

Université Toulouse III - Paul Sabatier UT3

Calcul scientifique et apprentissage automatique



Projet Beatbox

Auteurs

EL AMRANI Wadie
RATABOUIL Guilhem

Professeurs

MOUYSET SANDRINE

Toulouse, 20 octobre 2023

Conteúdo

1	Introduction	2
2	Méthodes Supervisées	3
2.1	Classificateur à Vecteurs de Support (SVC)	3
2.1.1	Théorie	3
2.1.2	Évaluation	3
2.2	Classificateur à Vecteurs de Support (SVC) avec ACP	4
2.2.1	Analyse en Composantes Principales (ACP)	4
2.2.2	Approche en 2D	5
2.2.3	Frontière de décision en 2D	5
2.2.4	Approche en 3D	5
2.2.5	Comparaison des Performances	6
2.3	Conclusion	6
3	Méthodes Non Supervisées	7
3.1	Modèle k-means	7
3.2	Classification sans prétraitement	7
3.2.1	Prétraitement des Données	7
3.2.2	Entraînement du modèle	8
3.2.3	Résultats obtenus	8
3.3	Classification avec prétraitement	8
3.3.1	Prétraitement des Données	9
3.3.2	Entraînement du modèle	9
3.3.3	Résultats obtenus	9
4	Conclusion	10

1 Introduction

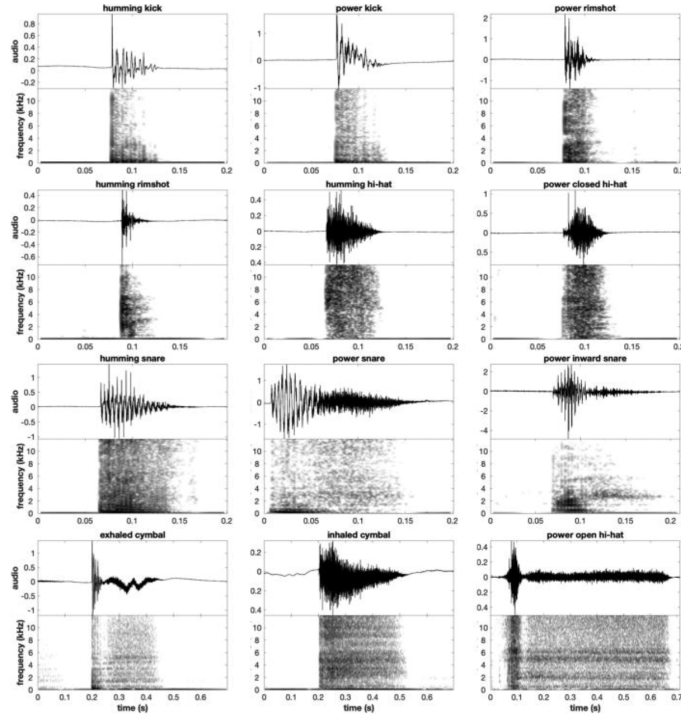


Figure 1: 12 sons beatbox

Dans l'ère moderne de la technologie et des données volumineuses, l'analyse et l'interprétation de grandes quantités de données sont devenues essentielles pour divers domaines, de la finance à la santé. Notre projet s'inscrit précisément dans ce contexte. Il vise à détecter et reconnaître les différents sons de beatbox à partir d'enregistrements audio au format .wav grâce à l'analyse cepstrale. Ces enregistrements proviennent d'une base de données contenant divers échantillons de beatbox, et le défi réside dans la reconnaissance de 12 sons beatbox distincts.

Nous utilisons à la fois des méthodes supervisées et non supervisées pour aborder ce problème. Les méthodes supervisées nécessitent une sortie connue pour chaque entrée, et le modèle est formé sur cet ensemble d'entraînements. Les méthodes les plus courantes dans cette catégorie comprennent la régression et la classification. Pour ce projet, nous avons choisi d'utiliser la machine à vecteurs de support (SVM) comme principal algorithme de classification.

