

des classes minoritaires et de garantir un nombre minimum de 10 échantillons par classe, même pour la classe majoritaire initiale.

Différentes méthodes d'oversampling ont été explorées, notamment l'oversampling aléatoire (RandomOverSampler - ROS) et le SMOTE (Synthetic Minority Over-sampling Technique). Le ROS crée de nouvelles instances de la classe minoritaire en répliquant de manière aléatoire des échantillons existants de cette classe avec la possibilité d'ajouter un bruit pour ne pas avoir les mêmes échantillons afin d'éviter le surajustement. En revanche, le SMOTE génère de nouveaux échantillons synthétiques en interpolant entre les échantillons de la classe minoritaire. Le SMOTE a été choisi, car il augmente la diversité de l'ensemble des données, ce qui contribue à réduire le surajustement et préserve la séparabilité des classes.

La figure 4.20 illustre la répartition des classes avant et après le rééquilibrage avec les deux approches. On constate que le SMOTE préserve la séparabilité des classes après le rééquilibrage, ce qui peut contribuer à une amélioration de la classification. En revanche, le ROS peut entraîner une certaine confusion entre les classes et une perte de séparabilité. Il est essentiel de trouver le bon équilibre pour améliorer les performances du classifieur tout en évitant les effets indésirables du rééquilibrage.

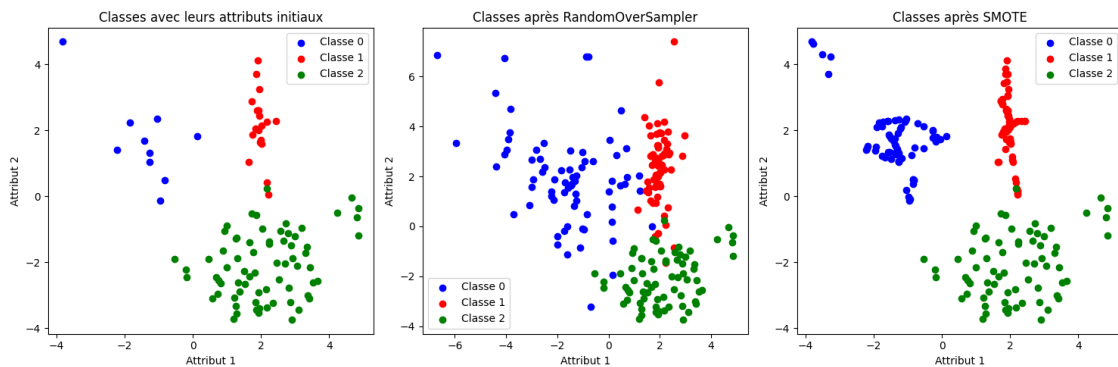


FIGURE 4.20 – Comparaison des classes avant et après le rééquilibrage

## 4.4 Détection de nouveautés

La détection de nouveauté, également appelée détection d'anomalies ou détection d'objets inconnus, est une tâche importante dans le domaine de l'apprentissage automatique. Alors que les classifieurs traditionnels ont tendance à renvoyer une prédiction basée sur les classes existantes, même pour des exemples inhabituels, la détection de nouveauté vise à identifier des exemples qui diffèrent considérablement des données d'entraînement.

Dans le cadre de notre étude, La détection de nouveauté dans le contexte de la reconnaissance d'instruments chirurgicaux sera abordée. L'objectif est d'identifier des instruments ou des configurations d'instruments qui ne sont pas présents dans notre ensemble de données d'entraînement, afin d'améliorer la capacité du système à reconnaître et à réagir à des situations nouvelles ou inattendues.

Cette section présentera en détail la méthode de détection de nouveauté utilisée et les stratégies d'évaluation. Nous discuterons également des avantages, des limitations et des perspectives d'amélioration de cette méthode dans le contexte de notre application.

### 4.4.0.1 La méthode ADDaMM

La méthode ADDaMM (Anomaly Detection using Density ratio and Mixture Models) est une approche probabiliste de détection de nouveauté qui combine les concepts de

modèles de mélange et d'estimation de densité par noyau.

Traditionnellement, les modèles de mélange gaussien sont utilisés pour modéliser la distribution des données en supposant qu'elle est composée d'un mélange de différentes distributions de probabilité. Chaque composante du mélange est caractérisée par ses propres paramètres, tels que les poids et les paramètres de la distribution. Cependant, dans le contexte de la méthode ADDaMM, une approche différente est employée pour estimer la distribution des données. Cette approche, connue sous le nom d'estimation de densité du noyau (Kernel Density Estimation, KDE), va encore plus loin dans le concept du mélange de gaussiennes. Contrairement aux modèles de mélange gaussien traditionnels qui supposent un nombre prédéfini de composantes, la KDE utilise un mélange composé d'une composante gaussienne par point de données. Cette approche se traduit par un estimateur de densité essentiellement non paramétrique, évitant ainsi de définir un nombre fixe de composantes.

