum_{j \in J} a_{j,t} + \sum_{k \in K} \sum_{a \in A(t)} c_{k,ta} = 1 \quad \forall t \in N; j \in J, k \in K; ta \in t \cup A(t) \quad (1.0.2)$$

- Si le sommet t effectue une séparation ($\sum_{j \in J} a_{j,t} = 1$), alors ni t , ni aucun de ses ancêtres n'est une feuille ($\sum_{k \in K} \sum_{a \in A(t)} c_{k,ta} = 0$), $k \in K$, $ta \in t \cup A(t)$.
- Si le sommet t n'effectue pas de séparation ($\sum_{j \in J} a_{j,t} = 0$), alors soit lui, soit un de ses ancêtres doit prédire une classe ($\sum_{k \in K} \sum_{a \in A(t)} c_{k,ta} = 1$).

Pour comparer ces deux approches, on utilisera `main()` avec un argument "inegalite" qui sera vrai si l'on souhaite incorporer les inégalités et faux sinon.

1.1 Formulation F sans inégalités valides

Dataset	iris	D	Univarié					Multivarié				
			Solve time	Gap(%)	Errors train/test	Node count	Constraint count	Solve time	Gap(%)	Errors train/test	Node count	Constraint count
Train size	120	2	7.6	0.0	5/1	2382	3133	2.4	0.0	1/3	921	2841
Test size	30	2	41.1	0.0	0/4	7351	6989	3.5	0.0	0/2	70	6309
Features count	4	3	56.7	0.0	0/1	7251	14701	6.8	0.0	0/0	151	13245
Time limit	1000	4										

TABLE 1.1 – Resultats pour les données Iris avec Formulation F

Dataset	seeds	D	Univarié					Multivarié				
			Solve time	Gap(%)	Errors train/test	Node count	Constraint count	Solve time	Gap(%)	Errors train/test	Node count	Constraint count
Train size	168	2	32.6	0.0	10/3	14464	4381	3.4	0.0	0/1	218	3975
Test size	42	2	1000.7	2.4	4/2	62837	9773	8.0	0.0	0/1	0	8827
Features count	6	3	472.6	0.0	0/4	49932	20557	46.6	0.0	0/6	4907	18531
Time limit	1000	4										

TABLE 1.2 – Resultats pour les données Seeds avec Formulation F

Dataset	wine	D	Univarié					Multivarié				
			Solve time	Gap(%)	Errors train/test	Node count	Constraint count	Solve time	Gap(%)	Errors train/test	Node count	Constraint count
Train size	142	2	27.1	0.0	5/2	6161	3705	1.1	0.0	0/2	0	3482
Test size	36	2	195.7	0.0	0/3	9893	8265	4.2	0.0	0/1	0	7746
Features count	13	3	177.1	0.0	0/3	14762	17385	9.5	0.0	0/3	0	16274
Time limit	1000	4										

TABLE 1.3 – Resultats pour les données Wine avec Formulation F

Dataset	dry_bean	D	Univarié					Multivarié				
			Solve time	Gap(%)	Errors train/test	Node count	Constraint count	Solve time	Gap(%)	Errors train/test	Node count	Constraint count
Train size	117	2	50.4	0.0	49/14	19189	3055	0.8	0.0	46/18	0	2952
Test size	30	2	1001.0	11.4	12/4	88362	6815	6.3	0.0	0/2	0	6576
Features count	16	3	1000.3	2.6	3/3	44262	14335	20.5	0.0	0/4	626	13824
Time limit	1000	4										

TABLE 1.4 – Resultats pour les données Dry_bean avec Formulation F

Dataset	mushroom	D	Univarié					Multivarié				
			Solve time	Gap(%)	Errors train/test	Node count	Constraint count	Solve time	Gap(%)	Errors train/test	Node count	Constraint count
Train size	120	2	73.9	0.0	6/1	14783	3133	0.7	0.0	0/2	0	3066
Test size	30	2	64.9	0.0	0/0	6057	6989	0.4	0.0	0/2	0	6834
Features count	19	3	122.7	0.0	0/2	8675	14701	0.6	0.0	0/2	0	14370
Time limit	1000	4										

TABLE 1.5 – Resultats pour les données Mushroom avec Formulation F

Dans les tableaux ci-dessus, les résultats obtenus en testant la fonction `main()` avec l'argument (`inegalité=false`) sur les ensembles de données nous montrent que le cas multivarié a généralement des performances supérieures au cas univarié, comme indiqué par les erreurs observées pour différentes valeurs de la séparation maximale D . De plus, le temps de calcul pour le cas multivarié est inférieur à celui du cas univarié. En règle générale, à mesure que la taille de l'arbre augmente, c'est-à-dire lorsque D est élevé, l'erreur Train va diminuer. On peut observer ces résultats dans les tableaux 1.1 et 1.2.

1.2 Formulation F avec inégalités valides

Dataset	iris	D	Univarié					Multivarié				
			Solve time	Gap(%)	Errors train/test	Node count	Constraint count	Solve time	Gap(%)	Errors train/test	Node count	Constraint count
Train size	120	2										
Test size	30	2	4.5	0.0	5/1	2269	3856	0.7	0.0	1/3	0	3564
Features count	4	3	9.1	0.0	0/3	2244	8676	3.8	0.0	0/4	4301	7996
Time limit	1000	4	14.8	0.0	0/1	1801	18316	14.7	0.0	0/5	4286	16860

TABLE 1.6 – Résultats pour les données Iris avec Formulation F avec inegalités valides

Dataset	seeds	D	Univarié					Multivarié				
			Solve time	Gap(%)	Errors train/test	Node count	Constraint count	Solve time	Gap(%)	Errors train/test	Node count	Constraint count
Train size	168	2										
Test size	42	2	15.2	0.0	10/3	9550	5392	5.8	0.0	4/4	5875	4986
Features count	6	3	1000.1	1.8	4/1	125748	12132	28.5	0.0	0/3	15200	11186
Time limit	1000	4	458.8	0.0	0/4	43589	25612	47.4	0.0	0/4	10955	23586

TABLE 1.7 – Résultats pour les données Seeds avec Formulation F avec inegalités valides

Dataset	wine	D	Univarié					Multivarié				
			Solve time	Gap(%)	Errors train/test	Node count	Constraint count	Solve time	Gap(%)	Errors train/test	Node count	Constraint count
Train size	142	2										
Test size	36	2	11.7	0.0	5/2	7351	4560	1.1	0.0	0/2	0	4337
Features count	13	3	30.1	0.7	1/1	7024	10260	2.2	0.0	0/1	0	9741
Time limit	30	4	17.8	0.0	0/3	1728	21660	7.8	0.0	0/3	1724	20549

TABLE 1.8 – Résultats pour les données Wine avec Formulation F avec inegalités valides

Dataset	dry_bean	D	Univarié					Multivarié				
Train size	117		Solve time	Gap(%)	Errors train/test	Node count	Constraint count	Solve time	Gap(%)	Errors train/test	Node count	Constraint count
Test size	30	2	4.7	0.0	49/14	2412	5164	0.4	0.0	46/18	0	5061
Features count	16	3	1001.7	6.0	12/2	487795	11736	1.7	0.0	0/1	171	11497
Time limit	1000	4	1000.1	0.9	1/6	207479	24880	5.6	0.0	0/3	977	24369

TABLE 1.9 – Résultats pour les données dry_bean avec Formulation F avec inegalités valides

Dataset	mushroom	D	Univarié					Multivarié				
Train size	120		Solve time	Gap(%)	Errors train/test	Node count	Constraint count	Solve time	Gap(%)	Errors train/test	Node count	Constraint count
Test size	30	2	8.6	0.0	2/3	979	3496	1.1	0.0	0/2	1312	3429
Features count	19	3	50.5	0.0	0/2	5040	7836	5.0	0.0	0/4	2006	7681
Time limit	1000	4	83.8	0.0	0/4	4544	16516	41.6	0.0	0/1	22559	16185

TABLE 1.10 – Résultats pour les données Mushroom avec Formulation F avec inegalités valides

Dans le contexte de la résolution sans regroupement, on a initié une évaluation de l'efficacité des inégalités valides ajoutées. on a observé que dans la majorité des cas, l'introduction de ces inégalités se traduit par une réduction de la vitesse de résolution, ainsi que par une augmentation du nombre de nœuds et de contraintes. Contrairement à ce que l'on pourrait attendre, ces inégalités ne semblent pas améliorer significativement la précision ni l'écart optimal pour l'ensemble des données. Cependant, une légère amélioration a été remarquée dans le cas du jeu de données "dry bean", où l'écart d'optimalité a diminué pour les valeurs de D égales à 3 et 4.

2

Formulation F_U

Dans cette partie, on adopte la formulation F_U qui correspond à la formulation F avec des regroupements. Comme vu en cours, cette approche permet de réduire le nombre de contraintes. Afin de construire des regroupements de données et de tester cette méthode sur les jeux de données utilisés, nous considérons deux algorithmes différents de regroupement : le premier est le *simpleMerge()*, qui construit des regroupements itérativement en fusionnant des clusters ; le deuxième est l'algorithme *Spectral Clustering*, qui se base sur l'analyse spectrale et l'algèbre linéaire.

2.1 Formulation F_U avec *simpleMerge*

Dans la fonction *main-merge()*, la sous routine *simpleMerge()* était utilisée pour créer les groupements de données (clusters) pour la formulation F_U . Cette fonction implémente un algorithme de *clustering* itératif basé sur la similarité des points de données, tout en assurant que seuls les points appartenant à la même classe soient fusionnés. Initialement, chaque point de donnée est considéré comme un cluster individuel. L'algorithme calcule ensuite les distances entre toutes les paires de points appartenant à la même classe et trie ces distances par ordre croissant. Il procède à la fusion des clusters en commençant par les paires de points les plus proches, c'est-à-dire ceux ayant la distance la plus courte. Cette opération est répétée itérativement, en fusionnant progressivement les clusters, jusqu'à atteindre un seuil spécifié par le paramètre *gamma*, qui détermine le nombre de clusters

souhaité en fonction du nombre initial de points de données.

2.1.1 Formulation F_U avec *simpleMerge* sans inégalité valide

Pour analyser l'impact de cette méthode de regroupement sur les performance du solveur, on effectue une liste de tests où l'on fait varier *gamma* entre 0.0 et 1.0 avec un pas de 0.2 afin de contrôler le nombre de clusters formés. Dans tous les tests, on choisit une limite de temps de calcul de 900s.