D'autre part, les méthodes non supervisées ne reposent pas sur des sorties pré-étiquetées pour former leurs modèles. Au lieu de cela, elles cherchent à identifier des structures intrinsèques dans les données. Dans cette catégorie, nous avons utilisé l'algorithme de clustering K-Means pour segmenter notre ensemble de données en clusters distincts, révélant ainsi des patterns sous-jacents. De plus, pour aborder les défis posés par la haute dimensionnalité de nos données, nous avons utilisé l'analyse en composantes principales (ACP). Cette technique permet de réduire la dimension des données tout en conservant le maximum de variance, facilitant ainsi leur visualisation et leur analyse.

Ce rapport détaillera notre approche et les résultats obtenus en utilisant ces méthodes sur notre ensemble de données.

2 Méthodes Supervisées

Dans cette sous-section, nous décrirons en détail la méthode supervisée que nous avons choisie pour notre projet, y compris le prétraitement des données, les caractéristiques extraites, et le modèle utilisé. Nous présenterons également les résultats, y compris la matrice de confusion et le score de performance.

2.1 Classificateur à Vecteurs de Support (SVC)

2.1.1 Théorie

Le classificateur à vecteurs de support (SVC, Support Vector Classifier) est un algorithme de classification supervisée qui est particulièrement efficace pour résoudre des problèmes de classification binaire ou multiclasse. L'idée fondamentale de SVC est de trouver l'hyperplan qui sépare de manière optimale les classes dans l'espace des caractéristiques.

L'algorithme tente de maximiser la marge entre les différentes classes en résolvant un problème d'optimisation convexe. La marge est définie comme la distance entre les vecteurs de support les plus proches de chaque classe.

Algorithm 1 Algorithme SVM

- 1: **Entrée:**
- 2: \mathbf{X} : ensemble de données d'entraînement
- 3: \mathbf{y} : étiquettes de l'ensemble de données d'entraînement, $\mathbf{y}_i \in \{-1, 1\}$
- 4: C : paramètre de régularisation
- 5: **Initialisation:** Choisir un hyperplan initial (ou aléatoirement)
- 6: **repeat**
- 7: Sélectionner un sous-ensemble d'exemples d'entraînement, appelé vecteurs de support
- 8: Minimiser la fonction objectif:

$$\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_i \xi_i$$

- 9: sous les contraintes:

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0$$

- 10: Mettre à jour l'hyperplan de séparation.
 - 11: **until** convergence ou nombre d'itérations maximum atteint
 - 12: **Sortie:** Vecteur des poids \mathbf{w} et biais b qui définissent l'hyperplan optimal.
-

2.1.2 Évaluation

Le modèle SVC a été entraîné sur un ensemble de données d'entraînement représenté par X_{train} et y_{train} . Après l'entraînement, nous avons utilisé le modèle pour prédire les étiquettes de l'ensemble de données de test X_{test} . La performance du modèle a été évaluée en utilisant une matrice de confusion et un score d'exactitude.

Matrice de confusion : Un outil utilisé en apprentissage automatique et en statistiques pour comprendre la performance d'un algorithme de classification. Elle présente

un tableau qui confronte les prédictions d'un modèle par rapport aux véritables valeurs observées. Les lignes de la matrice représentent les classes réelles, tandis que les colonnes représentent les classes prédites. Les éléments diagonaux de la matrice montrent le nombre de prédictions correctes, tandis que les éléments hors diagonale indiquent les erreurs de classification.

Accuracy : Nous avons évalué le score de performance de notre modèle SVM. Pour cela, nous avons utilisé l'exactitude (ou *accuracy* en anglais) comme métrique, qui est calculée comme la proportion des prédictions correctes par rapport à l'ensemble total des prédictions.

2.2 Classificateur à Vecteurs de Support (SVC) avec ACP

2.2.1 Analyse en Composantes Principales (ACP)

L'Analyse en Composantes Principales (ACP) est une méthode statistique de réduction de dimension. L'idée centrale est de transformer un ensemble initial de variables qui peuvent être corrélées entre elles en un nouvel ensemble de variables, appelées *composantes principales*, qui sont orthogonales (non corrélées) et qui reflètent la majeure partie de la variabilité présente dans les données.