Supposons que nous disposions d'un ensemble de données d'entraînement  $X = x_1, x_2, \dots, x_n$ , où chaque  $x_i$  est un vecteur d'attributs. L'objectif est de détecter les observations qui ne suivent pas la distribution des données d'entraînement, c'est-à-dire les nouveautés.

La méthode ADDaMM utilise l'estimation de densité par noyau Gaussien pour modéliser la distribution des données d'entraînement. L'estimation de densité par noyau estime la densité de probabilité  $p(x)$  à partir des données en considérant chaque observation  $x_i$  comme un échantillon potentiel d'une variable aléatoire  $X$ . La fonction de densité de probabilité estimée est donnée par :

$$\hat{p}(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - x_i) \quad (4.14)$$

où  $K_h(\cdot)$  est une fonction de noyau et  $h$  est le paramètre de bande passante (bandwidth) qui contrôle la taille de la fenêtre de lissage appliquée lors de l'estimation.

Pour détecter les nouveautés, nous évaluons la probabilité qu'une nouvelle observation  $x$  soit générée par le modèle appris. Cette probabilité est mesurée en utilisant le score log-vraisemblance, qui est défini comme le logarithme de la densité de probabilité estimée  $\hat{p}(x)$  pour l'observation  $x$ . Mathématiquement, le score log-vraisemblance est donné par :

$$\log(\hat{p}(x)) = \log \left( \frac{1}{n} \sum_{i=1}^n K_h(x - x_i) \right) \quad (4.15)$$

Une observation avec un score log-vraisemblance faible indique une probabilité élevée de nouveauté, car elle ne suit pas la distribution des données d'entraînement.

Il convient de noter que dans cette approche, le choix du paramètre de bande passante  $h$  est crucial dans l'estimation de densité par noyau. Une valeur plus petite de la bande passante conduit à une estimation plus précise, mais moins lissée, tandis qu'une valeur plus grande conduit à une estimation plus lissée, mais moins précise. Autrement dit, pour des valeurs très petites, le système devient agressif en termes de détection de nouveauté, surtout que les caractéristiques extraites à partir d'un échantillon ne sont pas stables. Pour des valeurs importantes, le système devient moins sensible, c'est-à-dire qu'il suffit que l'objet ressemble un peu à un des éléments présents dans les données d'entraînement pour qu'il ne soit pas détecté comme nouveauté.

La figure 4.21 illustre l'influence du paramètre de bande passante (BP) sur la détection de nouveauté à l'aide de la méthode ADDaMM. Les données synthétiques utilisées comprennent 150 échantillons répartis en 12 classes différents. Chaque échantillon est un vecteur bidimensionnel, représentant les attributs de l'objet.