Dataset=iris		Train size=120		Test size=30		Feature count =4		Time limit =900		Inequalities =Deactivated	
D 2		Univarié					Multivarié				
Gamma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.0	3	2.1	0.0	38/12	0	82	0.1	0.0	38/12	0	852
0.2	24	0.2	0.0	5/1	403	565	0.2	0.0	1/0	0	1209
0.4	48	0.6	0.0	5/2	320	1117	0.4	0.0	1/0	0	1617
0.6	72	0.8	0.0	5/1	707	1669	0.7	0.0	1/3	1424	2025
0.8	96	1.6	0.0	5/1	1364	2221	1.5	0.0	1/3	887	2433
1.0	120	3.0	0.0	5/1	2549	2773	1.3	0.0	1/3	396	2841
D 3		Univarié					Multivarié				
Gamma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.0	3	0.0	0.0	38/12	0	182	0.2	0.0	38/12	74	1980
0.2	24	1.6	0.0	4/1	2857	1253	0.3	0.0	0/0	0	2757
0.4	48	1.1	0.0	0/1	974	2477	1.0	0.0	0/1	0	3645
0.6	72	2.9	0.0	0/1	3020	3701	0.7	0.0	0/0	0	4533
0.8	96	11.8	0.0	0/4	5781	4925	2.3	0.0	0/6	216	5421
1.0	120	5.5	0.0	0/1	3755	6149	4.3	0.0	0/4	110	6309
D 4		Univarié					Multivarié				
Gamma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.0	3	0.1	0.0	38/12	0	382	1.4	0.0	38/12	138	4236
0.2	24	12.3	0.0	3/1	7951	2629	0.8	0.0	0/1	0	5853
0.4	48	3.7	0.0	0/2	1911	5197	1.7	0.0	0/0	0	7701
0.6	72	13.6	0.0	0/1	5372	7765	2.0	0.0	0/1	0	9549
0.8	96	36.7	0.0	0/4	6180	10333	17.9	0.0	0/0	4371	11397
1.0	120	67.8	0.0	0/2	5297	12901	7.0	0.0	0/6	440	13245

TABLE 2.1 – Résultats pour les données Iris avec Formulation F_U sans inégalités valides avec *simpleMerge*.

Dataset=wine		Train size=142		Test size=36		Feature count =13		Time limit =900		Inequalities =Deactivated	
D 2		Univarié					Multivarié				
Gamma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.0	3	0.0	0.0	56/15	0	82	0.1	0.0	0/2	0	1119
0.2	28	0.4	0.0	8/4	897	657	0.4	0.0	0/1	0	1544
0.4	56	1.4	0.0	7/3	1210	1301	0.1	0.0	0/2	0	2020
0.6	85	4.2	0.0	5/2	2741	1968	0.6	0.0	0/1	0	2513
0.8	113	30.3	0.0	5/2	9086	2612	0.6	0.0	0/1	0	2989
1.0	142	49.4	0.0	5/2	9635	3279	0.6	0.0	0/2	0	3482
D 3		Univarié					Multivarié				
Gamma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.0	3	0.0	0.0	56/15	0	182	0.2	0.0	0/1	0	2603
0.2	28	2.6	0.0	0/2	3764	1457	0.2	0.0	0/2	0	3528
0.4	56	12.7	0.0	1/2	6156	2885	1.7	0.0	0/4	0	4564
0.6	85	58.8	0.0	1/1	6695	4364	2.3	0.0	0/2	0	5637
0.8	113	102.6	0.0	0/5	13779	5792	2.8	0.0	0/3	0	6673
1.0	142	129.3	0.0	0/1	8029	7271	2.8	0.0	0/2	0	7746
D 4		Univarié					Multivarié				
Gamma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.0	3	0.1	0.0	56/15	99	382	0.3	0.0	0/1	0	5571
0.2	28	6.6	0.0	0/2	4090	3057	3.3	0.0	0/2	0	7496
0.4	56	74.2	0.0	0/1	9712	6053	3.1	0.0	0/3	0	9652
0.6	85	75.2	0.0	0/0	5793	9156	4.0	0.0	0/2	0	11885
0.8	113	298.7	0.0	0/2	22452	12152	5.5	0.0	0/3	0	14041
1.0	142	458.3	0.0	0/2	40660	15255	16.5	0.0	0/2	0	16274

TABLE 2.2 – Résultats pour les données Wine avec Formulation F_U sans inegalités valides avec *simpleMerge*.

Dataset=seeds		Train size=168		Test size=42		Feature count =6		Time limit =900		Inequalities =Deactivated	
D 2		Univarié					Multivarié				
Gamma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.0	3	0.0	0.0	53/17	0	82	0.3	0.0	53/17	0	1170
0.2	33	0.4	0.0	17/4	829	772	0.5	0.0	1/1	277	1680
0.4	67	0.8	0.0	14/3	503	1554	0.7	0.0	0/0	0	2258
0.6	100	2.7	0.0	11/2	3646	2313	1.4	0.0	0/1	279	2819
0.8	134	20.5	0.0	10/3	7828	3095	0.8	0.0	0/0	0	3397
1.0	168	35.3	0.0	10/3	11241	3877	2.8	0.0	0/2	400	3975
D 3		Univarié					Multivarié				
Gamma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.0	3	0.0	0.0	53/17	0	182	2.0	0.0	53/17	0	2722
0.2	33	7.7	0.0	19/7	7339	1712	0.9	0.0	0/2	0	3832
0.4	67	25.8	0.0	7/3	8470	3446	1.7	0.0	0/0	286	5090
0.6	100	151.0	0.0	4/2	31265	5129	2.8	0.0	0/2	104	6311
0.8	134	900.7	3.7	6/2	72317	6863	4.3	0.0	0/3	87	7569
1.0	168	916.8	2.4	4/2	42740	8597	4.8	0.0	0/3	0	8827
D 4		Univarié					Multivarié				
Gamma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.0	3	0.1	0.0	53/17	0	382	2.2	0.0	53/17	0	5826
0.2	33	14.5	0.0	19/4	6501	3592	1.8	0.0	0/2	0	8136
0.4	67	901.2	4.4	5/3	113805	7230	5.2	0.0	0/4	0	10754
0.6	100	548.1	0.0	0/1	74063	10761	9.7	0.0	0/2	0	13295
0.8	134	864.0	0.0	0/3	58991	14399	27.7	0.0	0/3	1392	15913
1.0	168	902.3	1.2	2/4	41442	18037	16.9	0.0	0/2	0	18531

TABLE 2.3 – Résultats pour les données Seeds avec Formulation F_U sans inegalités valides avec *simpleMerge*.

Dataset=mushroom		Train size=120		Test size=30		Feature count =22		Time limit =900		Inequalities =Deactivated	
D 2		Univarié					Multivarié				
Gamma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.0	2	0.0	0.0	60/15	0	59	0.1	0.0	0/3	0	1105
0.2	24	0.3	0.0	3/1	416	565	0.1	0.0	0/2	0	1479
0.4	48	0.7	0.0	3/1	773	1117	0.1	0.0	0/3	0	1887
0.6	72	2.0	0.0	3/1	1097	1669	0.1	0.0	0/3	0	2295
0.8	96	8.6	0.0	3/1	3036	2221	0.3	0.0	0/3	0	2703
1.0	120	13.9	0.0	3/1	3157	2773	0.2	0.0	0/3	0	3111
D 3		Univarié					Multivarié				
Gamma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.0	2	0.0	0.0	60/15	0	131	0.2	0.0	0/3	0	2573
0.2	24	1.2	0.0	0/0	1574	1253	0.2	0.0	0/5	0	3387
0.4	48	3.0	0.0	0/0	1199	2477	0.6	0.0	0/2	0	4275
0.6	72	33.7	0.0	0/0	7782	3701	1.0	0.0	0/3	0	5163
0.8	96	45.5	0.0	0/0	4398	4925	1.4	0.0	0/0	0	6051
1.0	120	43.8	0.0	0/0	5099	6149	1.1	0.0	0/3	0	6939
D 4		Univarié					Multivarié				
Gamma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.0	2	0.0	0.0	60/15	0	275	0.3	0.0	0/3	0	5509
0.2	24	0.8	0.0	0/0	30	2629	0.4	0.0	0/1	0	7203
0.4	48	6.4	0.0	0/2	2205	5197	2.3	0.0	0/1	0	9051
0.6	72	40.1	0.0	0/1	6781	7765	2.7	0.0	0/3	0	10899
0.8	96	10.8	0.0	0/0	1784	10333	3.2	0.0	0/3	0	12747
1.0	120	118.4	0.0	0/0	7480	12901	14.8	0.0	0/3	463	14595

TABLE 2.4 – Résultats pour les données Mushroom avec Formulation F_U sans inegalités valides avec *simpleMerge*.

Dataset=dry_bean		Train size=117		Test size=30			Feature count =16		Time limit =900		Inequalities =Deactivated		
D 2		Univarié						Multivarié					
Gamma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints		
0.0	7	0.1	0.0	48/16	12	174	0.6	0.0	46/18	0	1082		
0.2	23	0.6	0.0	48/16	761	542	1.7	0.0	46/18	0	1354		
0.4	46	1.7	0.0	82/24	0	1071	0.8	0.0	46/18	0	1745		
0.6	70	2.9	0.0	80/25	2334	1623	0.5	0.0	46/17	0	2153		
0.8	93	7.4	0.0	65/20	1798	2152	0.5	0.0	46/18	0	2544		
1.0	117	17.8	0.0	81/24	8322	2704	12.6	0.0	46/17	85	2952		
D 3		Univarié						Multivarié					
Gamma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints		
0.0	7	0.6	0.0	34/8	6920	386	4.0	0.0	15/6	733	2506		
0.2	23	10.6	0.0	11/7	8169	1202	1.6	0.0	0/5	0	3098		
0.4	46	155.0	0.0	24/9	45838	2375	2.2	0.0	0/5	0	3949		
0.6	70	902.8	3.7	26/11	170405	3599	14.0	0.0	0/4	4081	4837		
0.8	93	901.0	8.3	9/4	47662	4772	3.6	0.0	0/4	0	5688		
1.0	117	902.1	6.4	51/13	66843	5996	11.3	0.0	0/7	202	6576		
D 4		Univarié						Multivarié					
Gamma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints		
0.0	7	2.9	0.0	34/8	10383	810	15.7	0.0	15/6	6516	5354		
0.2	23	55.9	0.0	8/7	48690	2522	3.3	0.0	0/4	0	6586		
0.4	46	49.7	0.0	20/8	23279	4983	3.7	0.0	0/4	0	8357		
0.6	70	900.3	1.7	69/19	124722	7551	5.8	0.0	0/3	0	10205		
0.8	93	121.6	0.0	28/7	13046	10012	7.9	0.0	0/1	0	11976		
1.0	117	63.0	0.0	61/18	3514	12580	11.4	0.0	0/5	0	13824		

TABLE 2.5 – Résultats pour les données Dry-bean avec Formulation F_U sans inegalités valides avec *simpleMerge*.