La première composante principale reflète la plus grande part de la variabilité des données, la deuxième composante principale (perpendiculaire à la première) capte la plus grande part de la variabilité restante et ainsi de suite.

L'utilisation de l'ACP dans le contexte de la classification, comme avec la méthode SVC (Support Vector Classifier), présente plusieurs avantages. Avant tout, elle permet de réduire le nombre de dimensions des données, ce qui peut potentiellement accélérer le processus d'apprentissage du modèle. De plus, en projetant les données dans un espace de dimension réduite, on peut visualiser les données plus facilement, ce qui aide à comprendre les structures sous-jacentes.

Pour notre projet, l'ACP a été réalisée sur l'ensemble d'apprentissage. Après avoir obtenu les composantes principales, nous avons projeté les données de l'ensemble de test dans cette nouvelle base afin d'assurer que les deux ensembles sont traités dans le même espace réduit.

Algorithm 2 Algorithme de l'Analyse en Composantes Principales (ACP)

- 1: **procedure** ACP(Données X)
 - 2: $n, d \leftarrow$ dimensions de X \triangleright n est le nombre d'observations et d est le nombre de variables
 - 3: $m \leftarrow$ moyenne de chaque colonne de X
 - 4: Centraliser les données : $Z \leftarrow X - m$
 - 5: $S \leftarrow \frac{1}{n-1} Z^T Z$ \triangleright S est la matrice de covariance
 - 6: Calculer les valeurs propres $\lambda_1, \lambda_2, \dots, \lambda_d$ et les vecteurs propres v_1, v_2, \dots, v_d de S
 - 7: Trier les paires (valeur propre, vecteur propre) par les valeurs propres décroissantes
 - 8: Sélectionner les k premiers vecteurs propres $V_k = [v_1, v_2, \dots, v_k]$ \triangleright k est la nouvelle dimension choisie
 - 9: Projeter Z sur les k premiers vecteurs propres : $T = ZV_k$ \triangleright T est le nouvel ensemble de données de dimension réduite
 - 10: **return** T, V_k
-

2.2.2 Approche en 2D

Lorsque les données sont projetées en 2D à l'aide de l'ACP, nous réduisons l'espace des caractéristiques à deux dimensions. Cela facilite la visualisation des données et la compréhension des relations potentielles entre les différentes classes.

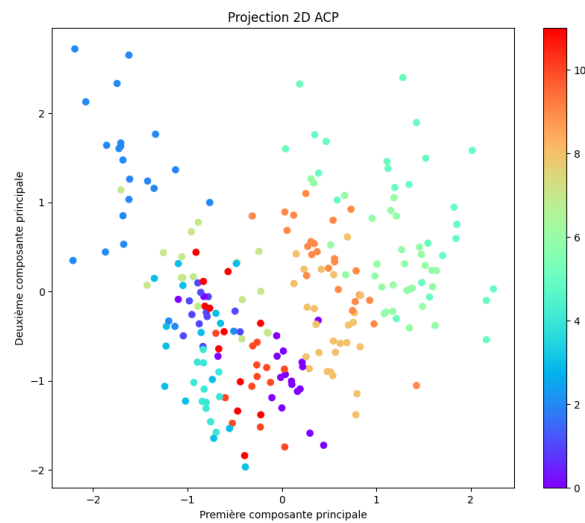


Figura 2: Visualisation des données après projection en 2D

2.2.3 Frontière de décision en 2D

La frontière de décision du SVM dans l'espace projeté en 2D nous donne une idée de comment le SVM classe les points dans cet espace réduit.