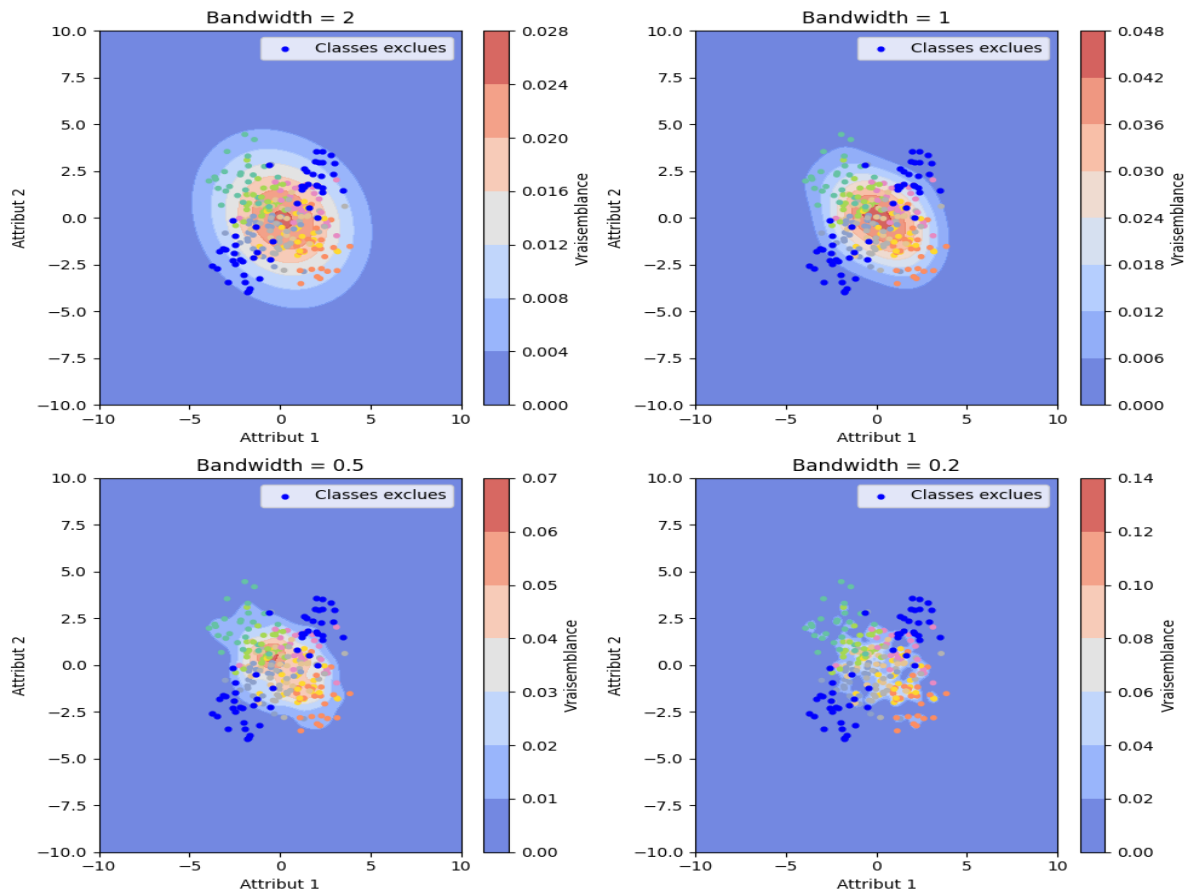


FIGURE 4.21 – Effet de la bande passante sur l'estimation de la densité

Pour chaque valeur de bande passante, le modèle ADDaMM est entraîné en excluant deux classes spécifiques des données d'entraînement. Les classes exclues sont identifiées dans la légende du graphique. Les zones colorées sur les graphiques représentent les contours de densité estimée à l'aide du modèle ADDaMM. Plus la couleur est foncée, plus la densité estimée est élevée, indiquant une probabilité plus élevée de correspondance à la distribution des données d'entraînement.

Sur la figure, on peut observer l'effet de la valeur de la BP sur l'estimation de la densité et la délimitation des régions de décision. Une valeur plus élevée de la BP conduit à une estimation plus lissée de la densité, ce qui se traduit par des contours de densité plus étendus et des régions de décision plus grandes. Dans ce cas, les classes exclues peuvent être considérées comme des habituelles, ce qui n'est pas souhaité.

En revanche, une valeur plus faible de la BP donne une estimation plus précise, mais moins lissée de la densité. Cela se traduit par des contours de densité plus serrés et des régions de décision plus petites. Dans ce scénario, les classes exclues peuvent être correctement isolées. Cependant, si l'un des échantillons s'éloigne légèrement de la distribution de sa classe d'origine, il peut être considéré comme une nouveauté, même si cette classe est présente dans les données d'entraînement.

Pour choisir une valeur de la BP, plusieurs tests ont été effectués sur différents jeux de données. Après évaluation, nous avons opté pour une valeur de bande passante de 0.05 qui offre un équilibre entre précision et sensibilité dans la détection des objets anormaux.

Nous avons mis en place un système de classification des vraisemblances prédites en trois catégories distinctes :

- Si la vraisemblance prédite est inférieure à 40% de la vraisemblance moyenne des données d'entraînement, l'objet est considéré comme étrange et le classifieur ne fournit aucune prédiction. Cette catégorie indique une forte probabilité d'anomalie.

- Si la vraisemblance prédite se situe dans l'intervalle de 40% à 85% de la vraisemblance moyenne des données d'entraînement, le classifieur fournit une prédiction, mais avec un carré orange pour indiquer une incertitude quant à l'appartenance de l'objet à une classe présente dans les données d'entraînement. Cela permet de signaler une certaine ambiguïté dans la classification.
- Enfin, si la vraisemblance prédite est supérieure à 85% de la vraisemblance moyenne des données d'entraînement, le classifieur fournit une prédiction normale. Cette catégorie indique une forte probabilité de conformité de l'objet à la classe prédite.