D'après les resultats des test présentés ci dessus, on remarque que pour tous les jeux de données, le cas *Multivarié* est plus efficace en termes de nombre de points misclassifiés, temps de calcul (*Solve time*), nombre de neouds avec un nombre de contraintes légèrement plus grand que celui dans le cas *univarié*. En outre, on remarque qu'un nombre de clusters maximal n'est pas toujours optimal comme dans le cas univarié pour les jeux de données Iris, Wine et Seeds et dans le cas multivarié pour Mushroom et Dry_bean.

Pour le jeu de donné *Dry_bean*, on remarque que pour un arbre de profondeur $D = 2$, même le cas multivarié exhibe un grand nombre d'erreurs de misclassifications. Cette observation peut être expliqué par la nature des données et leur complexité, ce qui n'était pas le cas avec les autres jeux de données.

Finalement, en comparant ces résultats de la formulation F_U avec *simpleMerge* par rapport à la Formulation F dans la section précédente, on remarque que les performances ont deterioré montrant l'inefficacité de ce merge.

2.1.2 Formulation F_U avec *simpleMerge* avec inegalité valide

Dataset=iris		Train size=120		Test size=30		Feature count =4		Time limit =900		Inequalities =Activated	
D 2		Univarié					Multivarié				
Gamma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.0	3	0.0	0.0	38/12	0	103	0.1	0.0	38/12	0	873
0.2	24	0.1	0.0	5/1	0	712	0.1	0.0	4/4	0	1356
0.4	48	0.3	0.0	5/1	0	1408	3.8	0.0	1/3	0	1908
0.6	72	0.9	0.0	5/1	534	2104	0.4	0.0	1/3	0	2460
0.8	96	1.6	0.0	5/1	934	2800	0.7	0.0	1/3	153	3012
1.0	120	2.3	0.0	5/1	1465	3496	0.6	0.0	1/3	372	3564
D 3		Univarié					Multivarié				
Gamma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.0	3	0.0	0.0	38/12	0	231	0.1	0.0	38/12	0	2029
0.2	24	0.8	0.0	4/1	1822	1596	0.5	0.0	1/1	931	3100
0.4	48	1.4	0.0	0/4	768	3156	1.0	0.0	0/3	1127	4324
0.6	72	3.5	0.0	0/4	2555	4716	1.2	0.0	0/4	342	5548
0.8	96	5.9	0.0	0/4	5295	6276	2.1	0.0	0/5	1418	6772
1.0	120	12.6	0.0	0/1	6630	7836	1.9	0.0	0/4	778	7996
D 4		Univarié					Multivarié				
Gamma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.0	3	0.0	0.0	38/12	0	487	0.2	0.0	38/12	0	4341
0.2	24	3.9	0.0	3/2	5978	3364	1.3	0.0	0/4	963	6588
0.4	48	3.6	0.0	0/1	3176	6652	1.8	0.0	0/2	0	9156
0.6	72	6.0	0.0	0/1	2607	9940	3.4	0.0	0/2	1414	11724
0.8	96	5.2	0.0	0/2	1421	13228	6.1	0.0	0/5	5189	14292
1.0	120	11.8	0.0	0/4	3913	16516	6.7	0.0	0/2	4072	16860

TABLE 2.6 – Résultats pour les données Iris avec Formulation F_U avec inegalités valides avec *simpleMerge*.

Dataset=seeds		Train size=168		Test size=42		Feature count =6		Time limit =900		Inequalities =Activated	
D 2		Univarié					Multivarié				
Gamma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.0	3	0.0	0.0	53/17	0	103	0.1	0.0	53/17	0	1191
0.2	33	0.3	0.0	17/4	0	973	0.2	0.0	23/8	0	1881
0.4	67	1.1	0.0	14/3	187	1959	0.9	0.0	8/3	1105	2663
0.6	100	3.2	0.0	10/3	3096	2916	1.1	0.0	3/4	3328	3422
0.8	134	10.0	0.0	10/3	9757	3902	3.2	0.0	3/4	3747	4204
1.0	168	13.2	0.0	10/3	12956	4888	3.3	0.0	3/3	7492	4986
D 3		Univarié					Multivarié				
Gamma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.0	3	0.0	0.0	53/17	0	231	0.2	0.0	53/17	0	2771
0.2	33	3.5	0.0	21/6	6668	2181	1.5	0.0	20/5	1892	4301
0.4	67	9.9	0.0	7/3	4835	4391	3.1	0.0	0/3	4718	6035
0.6	100	40.7	0.0	4/2	12476	6536	18.0	0.0	0/3	15673	7718
0.8	134	213.9	0.0	3/2	54435	8746	22.8	0.0	0/2	16345	9452
1.0	168	586.0	0.0	3/4	105575	10956	17.4	0.0	0/1	8729	11186
D 4		Univarié					Multivarié				
Gamma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.0	3	0.0	0.0	53/17	0	487	0.4	0.0	53/17	0	5931
0.2	33	4.9	0.0	11/3	4436	4597	2.5	0.0	17/3	2191	9141
0.4	67	114.0	0.0	3/3	46925	9255	6.5	0.0	0/3	5329	12779
0.6	100	170.1	0.0	0/3	46166	13776	23.6	0.0	0/2	8626	16310
0.8	134	357.4	0.0	0/5	45811	18434	41.9	0.0	0/3	23334	19948
1.0	168	417.5	0.0	0/4	78888	23092	43.8	0.0	0/4	20417	23586

TABLE 2.7 – Résultats pour les données Seeds avec Formulation F_U avec inegalités valides avec *simpleMerge*.

Dataset=wine		Train size=142		Test size=36		Feature count =13		Time limit =900		Inequalities =Activated	
D	2	Univarié					Multivarié				
Gamma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.0	3	0.0	0.0	56/15	0	103	0.2	0.0	0/1	0	1140
0.2	28	0.2	0.0	10/5	0	828	0.2	0.0	0/1	0	1715
0.4	56	0.7	0.0	7/3	0	1640	0.3	0.0	0/1	0	2359
0.6	85	2.1	0.0	5/2	1093	2481	0.5	0.0	0/2	0	3026
0.8	113	5.6	0.0	5/2	5385	3293	0.7	0.0	0/1	0	3670
1.0	142	14.8	0.0	5/2	6919	4134	1.1	0.0	0/0	0	4337
D	3	Univarié					Multivarié				
Gamma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.0	3	0.0	0.0	56/15	0	231	0.4	0.0	0/1	0	2652
0.2	28	1.5	0.0	0/2	2671	1856	0.5	0.0	0/0	0	3927
0.4	56	4.6	0.0	1/1	5029	3676	0.7	0.0	0/1	0	5355
0.6	85	14.2	0.0	1/4	4866	5561	0.7	0.0	0/1	0	6834
0.8	113	18.9	0.0	0/3	6425	7381	2.3	0.0	0/1	304	8262
1.0	142	32.7	0.0	0/5	8961	9266	2.2	0.0	0/1	0	9741
D	4	Univarié					Multivarié				
Gamma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.0	3	0.1	0.0	56/15	0	487	0.7	0.0	0/1	0	5676
0.2	28	4.1	0.0	6/5	5459	3912	1.1	0.0	0/3	0	8351
0.4	56	45.1	0.0	0/2	13706	7748	1.4	0.0	0/1	0	11347
0.6	85	27.2	0.0	0/2	6123	11721	4.0	0.0	0/2	1202	14450
0.8	113	47.1	0.0	0/2	8510	15557	12.6	0.0	0/1	2959	17446
1.0	142	52.0	0.0	0/3	8249	19530	16.2	0.0	0/6	7030	20549

TABLE 2.8 – Résultats pour les données Wine avec Formulation F_U avec inegalités valides avec *simpleMerge*.

Dataset=mushroom		Train size=120		Test size=30		Feature count =19		Time limit =900		Inequalities =Activated	
D 2		Univarié					Multivarié				
Gamma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.0	2	0.0	0.0	59/16	0	68	0.1	0.0	59/16	0	1069
0.2	24	0.3	0.0	5/0	0	640	0.2	0.0	0/3	0	1509
0.4	48	1.4	0.0	5/0	593	1264	0.2	0.0	0/3	0	1989
0.6	72	3.3	0.0	5/0	1540	1888	0.8	0.0	0/3	3137	2469
0.8	96	6.4	0.0	5/0	3296	2512	0.6	0.0	0/1	1586	2949
1.0	120	12.7	0.0	5/0	6128	3136	1.1	0.0	0/3	1629	3429
D 3		Univarié					Multivarié				
Gamma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.0	2	0.0	0.0	59/16	0	152	0.3	0.0	59/16	0	2489
0.2	24	1.4	0.0	0/0	802	1428	0.6	0.0	0/0	0	3457
0.4	48	5.3	0.0	0/0	2942	2820	0.5	0.0	0/1	0	4513
0.6	72	11.0	0.0	0/1	5349	4212	1.0	0.0	0/2	0	5569
0.8	96	24.7	0.0	0/0	6706	5604	1.5	0.0	0/3	1832	6625
1.0	120	22.5	0.0	0/0	4724	6996	1.5	0.0	0/3	1138	7681
D 4		Univarié					Multivarié				
Gamma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.0	2	0.0	0.0	59/16	0	320	0.6	0.0	59/16	0	5329
0.2	24	1.9	0.0	0/0	1447	3004	0.8	0.0	0/2	0	7353
0.4	48	5.0	0.0	0/2	2655	5932	1.5	0.0	0/2	72	9561
0.6	72	31.8	0.0	0/2	8788	8860	2.2	0.0	0/2	0	11769
0.8	96	30.7	0.0	0/0	3923	11788	4.9	0.0	0/1	2656	13977
1.0	120	51.4	0.0	0/0	5212	14716	7.1	0.0	0/0	6195	16185

TABLE 2.9 – Résultats pour les données Mushroom avec Formulation F_U avec inegalités valides avec *simpleMerge*.