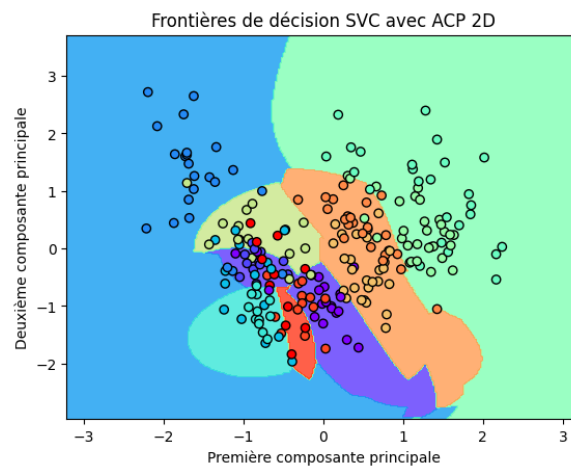


Figura 3: Frontière de décision du SVM dans l'espace projeté en 2D

2.2.4 Approche en 3D

Une projection en 3D capture davantage de variance des données que la projection en 2D. Cela pourrait donc potentiellement améliorer la performance du SVM, puisque davantage d'informations sont conservées.

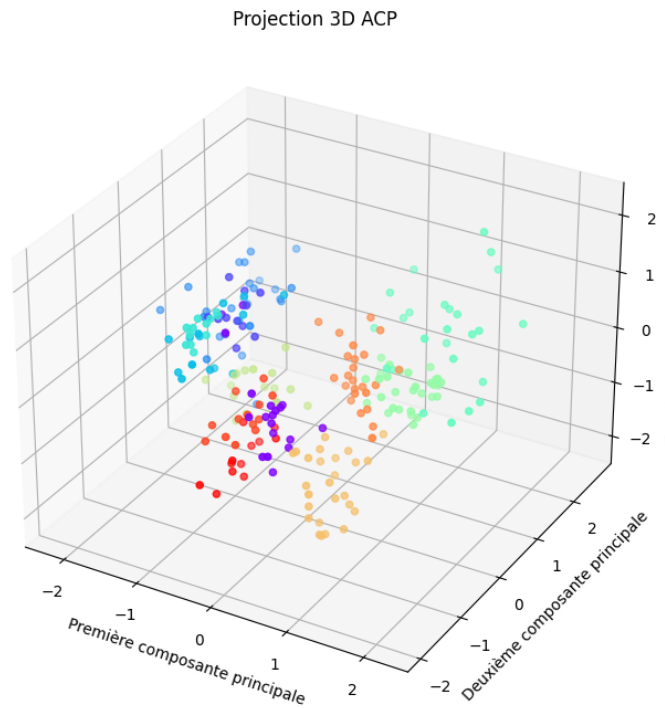


Figura 4: Visualisation des données après projection en 3D

2.2.5 Comparaison des Performances

Il est possible que le score obtenu avec une projection en 3D soit meilleur que celui en 2D car plus de variance et d'information sont capturées avec trois composantes principales qu'avec deux. Cela donne au SVM une meilleure capacité à différencier et classer les données correctement.

Score en 2D : 0.63

Score en 3D : 0.76

2.3 Conclusion

Le tableau ci-dessous résume les résultats de l'accuracy des trois modèles :

	SVM sans prétraitement	SVM avec prétraitement	
		ACP 2D	ACP 3D
Accuracy	0.84	0.63	0.76

3 Méthodes Non Supervisées

Ici, nous expliquerons la méthode non supervisée que nous avons utilisée pour la reconnaissance des sons Beatbox. Nous discuterons des détails de l'analyse non supervisée, des résultats obtenus, et de la manière dont ces résultats sont interprétés.

3.1 Modèle k-means

KMeans, l'un des algorithmes de clustering les plus répandus en apprentissage automatique non supervisé, se distingue par son objectif fondamental : regrouper un ensemble de données en plusieurs clusters distincts en fonction de leurs caractéristiques. Cette démarche vise à minimiser la variance au sein de chaque cluster, tout en maximisant la variance entre les clusters. En d'autres termes, KMeans cherche à créer des groupes d'éléments de telle manière que les éléments à l'intérieur d'un même groupe se ressemblent davantage les uns les autres, tout en étant distincts des éléments appartenant à d'autres groupes. Cette notion de séparation optimale des données en clusters cohérents en fait un outil puissant pour l'exploration de structures sous-jacentes dans des ensembles de données non étiquetés. KMeans repose sur un processus itératif qui réaffecte les échantillons aux clusters en fonction de la proximité de leurs caractéristiques, jusqu'à ce qu'une convergence soit atteinte et que les clusters finaux soient établis.