Dataset=dry_bean		Train size=117		Test size=30		Feature count =16		Time limit =900		Inequalities =Activated	
D 2		Univarié					Multivarié				
Gamma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.0	7	0.1	0.0	46/18	0	303	0.2	0.0	43/23	0	1211
0.2	23	0.3	0.0	46/18	0	959	0.3	0.0	43/23	0	1771
0.4	46	0.7	0.0	46/19	0	1902	0.3	0.0	43/23	0	2576
0.6	70	1.3	0.0	45/21	0	2886	0.4	0.0	43/22	0	3416
0.8	93	1.7	0.0	45/21	0	3829	0.4	0.0	43/21	0	4221
1.0	117	3.2	0.0	45/21	497	4813	0.4	0.0	43/22	0	5061
D 3		Univarié					Multivarié				
Gamma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.0	7	0.1	0.0	33/11	0	687	0.6	0.0	0/4	0	2807
0.2	23	2.5	0.0	9/7	3641	2175	0.7	0.0	0/3	0	4071
0.4	46	13.5	0.0	6/7	5263	4314	0.9	0.0	0/2	0	5888
0.6	70	18.9	0.0	5/7	6450	6546	1.6	0.0	0/3	0	7784
0.8	93	80.2	0.0	5/6	21903	8685	1.7	0.0	0/3	0	9601
1.0	117	230.7	0.0	5/6	51824	10917	1.4	0.0	0/2	0	11497
D 4		Univarié					Multivarié				
Gamma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.0	7	0.6	0.0	33/11	2686	1455	1.4	0.0	0/3	0	5999
0.2	23	22.7	0.0	10/7	13237	4607	1.4	0.0	0/8	0	8671
0.4	46	93.9	0.0	4/7	27901	9138	2.0	0.0	0/6	0	12512
0.6	70	53.2	0.0	0/7	10214	13866	2.5	0.0	0/3	0	16520
0.8	93	236.8	0.0	5/7	43942	18397	10.2	0.0	0/5	1576	20361
1.0	117	71.2	0.0	0/7	7865	23125	10.3	0.0	0/4	1736	24369

TABLE 2.10 – Résultats pour les données Dry_bean avec Formulation F_U avec inegalités valides avec *simpleMerge*.

Il est remarquable que l’ajout des inégalités entraîne une diminution de la vitesse de résolution ainsi qu’une augmentation du nombre de nœuds et de contraintes effectués. Contrairement à ce qui était prévu, ces inégalités ne contribuent ni à l’amélioration de la précision ni à la réduction de l’écart d’optimalité pour l’ensemble des données.

2.2 Formulation F_U avec *Spectral Clustering*

Pour varier les méthodes de *clustering* utilisées, On propose de considérer l’algorithme de *Spectral Clustering* (*SC*). Bien que cet algorithme utilise une sous routine itérative dans son processus d’optimisation (K-means), il n’adopte pas une approche itérative pour la fusion de clusters basée sur la similarité ou la distance entre les clusters eux-mêmes comme le fait l’algorithme *simpleMerge*.

En outre, *SC* ne garantit pas l’homogénéité des clusters en termes de classes. En effet, l’idée de base de la methode de spectral clustering est de transformer les données d’entrée en un graphe pondéré, où chaque donnée est représentée par un noeud et le poids entre les noeuds mesure la similitude entre les données.

Cette transformation est effectuée en utilisant une mesure de similarité telle que la distance euclidienne. Une fois que le graphe est construit, l'algorithme de Spectral Clustering calcule les vecteurs propres de la matrice Laplacienne normalisée du graphe, qui est une matrice symétrique définie positive. Les vecteurs propres sont triés par leurs valeurs propres correspondantes, ce qui permet de sélectionner les premiers vecteurs qui représentent les directions principales du graphe. Les données sont ensuite projetées sur ces vecteurs propres pour obtenir une représentation en dimension réduite. Cette représentation est ensuite utilisée pour effectuer une classification, ici en utilisant l'algorithme *K-means*.

L'algorithme de *SC* dépend principalement de deux paramètres qu'on fait varier dans les tests suivants. Précisamment, on a le paramètre *sigma* qui controle la variance du noyau gaussien utilisé pour calculer la similarité entre les points de données et on a aussi le paramètre *k* de la sous routine *k-means* qui controle le nombre de regroupements.

2.2.1 Formulation F_U avec *Spectral Clustering* sans inégalités valides

Dans les expériences suivantes, nous fixons la profondeur de l'arbre à $D = 4$. Pour la majorité des jeux de données, nous remarquons que parmi les combinaisons de paramètres (σ, k) , celles ayant les meilleures performances sont celles avec $k = 15$ et un *sigma* particulier. Cependant, cette méthode de clustering n'améliore pas les résultats obtenus avec le *simpleMerge*.

Dataset=seeds		Train size=168		Test size=42		Feature count =6		Time limit =900		Inequalities =Deactivated	
k	1	Univarié					Multivarié				
Sigma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.06945894944063327	1	1.0	0.0	112/28	0	168	0.2	0.0	112/28	0	5702
0.7571025489029025	1	0.0	0.0	112/28	0	168	0.2	0.0	112/28	0	5702
1.4447461483651718	1	0.0	0.0	112/28	0	168	0.2	0.0	112/28	0	5702
2.132389747827441	1	0.0	0.0	112/28	0	168	0.2	0.0	112/28	0	5702
2.8200333472897103	1	0.0	0.0	112/28	0	168	0.2	0.0	112/28	0	5702
3.5076769467519795	1	0.0	0.0	112/28	0	168	0.2	0.0	112/28	0	5702
4.1953205462142495	1	0.0	0.0	112/28	0	168	0.2	0.0	112/28	0	5702
4.882964145676518	1	0.0	0.0	112/28	0	168	0.2	0.0	112/28	0	5702
5.570607745138788	1	0.0	0.0	112/28	0	168	0.2	0.0	112/28	0	5702
6.2582513446010575	1	0.0	0.0	112/28	0	168	0.2	0.0	112/28	0	5702
6.945894944063326	1	0.0	0.0	112/28	0	168	0.2	0.0	112/28	0	5702
k	5	Univarié					Multivarié				
Sigma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.06945894944063327	5	0.0	0.0	112/28	0	596	1.3	0.0	112/28	772	6130
0.7571025489029025	5	0.2	0.0	112/28	686	596	1.9	0.0	79/14	3876	6130
1.4447461483651718	5	0.2	0.0	112/28	1839	596	1.4	0.0	65/15	1582	6130
2.132389747827441	5	0.2	0.0	112/28	2236	596	3.1	0.0	112/28	3462	6130
2.8200333472897103	5	0.2	0.0	117/31	578	596	3.1	0.0	115/25	4623	6130
3.5076769467519795	5	0.2	0.0	112/28	836	596	1.8	0.0	65/15	1423	6130
4.1953205462142495	5	0.2	0.0	112/28	887	596	1.8	0.0	65/15	1430	6130
4.882964145676518	5	0.1	0.0	112/28	628	596	1.7	0.0	112/28	1832	6130
5.570607745138788	5	0.1	0.0	112/28	572	596	3.8	0.0	112/28	8121	6130
6.2582513446010575	5	0.0	0.0	112/28	0	596	2.0	0.0	112/28	684	6130
6.945894944063326	5	0.2	0.0	112/28	1623	596	1.8	0.0	65/15	3341	6130
k	10	Univarié					Multivarié				
Sigma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.06945894944063327	10	0.8	0.0	112/28	2798	1131	15.8	0.0	57/17	21740	6665
0.7571025489029025	10	2.4	0.0	112/28	9191	1131	27.7	0.0	54/17	43335	6665
1.4447461483651718	10	1.8	0.0	54/18	8881	1131	22.9	0.0	57/17	30776	6665
2.132389747827441	10	1.6	0.0	115/30	8225	1131	21.5	0.0	107/28	28419	6665
2.8200333472897103	10	1.7	0.0	69/16	7174	1131	22.6	0.0	78/20	25492	6665
3.5076769467519795	10	1.5	0.0	116/31	8499	1131	19.4	0.0	71/16	20603	6665
4.1953205462142495	10	1.7	0.0	112/28	10261	1131	17.0	0.0	57/17	16353	6665
4.882964145676518	10	2.2	0.0	99/19	8386	1131	4.1	0.0	53/17	5608	6665
5.570607745138788	10	1.9	0.0	112/28	9697	1131	18.7	0.0	37/6	31483	6665
6.2582513446010575	10	1.6	0.0	117/29	7009	1131	22.7	0.0	112/28	28118	6665
6.945894944063326	10	1.6	0.0	44/8	7330	1131	21.1	0.0	112/28	22273	6665
k	15	Univarié					Multivarié				
Sigma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.06945894944063327	15	1.0	0.0	112/28	1193	1666	11.3	0.0	112/28	7705	7200
0.7571025489029025	15	5.0	0.0	98/24	13628	1666	43.1	0.0	35/6	80385	7200
1.4447461483651718	15	4.7	0.0	64/21	13324	1666	79.9	0.0	91/23	144035	7200
2.132389747827441	15	4.5	0.0	117/29	11514	1666	63.9	0.0	110/27	111597	7200
2.8200333472897103	15	3.9	0.0	114/29	9357	1666	42.7	0.0	84/21	62823	7200
3.5076769467519795	15	4.6	0.0	112/28	10388	1666	50.7	0.0	75/20	93888	7200
4.1953205462142495	15	3.8	0.0	53/18	8255	1666	76.7	0.0	50/10	165390	7200
4.882964145676518	15	4.6	0.0	114/29	11153	1666	61.1	0.0	102/25	112104	7200
5.570607745138788	15	4.0	0.0	112/28	11453	1666	44.2	0.0	109/29	66071	7200
6.2582513446010575	15	3.4	0.0	112/29	8934	1666	79.5	0.0	112/28	173373	7200
6.945894944063326	15	4.1	0.0	98/25	9851	1666	90.6	0.0	107/27	188936	7200

TABLE 2.11 – Résultats pour les données Seeds avec Formulation F_U sans inegalités valides avec *Spectral Clustering*.