Algorithm 3 Algorithme KMeans

```
1: Entrée:
2:  $\mathbf{X}$  : ensemble de données à regrouper
3:  $K$  : nombre de clusters souhaité
4:  $\epsilon$  : tolérance pour la convergence
5: Initialisation: Choisir  $K$  centroïdes initiaux (peuvent être aléatoires ou basés sur
   des heuristiques)
6: repeat
7:   for chaque point  $\mathbf{x}$  dans  $\mathbf{X}$  do
8:     Affecter  $\mathbf{x}$  au cluster dont le centroïde est le plus proche
9:   for chaque cluster do
10:    Recalculer le centroïde du cluster comme la moyenne de ses points
11: until convergence (lorsque les centroïdes ne changent que de façon marginale) ou
   nombre d'itérations maximum atteint
12: Sortie: Les clusters formés et les centroïdes correspondants.
```

3.2 Classification sans prétraitement

La classification sans prétraitement signifie que nous utilisons directement les données brutes sans effectuer de transformations ou d'extraction de caractéristiques préalables. Dans ce contexte, nous travaillons avec les coefficients MFCC (Mel Frequency Cepstral Coefficient), qui sont des paramètres permettant d'extraire des caractéristiques fréquentielles à partir des enregistrements audio.

3.2.1 Prétraitement des Données

Contrairement à de nombreuses approches de classification, aucune transformation ou prétraitement des données n'est effectué dans cette méthode. Les données d'entrée

consistent en des vecteurs de 13 coefficients MFCC extraits directement des enregistrements audio. Cela signifie que le modèle k-means opère sur les données brutes sans modification.

3.2.2 Entraînement du modèle

Le modèle KMeans est entraîné sur les données d'entraînement. Lors de l'entraînement, l'algorithme répartit automatiquement les données dans les clusters en fonction de leurs caractéristiques MFCC. Il le fait sans avoir connaissance des étiquettes de classe réelles, car c'est une méthode non supervisée.

[16	3	0	0	0	0	0	0	0	0	0	0]
[0	12	0	0	0	0	0	0	0	0	0	0]
[0	0	19	1	0	0	0	1	0	0	2	0]
[1	0	1	13	0	0	0	0	0	0	0	0]
[0	0	0	12	0	0	0	0	0	0	0	0]
[0	0	0	0	0	18	2	0	0	5	0	0]
[0	0	0	0	0	0	30	0	0	0	0	0]
[0	2	0	0	0	0	0	15	0	0	0	0]
[0	0	0	0	0	0	0	0	26	0	0	0]
[0	0	0	0	0	3	2	2	0	17	0	0]
[0	0	0	0	0	0	0	0	0	0	13	0]
[0	0	0	0	0	0	0	0	0	0	0	12]]

Figura 5: Matrice de confusion

3.2.3 Résultats obtenus

Les résultats de la classification k-means sont évalués en utilisant deux métriques clés : la matrice de confusion et le score de performance.

La matrice de confusion est une représentation tabulaire qui montre comment les données réelles correspondent aux clusters prédits. Chaque entrée de la matrice indique combien d'échantillons appartiennent à une classe particulière, et elle permet d'évaluer la qualité de la classification.

Le score de performance quantifie la précision globale du modèle. Il mesure la proportion d'échantillons correctement classés parmi l'ensemble des échantillons. Dans ce cas, l'exactitude est de 0.84, ce qui signifie que le modèle a correctement classé 84/100 des données d'entraînement.