Dataset=mushroom		Train size=120		Test size=30		Feature count =19		Time limit =900		Inequalities =Deactivated					
k		1		Univarié						Multivarié					
Sigma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints				
0.13933952957362017	1	0.0	0.0	60/15	0	168	0.2	0.0	60/15	0	5237				
1.5188008723524598	1	0.0	0.0	60/15	0	168	0.2	0.0	60/15	0	5237				
2.8982622151312993	1	0.0	0.0	60/15	0	168	0.2	0.0	60/15	0	5237				
4.277723557910139	1	0.0	0.0	60/15	0	168	0.2	0.0	60/15	0	5237				
5.6571849006889785	1	0.0	0.0	60/15	0	168	0.2	0.0	60/15	0	5237				
7.036646243467818	1	0.0	0.0	60/15	0	168	0.2	0.0	60/15	0	5237				
8.416107586246659	1	0.0	0.0	60/15	0	168	0.2	0.0	60/15	0	5237				
9.795568929025498	1	0.0	0.0	60/15	0	168	0.2	0.0	60/15	0	5237				
11.175030271804337	1	0.0	0.0	60/15	0	168	0.2	0.0	60/15	0	5237				
12.554491614583176	1	0.0	0.0	60/15	0	168	0.2	0.0	60/15	0	5237				
13.933952957362015	1	0.0	0.0	60/15	0	168	0.2	0.0	60/15	0	5237				
k		5		Univarié						Multivarié					
Sigma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints				
0.13933952957362017	5	0.3	0.0	60/15	2821	596	2.1	0.0	60/15	1575	5665				
1.5188008723524598	5	0.2	0.0	60/15	1324	596	1.0	0.0	47/11	510	5665				
2.8982622151312993	5	0.1	0.0	60/15	0	596	2.7	0.0	56/13	1852	5665				
4.277723557910139	5	0.3	0.0	60/15	1333	596	1.9	0.0	60/15	1575	5665				
5.6571849006889785	5	0.2	0.0	56/13	439	596	4.4	0.0	81/19	7903	5665				
7.036646243467818	5	0.1	0.0	60/15	0	596	1.4	0.0	56/14	1100	5665				
8.416107586246659	5	0.1	0.0	60/15	0	596	0.7	0.0	42/12	0	5665				
9.795568929025498	5	0.3	0.0	60/15	408	596	2.5	0.0	60/15	3150	5665				
11.175030271804337	5	0.1	0.0	60/15	0	596	2.5	0.0	52/15	2725	5665				
12.554491614583176	5	0.1	0.0	60/15	0	596	3.3	0.0	60/15	4253	5665				
13.933952957362015	5	0.2	0.0	60/15	2219	596	1.8	0.0	40/10	1335	5665				
k		10		Univarié						Multivarié					
Sigma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints				
0.13933952957362017	10	1.8	0.0	60/15	7420	1131	12.7	0.0	61/13	16627	6200				
1.5188008723524598	10	1.0	0.0	60/15	5730	1131	1.1	0.0	29/10	0	6200				
2.8982622151312993	10	1.0	0.0	60/15	4833	1131	22.1	0.0	59/14	35326	6200				
4.277723557910139	10	1.3	0.0	60/15	6745	1131	11.0	0.0	38/8	13888	6200				
5.6571849006889785	10	1.3	0.0	60/15	7258	1131	13.8	0.0	25/7	11220	6200				
7.036646243467818	10	1.2	0.0	60/15	3725	1131	15.9	0.0	49/16	19711	6200				
8.416107586246659	10	1.4	0.0	60/15	7714	1131	18.8	0.0	31/12	25842	6200				
9.795568929025498	10	1.7	0.0	60/15	6545	1131	14.8	0.0	55/16	25095	6200				
11.175030271804337	10	1.7	0.0	60/15	8991	1131	18.4	0.0	60/15	42137	6200				
12.554491614583176	10	1.1	0.0	60/15	5215	1131	25.6	0.0	73/20	63963	6200				
13.933952957362015	10	1.0	0.0	60/15	4580	1131	11.8	0.0	54/15	12362	6200				
k		15		Univarié						Multivarié					
Sigma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints				
0.13933952957362017	15	2.3	0.0	60/15	4696	1666	9.5	0.0	60/15	11635	6735				
1.5188008723524598	15	2.8	0.0	36/10	9039	1666	134.8	0.0	55/10	435867	6735				
2.8982622151312993	15	3.2	0.0	60/15	8400	1666	305.8	0.0	39/13	785830	6735				
4.277723557910139	15	4.2	0.0	60/15	13294	1666	139.5	0.0	44/10	114617	6735				
5.6571849006889785	15	3.3	0.0	60/15	12223	1666	184.6	0.0	76/14	567670	6735				
7.036646243467818	15	3.7	0.0	25/3	14610	1666	39.4	0.0	76/19	60700	6735				
8.416107586246659	15	4.2	0.0	60/15	12806	1666	157.9	0.0	49/11	557984	6735				
9.795568929025498	15	2.8	0.0	60/15	10510	1666	39.8	0.0	58/17	50030	6735				
11.175030271804337	15	4.4	0.0	60/15	12717	1666	97.3	0.0	68/18	313818	6735				
12.554491614583176	15	4.0	0.0	60/15	13342	1666	196.8	0.0	45/10	868748	6735				
13.933952957362015	15	4.0	0.0	60/15	12715	1666	19.2	0.0	44/9	13301	6735				

TABLE 2.12 – Résultats pour les données Mushroom avec Formulation F_U sans inegalités valides avec *Spectral Clustering*.

La performance de l'algorithme spectral clustering est très sensible au choix de σ . Une recherche plus fine l'intervalle de variation de σ peut nous permettre d'améliorer les résultats obtenus.

Dataset=iris		Train size=120		Test size=30		Feature count =4		Time limit =900		Inequalities =Deactivated					
k		2		Univarié						Multivarié					
Sigma		Nb Clusters		Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints		
0.06499666132278292		2		0.0	0.0	76/24	0	275	0.5	0.0	76/24	0	4219		
0.7084636084183339		2		0.0	0.0	76/24	0	275	0.6	0.0	76/24	0	4219		
1.351930555513885		2		0.0	0.0	83/26	0	275	0.5	0.0	76/24	0	4219		
1.995397502609436		2		0.0	0.0	38/12	0	275	0.4	0.0	76/24	0	4219		
2.638864449704987		2		0.0	0.0	82/18	0	275	0.4	0.0	76/24	0	4219		
3.282331396800538		2		0.0	0.0	83/26	0	275	0.5	0.0	76/24	0	4219		
3.925798343896089		2		0.0	0.0	82/18	0	275	0.4	0.0	76/24	0	4219		
4.56926529099164		2		0.0	0.0	76/24	0	275	0.1	0.0	38/12	0	4219		
5.212732238087191		2		0.0	0.0	83/26	0	275	0.2	0.0	38/12	0	4219		
5.856199185182741		2		0.0	0.0	82/18	0	275	0.4	0.0	76/24	0	4219		
6.499666132278293		2		0.0	0.0	82/18	0	275	0.1	0.0	38/12	0	4219		
k		5		Univarié						Multivarié					
Sigma		Nb Clusters		Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints		
0.06499666132278292		5		0.1	0.0	38/12	0	596	3.7	0.0	76/24	4515	4540		
0.7084636084183339		5		0.1	0.0	82/18	0	596	2.4	0.0	76/24	1180	4540		
1.351930555513885		5		1.3	0.0	38/12	298	596	3.3	0.0	82/18	3783	4540		
1.995397502609436		5		0.2	0.0	82/18	88	596	0.3	0.0	39/13	0	4540		
2.638864449704987		5		0.1	0.0	38/12	0	596	3.0	0.0	38/12	3830	4540		
3.282331396800538		5		0.2	0.0	38/12	358	596	4.1	0.0	76/24	4816	4540		
3.925798343896089		5		0.4	0.0	76/24	2375	596	0.4	0.0	39/12	0	4540		
4.56926529099164		5		0.1	0.0	82/18	0	596	1.9	0.0	38/12	1953	4540		
5.212732238087191		5		0.1	0.0	76/24	0	596	2.1	0.0	76/24	3148	4540		
5.856199185182741		5		0.1	0.0	38/12	0	596	2.3	0.0	82/18	2454	4540		
6.499666132278293		5		0.1	0.0	64/21	0	596	0.2	0.0	38/12	0	4540		
k		10		Univarié						Multivarié					
Sigma		Nb Clusters		Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints		
0.06499666132278292		10		0.4	0.0	82/18	0	1131	5.1	0.0	82/18	6745	5075		
0.7084636084183339		10		2.2	0.0	76/24	5648	1131	16.6	0.0	44/12	17256	5075		
1.351930555513885		10		1.7	0.0	76/24	5787	1131	24.6	0.0	76/24	20696	5075		
1.995397502609436		10		0.5	0.0	76/24	1883	1131	7.1	0.0	76/24	5861	5075		
2.638864449704987		10		1.3	0.0	76/24	6205	1131	12.6	0.0	76/24	14763	5075		
3.282331396800538		10		0.9	0.0	76/24	3609	1131	22.0	0.0	76/24	14980	5075		
3.925798343896089		10		1.6	0.0	76/24	5165	1131	5.3	0.0	76/24	5308	5075		
4.56926529099164		10		1.4	0.0	6/2	5431	1131	16.4	0.0	41/12	9643	5075		
5.212732238087191		10		0.9	0.0	75/24	4548	1131	19.3	0.0	76/24	11277	5075		
5.856199185182741		10		1.6	0.0	76/24	5095	1131	24.5	0.0	76/24	19404	5075		
6.499666132278293		10		1.7	0.0	76/24	5330	1131	3.5	0.0	76/24	4057	5075		
k		15		Univarié						Multivarié					
Sigma		Nb Clusters		Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints		
0.06499666132278292		15		1.0	0.0	82/18	1910	1666	18.2	0.0	82/18	11506	5610		
0.7084636084183339		15		4.6	0.0	76/24	5347	1666	18.0	0.0	82/18	11506	5610		
1.351930555513885		15		4.2	0.0	76/24	6898	1666	32.3	0.0	52/19	17972	5610		
1.995397502609436		15		0.8	0.0	82/18	1910	1666	123.0	0.0	35/16	115285	5610		
2.638864449704987		15		4.8	0.0	76/24	5884	1666	71.1	0.0	45/15	61155	5610		
3.282331396800538		15		5.4	0.0	76/24	8145	1666	42.0	0.0	60/20	22782	5610		
3.925798343896089		15		5.3	0.0	76/24	6298	1666	83.6	0.0	13/6	57188	5610		
4.56926529099164		15		4.7	0.0	17/5	5828	1666	18.3	0.0	82/18	11506	5610		
5.212732238087191		15		3.6	0.0	76/24	4891	1666	74.9	0.0	38/12	47966	5610		
5.856199185182741		15		0.9	0.0	82/18	1910	1666	48.9	0.0	17/4	23956	5610		
6.499666132278293		15		4.5	0.0	76/24	5681	1666	65.7	0.0	42/12	28691	5610		

TABLE 2.13 – Résultats pour les données Iris avec Formulation F_U sans inegalités valides avec *Spectral Clustering*.

Dataset=wine		Train size=142		Test size=36		Feature count =13		Time limit =900		Inequalities =Deactivated	
k 2		Univarié					Multivarié				
Sigma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.10160708762402278	2	0.0	0.0	102/28	0	275	1.1	0.0	102/28	0	5554
1.1075172551018482	2	0.0	0.0	96/23	0	275	0.9	0.0	102/28	0	5554
2.1134274225796736	2	0.0	0.0	96/23	0	275	1.4	0.0	102/28	0	5554
3.119337590057499	2	0.0	0.0	96/23	0	275	1.5	0.0	102/28	0	5554
4.125247757535325	2	0.0	0.0	96/23	0	275	1.2	0.0	102/28	0	5554
5.1311579250131505	2	0.0	0.0	96/23	0	275	1.7	0.0	102/28	0	5554
6.137068092490976	2	0.0	0.0	96/23	0	275	1.4	0.0	102/28	0	5554
7.142978259968801	2	0.0	0.0	96/23	0	275	1.4	0.0	102/28	0	5554
8.148888427446627	2	0.0	0.0	96/23	0	275	1.6	0.0	102/28	0	5554
9.154798594924452	2	0.0	0.0	102/30	0	275	1.4	0.0	102/28	0	5554
10.160708762402278	2	0.0	0.0	96/23	0	275	0.4	0.0	57/15	0	5554
k 5		Univarié					Multivarié				
Sigma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.10160708762402278	5	0.3	0.0	102/28	1222	596	10.6	0.0	96/23	1642	5875
1.1075172551018482	5	0.1	0.0	102/28	0	596	8.5	0.0	76/20	1250	5875
2.1134274225796736	5	0.2	0.0	56/15	23	596	11.0	0.0	102/28	2402	5875
3.119337590057499	5	0.2	0.0	102/28	414	596	9.6	0.0	102/28	2542	5875
4.125247757535325	5	0.1	0.0	102/28	0	596	21.2	0.0	96/23	6848	5875
5.1311579250131505	5	0.2	0.0	102/28	1502	596	2.2	0.0	63/16	0	5875
6.137068092490976	5	0.3	0.0	102/28	1505	596	29.5	0.0	102/28	6442	5875
7.142978259968801	5	0.4	0.0	95/23	2174	596	24.0	0.0	102/28	8222	5875
8.148888427446627	5	0.3	0.0	102/28	133	596	18.7	0.0	71/20	4069	5875
9.154798594924452	5	0.2	0.0	58/15	0	596	7.9	0.0	102/28	823	5875
10.160708762402278	5	0.2	0.0	96/23	0	596	18.3	0.0	102/28	1860	5875
k 10		Univarié					Multivarié				
Sigma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.10160708762402278	10	2.7	0.0	102/28	4569	1131	27.7	0.0	102/28	7204	6410
1.1075172551018482	10	6.0	0.0	98/23	8619	1131	332.5	0.0	115/32	152227	6410
2.1134274225796736	10	6.3	0.0	86/21	9080	1131	105.4	0.0	61/18	15051	6410
3.119337590057499	10	6.9	0.0	84/21	10246	1131	89.0	0.0	56/15	21713	6410
4.125247757535325	10	5.7	0.0	96/23	8720	1131	267.9	0.0	88/24	90503	6410
5.1311579250131505	10	5.2	0.0	56/15	6598	1131	110.8	0.0	51/18	12174	6410
6.137068092490976	10	5.5	0.0	88/21	8429	1131	176.0	0.0	72/19	39971	6410
7.142978259968801	10	5.3	0.0	84/21	8067	1131	167.9	0.0	82/19	46251	6410
8.148888427446627	10	6.8	0.0	102/28	9564	1131	130.2	0.0	83/23	48491	6410
9.154798594924452	10	5.8	0.0	64/16	10461	1131	113.0	0.0	41/9	20345	6410
10.160708762402278	10	7.3	0.0	86/21	9449	1131	159.4	0.0	98/25	50841	6410
k 15		Univarié					Multivarié				
Sigma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.10160708762402278	15	6.9	0.0	102/28	7085	1666	61.1	0.0	102/28	11949	6945
1.1075172551018482	15	13.3	0.0	102/28	19584	1666	177.2	0.0	51/18	41061	6945
2.1134274225796736	15	16.3	0.0	67/18	23630	1666	519.6	0.0	84/23	183111	6945
3.119337590057499	15	10.8	0.0	57/13	13386	1666	253.8	0.0	31/7	62396	6945
4.125247757535325	15	19.7	0.0	56/15	24162	1666	214.9	0.0	50/13	55188	6945
5.1311579250131505	15	10.3	0.0	80/22	13769	1666	965.9	46.4	76/20	84159	6945
6.137068092490976	15	12.6	0.0	55/14	17660	1666	769.9	0.0	97/27	369094	6945
7.142978259968801	15	16.4	0.0	96/23	18214	1666	647.1	0.0	91/24	264602	6945
8.148888427446627	15	14.7	0.0	80/21	19388	1666	401.5	0.0	70/16	150268	6945
9.154798594924452	15	10.8	0.0	56/15	11996	1666	900.3	8.7	99/27	434685	6945
10.160708762402278	15	15.1	0.0	86/21	17474	1666	829.9	0.0	51/10	313848	6945

TABLE 2.14 – Résultats pour les données Wine avec Formulation F_U sans inegalités valides avec *Spectral Clustering*.

2.2.2 Formulation F_U avec *Spectral Clustering* avec inegalités valides

Dataset=iris		Train size=120		Test size=30		Feature count =4		Time limit =900		Inequalities =Activated	
k	1	Univarié					Multivarié				
Sigma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.06499666132278292	1	0.0	0.0	82/18	0	213	0.1	0.0	82/18	0	4157
0.7084636084183339	1	0.0	0.0	82/18	0	213	0.1	0.0	82/18	0	4157
1.351930555513885	1	0.0	0.0	82/18	0	213	0.1	0.0	82/18	0	4157
1.995397502609436	1	0.0	0.0	82/18	0	213	0.1	0.0	82/18	0	4157
2.638864449704987	1	0.0	0.0	82/18	0	213	0.1	0.0	82/18	0	4157
3.282331396800538	1	0.0	0.0	82/18	0	213	0.1	0.0	82/18	0	4157
3.925798343896089	1	0.0	0.0	82/18	0	213	0.1	0.0	82/18	0	4157
4.56926529099164	1	0.0	0.0	82/18	0	213	0.1	0.0	82/18	0	4157
5.212732238087191	1	0.0	0.0	82/18	0	213	0.1	0.0	82/18	0	4157
5.856199185182741	1	0.0	0.0	82/18	0	213	0.1	0.0	82/18	0	4157
6.499666132278293	1	0.0	0.0	82/18	0	213	0.1	0.0	82/18	0	4157
k	5	Univarié					Multivarié				
Sigma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.06499666132278292	5	0.1	0.0	38/12	0	761	0.7	0.0	76/24	731	4705
0.7084636084183339	5	0.0	0.0	82/18	0	761	0.5	0.0	76/24	0	4705
1.351930555513885	5	0.1	0.0	38/12	0	761	0.6	0.0	82/18	0	4705
1.995397502609436	5	0.1	0.0	82/18	0	761	0.4	0.0	38/12	0	4705
2.638864449704987	5	0.0	0.0	38/12	0	761	0.8	0.0	38/12	947	4705
3.282331396800538	5	0.1	0.0	38/12	112	761	0.8	0.0	76/24	740	4705
3.925798343896089	5	0.2	0.0	76/24	1380	761	0.5	0.0	46/12	0	4705
4.56926529099164	5	0.1	0.0	80/18	0	761	0.4	0.0	38/12	0	4705
5.212732238087191	5	0.0	0.0	76/24	0	761	0.7	0.0	76/24	452	4705
5.856199185182741	5	0.1	0.0	38/12	0	761	0.7	0.0	82/18	512	4705
6.499666132278293	5	0.1	0.0	65/22	0	761	0.3	0.0	38/12	0	4705
k	10	Univarié					Multivarié				
Sigma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.06499666132278292	10	0.3	0.0	82/18	0	1446	1.5	0.0	82/18	2778	5390
0.7084636084183339	10	0.9	0.0	76/24	6303	1446	1.4	0.0	76/24	2837	5390
1.351930555513885	10	0.4	0.0	76/24	0	1446	1.6	0.0	76/24	3308	5390
1.995397502609436	10	0.4	0.0	79/26	0	1446	1.0	0.0	76/24	639	5390
2.638864449704987	10	0.8	0.0	76/24	2894	1446	1.2	0.0	76/24	1439	5390
3.282331396800538	10	0.6	0.0	77/24	1530	1446	1.1	0.0	76/24	1565	5390
3.925798343896089	10	0.6	0.0	76/24	1952	1446	1.1	0.0	76/24	1324	5390
4.56926529099164	10	0.9	0.0	7/2	3235	1446	1.6	0.0	76/24	2199	5390
5.212732238087191	10	0.6	0.0	75/24	1170	1446	1.2	0.0	76/24	1507	5390
5.856199185182741	10	0.5	0.0	76/23	1328	1446	1.4	0.0	76/24	2771	5390
6.499666132278293	10	0.5	0.0	76/24	475	1446	0.9	0.0	76/24	699	5390
k	15	Univarié					Multivarié				
Sigma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.06499666132278292	15	0.5	0.0	82/19	0	2131	2.3	0.0	82/18	6429	6075
0.7084636084183339	15	2.4	0.0	76/24	3860	2131	2.3	0.0	82/18	6429	6075
1.351930555513885	15	1.9	0.0	76/23	5396	2131	3.0	0.0	52/19	8394	6075
1.995397502609436	15	0.5	0.0	82/19	0	2131	2.8	0.0	76/24	7609	6075
2.638864449704987	15	1.8	0.0	76/24	7148	2131	5.6	0.0	52/15	6960	6075
3.282331396800538	15	2.2	0.0	76/24	4348	2131	2.4	0.0	76/24	6440	6075
3.925798343896089	15	1.7	0.0	76/24	5978	2131	8.9	0.0	8/3	8060	6075
4.56926529099164	15	2.4	0.0	21/6	6239	2131	2.2	0.0	82/18	6429	6075
5.212732238087191	15	2.2	0.0	77/24	5495	2131	7.1	0.0	62/21	7794	6075
5.856199185182741	15	0.5	0.0	82/19	0	2131	4.8	0.0	14/1	7180	6075
6.499666132278293	15	2.7	0.0	77/24	5112	2131	5.2	0.0	42/12	5288	6075

TABLE 2.15 – Résultats pour les données Iris avec Formulation F_U avec inegalités valides et *Spectral Clustering*.