En résumé, la classification des sons de beatbox sans prétraitement en utilisant la méthode k-means a donné des résultats satisfaisants, avec une exactitude de 0.84. Cependant, il est important de noter que des prétraitements des données pourraient potentiellement améliorer la performance du modèle en extrayant des caractéristiques plus discriminantes à partir des coefficients MFCC.

3.3 Classification avec prétraitement

Dans cette section, nous explorons la classification des sons de beatbox en utilisant l'algorithme KMeans avec prétraitement des données. Le prétraitement consiste en une réduction de dimension par Analyse en Composantes Principales (ACP) sur l'ensemble des données. L'objectif est de regrouper les données en clusters en utilisant des caractéristiques extraites par l'ACP, puis d'évaluer la performance du modèle.

3.3.1 Prétraitement des Données

Pour améliorer la qualité de la classification, nous utilisons l'ACP pour réduire la dimension des données tout en préservant l'essentiel de l'information. L'ACP transforme l'ensemble des données en un espace de dimension réduite en sélectionnant les composantes principales qui expliquent la variance maximale des données, ce qui signifie que KMeans opère dans un espace de caractéristiques de plus faible dimension.

3.3.2 Entraînement du modèle

Le modèle KMeans est entraîné sur les données transformées par l'ACP. Il effectue le regroupement des données en clusters, sans avoir connaissance des étiquettes de classe réelles. Les centroïdes de chaque cluster sont calculés de manière itérative.

3.3.3 Résultats obtenus

Les résultats de la classification KMeans sont évalués en utilisant deux mesures clés, la matrice de confusion et le score de performance.

[22	4	0	0	0	0	0	0	0	0	2	0]
[0	18	0	0	0	0	0	0	0	0	0	0]
[0	0	36	0	1	0	0	0	0	0	0	0]
[0	0	1	0	18	0	0	1	0	0	0	0]
[0	0	0	0	24	0	0	0	0	0	0	0]
[0	0	0	0	0	18	19	0	0	0	0	0]
[0	0	0	0	0	0	42	0	0	0	0	0]
[0	3	3	0	0	0	0	16	0	0	2	1]
[7	0	0	0	0	0	0	0	31	0	0	0]
[5	0	0	0	0	1	22	0	0	0	2	0]
[2	0	0	0	0	0	0	0	0	0	20	0]
[0	0	0	0	0	0	0	2	0	0	0	18]]

Figura 6: Matrice de confusion

La matrice de confusion est un outil essentiel pour évaluer la performance du modèle. Elle révèle comment les données réelles correspondent aux clusters prédits. Chaque entrée de la matrice indique combien d'échantillons appartiennent à une classe particulière. Ainsi, elle met en évidence les vrais positifs, les vrais négatifs, les faux positifs et les faux négatifs, permettant ainsi d'analyser en détail la qualité de la classification.

Le score de performance est une mesure globale de la précision du modèle. Il mesure la proportion d'échantillons correctement classés parmi l'ensemble des échantillons. Dans ce cas, l'exactitude est de 0.72, ce qui signifie que le modèle a correctement classé 72/100 des données. Cela signifie que 72 échantillons sur 100 ont été attribués à la classe correcte.

En résumé, l'utilisation de l'ACP en prétraitement des données combinée à l'algorithme KMeans a abouti à une classification des sons de beatbox. Cependant, il est important de noter que l'exactitude obtenue (0.72) est légèrement inférieure à celle de la méthode non supervisée sans ACP (0.84). Cela suggère que l'ajout de l'ACP, bien qu'il ait réduit la dimension des données, n'a pas nécessairement amélioré la performance du modèle dans ce cas particulier. D'autres approches de prétraitement et de clustering pourraient être explorées pour rechercher des améliorations potentielles.

4 Conclusion

Les tableaux ci-dessous reprennent les resultats de l'accuracy de tous les modèles de ce projet :

	SVM sans prétraitement	SVM avec prétraitement	
		ACP 2D	ACP 3D
Accuracy	0.84	0.63	0.76

	KMeans sans prétraitement	KMeans avec prétraitement
Accuracy	0.84	0.72