Dataset=seeds		Train size=168		Test size=42		Feature count =6		Time limit =900		Inequalities =Activated	
k	1	Univarié					Multivarié				
Sigma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.06945894944063327	1	0.0	0.0	112/28	0	213	0.4	0.0	112/28	0	5747
0.7571025489029025	1	0.0	0.0	112/28	0	213	0.4	0.0	112/28	0	5747
1.4447461483651718	1	0.0	0.0	112/28	0	213	0.4	0.0	112/28	0	5747
2.132389747827441	1	0.0	0.0	112/28	0	213	0.4	0.0	112/28	0	5747
2.8200333472897103	1	0.0	0.0	112/28	0	213	0.4	0.0	112/28	0	5747
3.5076769467519795	1	0.0	0.0	112/28	0	213	0.4	0.0	112/28	0	5747
4.1953205462142495	1	0.0	0.0	112/28	0	213	0.4	0.0	112/28	0	5747
4.882964145676518	1	0.0	0.0	112/28	0	213	0.4	0.0	112/28	0	5747
5.570607745138788	1	0.0	0.0	112/28	0	213	0.4	0.0	112/28	0	5747
6.2582513446010575	1	0.0	0.0	112/28	0	213	0.4	0.0	112/28	0	5747
6.945894944063326	1	0.0	0.0	112/28	0	213	0.4	0.0	112/28	0	5747
k	5	Univarié					Multivarié				
Sigma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.06945894944063327	5	0.0	0.0	112/28	0	761	1.0	0.0	112/28	0	6295
0.7571025489029025	5	0.1	0.0	112/28	0	761	1.1	0.0	79/14	0	6295
1.4447461483651718	5	0.2	0.0	116/29	636	761	0.8	0.0	65/15	0	6295
2.132389747827441	5	0.2	0.0	112/28	760	761	0.9	0.0	112/28	0	6295
2.8200333472897103	5	0.2	0.0	116/31	452	761	1.0	0.0	115/25	856	6295
3.5076769467519795	5	0.2	0.0	111/28	548	761	0.7	0.0	65/15	0	6295
4.1953205462142495	5	0.3	0.0	110/28	992	761	1.1	0.0	65/15	0	6295
4.882964145676518	5	0.2	0.0	112/28	467	761	0.8	0.0	112/28	0	6295
5.570607745138788	5	0.1	0.0	116/31	549	761	1.5	0.0	112/28	0	6295
6.2582513446010575	5	0.0	0.0	112/28	0	761	0.8	0.0	112/28	0	6295
6.945894944063326	5	0.2	0.0	113/29	337	761	0.6	0.0	65/15	0	6295
k	10	Univarié					Multivarié				
Sigma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.06945894944063327	10	0.6	0.0	112/27	366	1446	2.5	0.0	56/17	3120	6980
0.7571025489029025	10	1.7	0.0	111/27	6052	1446	5.8	0.0	55/17	10448	6980
1.4447461483651718	10	1.8	0.0	53/18	6681	1446	3.1	0.0	57/17	4305	6980
2.132389747827441	10	1.8	0.0	111/28	7518	1446	3.5	0.0	112/28	6145	6980
2.8200333472897103	10	1.4	0.0	69/17	6800	1446	3.4	0.0	77/18	6232	6980
3.5076769467519795	10	1.7	0.0	115/30	7846	1446	3.6	0.0	53/17	6564	6980
4.1953205462142495	10	2.1	0.0	117/29	7493	1446	4.0	0.0	112/28	8562	6980
4.882964145676518	10	1.4	0.0	91/16	5631	1446	1.9	0.0	53/17	1005	6980
5.570607745138788	10	1.9	0.0	112/29	8940	1446	3.9	0.0	81/14	8238	6980
6.2582513446010575	10	2.2	0.0	116/31	11995	1446	2.8	0.0	112/28	3971	6980
6.945894944063326	10	1.2	0.0	44/8	8052	1446	3.0	0.0	112/28	3675	6980
k	15	Univarié					Multivarié				
Sigma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.06945894944063327	15	0.8	0.0	112/28	0	2131	4.0	0.0	112/28	5491	7665
0.7571025489029025	15	3.7	0.0	97/24	6728	2131	10.2	0.0	16/5	5566	7665
1.4447461483651718	15	4.1	0.0	64/19	7244	2131	13.0	0.0	70/17	5565	7665
2.132389747827441	15	4.4	0.0	112/28	7809	2131	14.1	0.0	53/17	5524	7665
2.8200333472897103	15	3.5	0.0	114/30	5281	2131	4.5	0.0	84/21	7450	7665
3.5076769467519795	15	3.9	0.0	114/29	5837	2131	11.0	0.0	76/20	5482	7665
4.1953205462142495	15	2.7	0.0	53/17	6705	2131	17.0	0.0	109/31	13503	7665
4.882964145676518	15	3.4	0.0	110/28	8372	2131	15.2	0.0	112/28	12024	7665
5.570607745138788	15	3.2	0.0	111/27	6530	2131	8.2	0.0	112/28	9802	7665
6.2582513446010575	15	3.4	0.0	112/28	6457	2131	11.8	0.0	112/28	25372	7665
6.945894944063326	15	3.2	0.0	97/25	6194	2131	15.3	0.0	112/28	5415	7665

TABLE 2.16 – Résultats pour les données Seeds avec Formulation F_U avec inegalités valides et *Spectral Clustering*.

Dataset=wine		Train size=142		Test size=36		Feature count =13		Time limit =900		Inequalities =Activated	
k	1	Univarié					Multivarié				
Sigma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.10160708762402278	1	0.0	0.0	102/29	0	213	0.6	0.0	102/28	0	5492
1.1075172551018482	1	0.0	0.0	102/29	0	213	0.6	0.0	102/28	0	5492
2.1134274225796736	1	0.0	0.0	102/29	0	213	0.6	0.0	102/28	0	5492
3.119337590057499	1	0.0	0.0	102/29	0	213	0.6	0.0	102/28	0	5492
4.125247757535325	1	0.0	0.0	102/29	0	213	0.6	0.0	102/28	0	5492
5.1311579250131505	1	0.0	0.0	102/29	0	213	0.6	0.0	102/28	0	5492
6.137068092490976	1	0.0	0.0	102/29	0	213	0.6	0.0	102/28	0	5492
7.142978259968801	1	0.0	0.0	102/29	0	213	0.6	0.0	102/28	0	5492
8.148888427446627	1	0.0	0.0	102/29	0	213	0.6	0.0	102/28	0	5492
9.154798594924452	1	0.0	0.0	102/29	0	213	0.6	0.0	102/28	0	5492
10.160708762402278	1	0.0	0.0	102/29	0	213	0.7	0.0	102/28	0	5492
k	5	Univarié					Multivarié				
Sigma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.10160708762402278	5	0.2	0.0	102/28	0	761	1.3	0.0	96/23	0	6040
1.1075172551018482	5	0.0	0.0	102/28	0	761	2.6	0.0	96/23	1002	6040
2.1134274225796736	5	0.1	0.0	56/15	0	761	1.4	0.0	102/28	0	6040
3.119337590057499	5	0.2	0.0	102/28	0	761	1.0	0.0	102/28	0	6040
4.125247757535325	5	0.0	0.0	102/28	0	761	2.0	0.0	96/23	0	6040
5.1311579250131505	5	0.3	0.0	102/28	795	761	1.4	0.0	102/28	0	6040
6.137068092490976	5	0.3	0.0	99/28	528	761	1.4	0.0	102/28	0	6040
7.142978259968801	5	0.3	0.0	96/22	1091	761	1.2	0.0	102/28	0	6040
8.148888427446627	5	0.1	0.0	102/28	0	761	2.1	0.0	87/22	777	6040
9.154798594924452	5	0.2	0.0	58/16	0	761	1.3	0.0	102/28	0	6040
10.160708762402278	5	0.0	0.0	96/23	0	761	1.2	0.0	102/28	0	6040
k	10	Univarié					Multivarié				
Sigma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.10160708762402278	10	0.8	0.0	102/28	1363	1446	5.2	0.0	67/16	6948	6725
1.1075172551018482	10	2.0	0.0	97/23	8548	1446	5.4	0.0	79/19	6550	6725
2.1134274225796736	10	2.2	0.0	86/21	7139	1446	4.3	0.0	53/15	3352	6725
3.119337590057499	10	1.8	0.0	84/21	7192	1446	11.8	0.0	66/14	7372	6725
4.125247757535325	10	1.5	0.0	95/23	8288	1446	11.8	0.0	131/33	5678	6725
5.1311579250131505	10	2.1	0.0	56/15	8045	1446	6.1	0.0	54/13	3751	6725
6.137068092490976	10	1.7	0.0	87/21	5212	1446	5.6	0.0	65/19	8050	6725
7.142978259968801	10	1.9	0.0	87/21	7716	1446	5.8	0.0	80/19	7210	6725
8.148888427446627	10	1.9	0.0	100/28	7232	1446	15.2	0.0	111/28	9457	6725
9.154798594924452	10	2.3	0.0	64/16	8833	1446	5.6	0.0	57/14	8429	6725
10.160708762402278	10	1.7	0.0	88/21	5608	1446	13.4	0.0	87/20	22394	6725
k	15	Univarié					Multivarié				
Sigma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.10160708762402278	15	1.5	0.0	102/28	2382	2131	5.3	0.0	102/28	6358	7410
1.1075172551018482	15	5.3	0.0	99/28	8980	2131	19.6	0.0	63/18	5621	7410
2.1134274225796736	15	6.7	0.0	59/15	12704	2131	29.8	0.0	98/28	20274	7410
3.119337590057499	15	5.1	0.0	56/13	9837	2131	15.2	0.0	25/5	5485	7410
4.125247757535325	15	5.8	0.0	57/16	13417	2131	21.4	0.0	30/7	5564	7410
5.1311579250131505	15	3.8	0.0	83/22	7659	2131	30.4	0.0	76/18	12993	7410
6.137068092490976	15	6.4	0.0	55/14	13532	2131	25.4	0.0	93/21	15019	7410
7.142978259968801	15	5.2	0.0	96/23	8687	2131	13.7	0.0	91/23	12620	7410
8.148888427446627	15	5.3	0.0	79/21	9918	2131	19.5	0.0	60/16	12018	7410
9.154798594924452	15	4.9	0.0	54/16	7897	2131	30.8	0.0	68/14	20445	7410
10.160708762402278	15	4.8	0.0	87/21	10370	2131	37.2	0.0	63/13	31110	7410

TABLE 2.17 – Résultats pour les données Wine avec Formulation F_U avec inegalités valides et *Spectral Clustering*.

Dataset=dry_bean		Train size=117		Test size=30		Feature count =16		Time limit =900		Inequalities =Activated	
k	1	Univari�					Multivari�				
Sigma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.054502084910682334	1	0.0	0.0	99/27	0	273	0.7	0.0	99/27	0	5027
0.5940727255264374	1	0.0	0.0	99/27	0	273	0.7	0.0	99/27	0	5027
1.1336433661421925	1	0.0	0.0	99/27	0	273	0.7	0.0	99/27	0	5027
1.6732140067579477	1	0.0	0.0	99/27	0	273	0.7	0.0	99/27	0	5027
2.2127846473737027	1	0.0	0.0	99/27	0	273	0.7	0.0	99/27	0	5027
2.7523552879894577	1	0.0	0.0	99/27	0	273	0.7	0.0	99/27	0	5027
3.291925928605213	1	0.0	0.0	99/27	0	273	0.7	0.0	99/27	0	5027
3.831496569220968	1	0.0	0.0	99/27	0	273	0.7	0.0	99/27	0	5027
4.371067209836723	1	0.0	0.0	99/27	0	273	0.7	0.0	99/27	0	5027
4.910637850452479	1	0.0	0.0	99/27	0	273	0.7	0.0	99/27	0	5027
5.450208491068233	1	0.0	0.0	99/27	0	273	0.7	0.0	99/27	0	5027
k	5	Univari�					Multivari�				
Sigma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.054502084910682334	5	0.0	0.0	92/26	0	1061	1.5	0.0	68/19	0	5815
0.5940727255264374	5	0.3	0.0	86/24	1374	1061	2.8	0.0	96/24	1578	5815
1.1336433661421925	5	0.0	0.0	99/27	0	1061	3.5	0.0	104/24	1230	5815
1.6732140067579477	5	0.3	0.0	80/25	2993	1061	2.4	0.0	88/17	0	5815
2.2127846473737027	5	0.3	0.0	91/18	2567	1061	3.6	0.0	72/21	2123	5815
2.7523552879894577	5	0.3	0.0	84/21	1137	1061	2.5	0.0	96/23	0	5815
3.291925928605213	5	0.3	0.0	99/26	2018	1061	2.4	0.0	90/23	0	5815
3.831496569220968	5	0.3	0.0	65/19	1280	1061	4.7	0.0	76/16	3034	5815
4.371067209836723	5	0.3	0.0	85/22	2173	1061	3.4	0.0	80/20	2985	5815
4.910637850452479	5	0.0	0.0	99/27	0	1061	3.5	0.0	93/24	1598	5815
5.450208491068233	5	0.3	0.0	102/24	1898	1061	3.2	0.0	81/18	1471	5815
k	10	Univari�					Multivari�				
Sigma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.054502084910682334	10	1.3	0.0	97/28	8064	2046	24.3	0.0	75/18	11635	6800
0.5940727255264374	10	2.6	0.0	83/22	11679	2046	25.3	0.0	87/18	11963	6800
1.1336433661421925	10	2.3	0.0	75/18	10131	2046	249.0	0.0	107/27	15250	6800
1.6732140067579477	10	2.3	0.0	87/22	14185	2046	29.9	0.0	98/24	24777	6800
2.2127846473737027	10	1.2	0.0	91/27	7387	2046	14.7	0.0	86/22	7175	6800
2.7523552879894577	10	2.0	0.0	86/23	10358	2046	266.3	0.0	91/20	13391	6800
3.291925928605213	10	1.4	0.0	54/15	5797	2046	29.8	0.0	83/19	12177	6800
3.831496569220968	10	2.0	0.0	82/24	6834	2046	23.4	0.0	82/21	18274	6800
4.371067209836723	10	2.3	0.0	87/20	13274	2046	237.7	0.0	82/15	21734	6800
4.910637850452479	10	1.5	0.0	62/11	7383	2046	261.0	0.0	99/23	11431	6800
5.450208491068233	10	1.9	0.0	84/18	7411	2046	12.3	0.0	64/16	5631	6800
k	15	Univari�					Multivari�				
Sigma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints
0.054502084910682334	15	5.8	0.0	100/26	10608	3031	120.3	0.0	76/24	75233	7785
0.5940727255264374	15	9.3	0.0	65/15	48358	3031	96.3	0.0	73/20	125022	7785
1.1336433661421925	15	8.9	0.0	72/20	24674	3031	404.8	0.0	78/18	33089	7785
1.6732140067579477	15	7.2	0.0	81/17	21646	3031	228.5	0.0	100/27	50060	7785
2.2127846473737027	15	6.8	0.0	84/21	15741	3031	289.8	0.0	79/20	37630	7785
2.7523552879894577	15	8.3	0.0	66/18	26888	3031	84.8	0.0	90/21	78226	7785
3.291925928605213	15	5.6	0.0	89/27	18144	3031	417.9	0.0	86/23	65675	7785
3.831496569220968	15	7.6	0.0	58/15	15711	3031	105.4	0.0	74/18	84022	7785
4.371067209836723	15	10.2	0.0	80/25	35057	3031	326.3	0.0	71/16	53667	7785
4.910637850452479	15	7.3	0.0	71/21	25929	3031	86.2	0.0	88/25	64620	7785
5.450208491068233	15	8.6	0.0	80/25	19302	3031	338.5	0.0	82/23	112214	7785

TABLE 2.18 – Résultats pour les données Dry_Bean avec Formulation F_U avec inegalités valides et *Spectral Clustering*.

Dataset=mushroom		Train size=120		Test size=30		Feature count =19		Time limit =900		Inequalities =Activated		
k	1	Univarié						Multivarié				
Sigma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints	
0.142621136182802	1	0.0	0.0	59/16	0	198	0.3	0.0	59/16	0	5267	
1.554570384392542	1	0.0	0.0	59/16	0	198	0.3	0.0	59/16	0	5267	
2.966519632602282	1	0.0	0.0	59/16	0	198	0.3	0.0	59/16	0	5267	
4.378468880812022	1	0.1	0.0	59/16	0	198	0.3	0.0	59/16	0	5267	
5.790418129021762	1	0.0	0.0	59/16	0	198	0.3	0.0	59/16	0	5267	
7.202367377231502	1	0.0	0.0	59/16	0	198	0.3	0.0	59/16	0	5267	
8.614316625441242	1	0.0	0.0	59/16	0	198	0.3	0.0	59/16	0	5267	
10.026265873650981	1	0.0	0.0	59/16	0	198	0.3	0.0	59/16	0	5267	
11.438215121860722	1	0.0	0.0	59/16	0	198	0.3	0.0	59/16	0	5267	
12.850164370070463	1	0.0	0.0	59/16	0	198	0.3	0.0	59/16	0	5267	
14.262113618280202	1	0.0	0.0	59/16	0	198	0.4	0.0	59/16	0	5267	
k	5	Univarié						Multivarié				
Sigma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints	
0.142621136182802	5	0.2	0.0	15/0	626	686	1.4	0.0	59/16	1725	5755	
1.554570384392542	5	0.1	0.0	61/14	0	686	1.9	0.0	61/14	1307	5755	
2.966519632602282	5	0.2	0.0	61/14	1434	686	1.5	0.0	17/1	608	5755	
4.378468880812022	5	0.1	0.0	59/16	0	686	0.4	0.0	17/1	0	5755	
5.790418129021762	5	0.1	0.0	59/16	0	686	2.4	0.0	58/13	2939	5755	
7.202367377231502	5	0.1	0.0	30/3	0	686	2.2	0.0	59/16	1398	5755	
8.614316625441242	5	0.1	0.0	59/16	0	686	0.7	0.0	19/1	0	5755	
10.026265873650981	5	0.1	0.0	59/16	0	686	1.3	0.0	74/20	296	5755	
11.438215121860722	5	0.1	0.0	61/14	0	686	2.0	0.0	61/14	760	5755	
12.850164370070463	5	0.3	0.0	42/7	1177	686	2.5	0.0	69/18	1057	5755	
14.262113618280202	5	0.1	0.0	59/16	0	686	2.2	0.0	59/16	2339	5755	
k	10	Univarié						Multivarié				
Sigma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints	
0.142621136182802	10	1.4	0.0	59/16	5359	1296	2.2	0.0	59/16	4033	6365	
1.554570384392542	10	0.8	0.0	59/16	2601	1296	4.1	0.0	41/7	5920	6365	
2.966519632602282	10	1.1	0.0	59/16	8388	1296	3.9	0.0	57/14	4695	6365	
4.378468880812022	10	2.1	0.0	61/14	8958	1296	7.8	0.0	29/10	7939	6365	
5.790418129021762	10	0.9	0.0	61/14	1688	1296	10.9	0.0	55/12	18569	6365	
7.202367377231502	10	1.4	0.0	30/3	4651	1296	4.0	0.0	59/16	4412	6365	
8.614316625441242	10	0.8	0.0	59/16	776	1296	4.2	0.0	37/7	7495	6365	
10.026265873650981	10	1.4	0.0	59/16	8539	1296	7.4	0.0	30/9	9976	6365	
11.438215121860722	10	1.6	0.0	61/14	8003	1296	6.6	0.0	15/0	8990	6365	
12.850164370070463	10	1.4	0.0	14/0	5336	1296	8.2	0.0	51/14	14337	6365	
14.262113618280202	10	1.2	0.0	59/14	3710	1296	3.6	0.0	59/13	3261	6365	
k	15	Univarié						Multivarié				
Sigma	Nb Clusters	Solve time	Gap(%)	Train/test	Node count	Nb constraints	Solve time	Gap(%)	Train/test	Node count	nb constraints	
0.142621136182802	15	2.9	0.0	59/16	3503	1906	8.8	0.0	59/16	6455	6975	
1.554570384392542	15	5.0	0.0	23/4	11329	1906	20.6	0.0	25/9	25994	6975	
2.966519632602282	15	4.1	0.0	59/16	10901	1906	29.3	0.0	57/14	68420	6975	
4.378468880812022	15	4.8	0.0	61/14	11245	1906	12.1	0.0	39/7	15601	6975	
5.790418129021762	15	5.0	0.0	59/16	11180	1906	11.0	0.0	43/12	16298	6975	
7.202367377231502	15	3.8	0.0	59/16	8560	1906	11.0	0.0	61/14	16533	6975	
8.614316625441242	15	4.1	0.0	61/14	9525	1906	25.5	0.0	50/10	44611	6975	
10.026265873650981	15	3.2	0.0	59/16	8574	1906	24.0	0.0	59/16	41273	6975	
11.438215121860722	15	3.9	0.0	59/14	10169	1906	20.0	0.0	48/15	35553	6975	
12.850164370070463	15	5.1	0.0	30/3	11709	1906	14.7	0.0	40/6	15317	6975	
14.262113618280202	15	5.6	0.0	59/16	11679	1906	14.3	0.0	60/14	16517	6975	

TABLE 2.19 – Résultats pour les données Mushroom avec Formulation F_U avec inegalités valides et *Spectral Clustering*.

En ajoutant les inegalités valides avec le spectral Clustering on a diminué le temps d'excution(solve time) ,on a augmenté le nombres de contraintes et le nombre de noeuds explorés mais on n'a pas pu arriver a ameliorer les performances .

Conclusion

Il convient de noter que la méthode de classification F a montré une bonne efficacité dans la classification des données "iris", "wine", "seeds" et "mushroom". Cependant, les résultats pour les données "dry_bean" n'étaient pas satisfaisants. En ce qui concerne le regroupement simple merge, il a donné des performances satisfaisantes pour toutes les données, sauf pour "dry_bean", où une précision très faible a été obtenue. L'ajout des inégalités valides a entraîné une légère amélioration en termes de temps de calcul, de nombre de nœuds développés et de nombre de contraintes, mais cela n'a pas conduit à une amélioration satisfaisante des résultats. De même, l'ajout de la méthode Spectral Clustering n'a pas significativement amélioré les performances, même avec l'utilisation des inégalités valides, ce qui souligne la sensibilité de cette méthode au choix de sigma.