



ALGORITHME_MEAN_SHIFT.PPTX

****Résumé détaillé du document : “Algorithme Mean Shift – Présentation et Analyse”**** --- ## 1. Titre / sujet principal ****Mean Shift – Algorithme de clustering par déplacement de densité**** (Présenté par : Khalloua Mohamed – Encadré : Abdel-alim Sadiq – Master Génie Logiciel pour le Cloud, 2024-2026) --- ## 2. Structure du document (diapositives) | N° de slide | Titre / thème | Points clés / résumé | |-----|-----|-----| | ****1**** | Présentation générale | • Nom de l’algorithme, auteurs, cadre académique. | | ****2**** | Visualisation du Mean Shift – effet du ****paramètre de largeur de bande**** (h). | • ****Petite largeur**** (h = 0.8) → plus de clusters, risque de capturer le bruit.

- ****Grande largeur**** (h = 2.0) → moins de clusters, fusion possible de groupes distincts.
- Animation : points convergent vers les ****modes de densité****. | | ****3**** | Applications du Mean Shift | • ****Segmentation d’images****
- ****Clustering de données**** multidimensionnelles
- ****Suivi d’objets**** dans les vidéos
- ****Analyse de données spatiales**** (géographique). | | ****4**** | Effet détaillé du ****paramètre de largeur de bande**** | • ****Largeur de bande (h)**** : taille du voisinage considéré.
- ****Petite h**** → sur-apprentissage (trop de clusters).
- ****Grande h**** → sous-apprentissage (trop peu).
- ****Valeur optimale**** capture la vraie structure des données. | | ****5**** | ****Avantages**** vs ****Inconvénients**** | ****Avantages**** :
 - Pas de nombre de clusters prédéfini (non-paramétrique).
 - Détection de formes arbitraires (non sphériques).
 - Robustesse aux valeurs aberrantes.
 - Base statistique solide (estimation de densité par noyau).****Inconvénients**** :
 - Choix critique de la largeur de bande.
 - Complexité $O(n^2)$ → lent pour grands jeux de données.
 - Malédiction de la dimensionnalité.
 - Convergence parfois lente. | | ****6**** | ****Conclusion & références**** | ****Points clés**** :
 - Non-paramétrique, intuitif, flexible, robuste.****Stratégies d’optimisation**** :
 - Sélection automatique de h (validation croisée).
 - Noyaux adaptatifs (densité locale).
 - Implémentations rapides (structures spatiales).****Références**** :
 1. Comaniciu & Meer (2002).
 2. Documentation Scikit-learn (MeanShift).
 3. Fukunaga & Hostetler (1975). | | ****7**** | Remerciements | “Merci pour votre attention”. | | ****8**** | Introduction au Mean Shift (définition) | • Algorithme non-paramétrique, ne nécessite pas le nombre de clusters a priori.



• Déplacement itératif des points vers les **modes de densité**. | **9** | Applications clés (vision par ordinateur, analyse spatiale, exploration de données). | **10-11** | Formules mathématiques du vecteur de déplacement **$m(x)$** (définition du vecteur de déplacement). | **12-14** |

Implémentation Python (noyau gaussien, mise à jour des points, boucle d'itération). | **15** | Visualisation du noyau gaussien 2D + comparaison avec K-Means (K-Means impose un nombre fixe de clusters sphériques). | --- **3**. Résumé de chaque section **2**. Visualisation du Mean Shift - **Paramètre h** détermine le voisinage utilisé pour estimer la densité. - **Petite h** → nombreux petits clusters (risque de bruit). - **Grande h** → moins de clusters (risque de fusion de groupes distincts). - L'animation montre les points qui se déplacent progressivement vers les **modes de densité**. **3**. Applications - **Segmentation d'images** : partitionner une image en régions homogènes. - **Clustering de données** : découvrir des regroupements naturels dans des espaces multidimensionnels. - **Suivi d'objets** : suivre la trajectoire d'un objet à travers une séquence vidéo. - **Analyse spatiale** : identifier les zones d'intérêt dans des données géographiques (ex. cartes de chaleur). **4**. Largeur de bande (h) - **Petite largeur** → sur-segmentation, possible **over-fitting**. - **Grande largeur** → sous-segmentation, possible **under-fitting**. - **Largeur optimale** : équilibre qui capture la structure réelle des données. **5**. Avantages / Inconvénients | Avantages | Inconvénients | |-----|-----| **Pas de nombre de clusters fixé** → l'algorithme détermine automatiquement le nombre de clusters. | **Choix de h** très critique ; mauvaise valeur entraîne un mauvais clustering. | **Formes de clusters arbitraires** (non sphériques). | **Complexité $O(n^2)$** → difficile à appliquer sur de très grands ensembles. | | **Robuste aux outliers** (moins sensible au bruit). | **Malédiction de la dimensionnalité** : performance diminue en haute dimension. | **Fondement statistique** (estimation de densité par noyau). | **Convergence** parfois lente, surtout pour des données complexes. | **6**. Conclusion & Optimisations - **Points clés** : non-paramétrique, intuitif, flexible, robuste. - **Optimisations** proposées : 1. **Sélection automatique de h** (validation croisée). 2. **Noyaux adaptatifs** basés sur la densité locale. 3. **Structures de données spatiales** (KD-tree, ball-tree) pour accélérer le calcul. - **Références** majeures : - Comaniciu & Meer (2002) – article fondateur. - Documentation Scikit-learn (implémentation Python). - Fukunaga & Hostetler (1975) – estimation du gradient de densité. **7-15** (Introduction, Formules, Code) - **Définition** : Mean Shift déplace chaque point vers le **mode** (maximum de densité) via un **vecteur de déplacement $m(x)$** calculé à partir d'un **noyau gaussien**. - **Formules** : présentation de la formule du vecteur de déplacement et de son implémentation en Python (étapes : noyau gaussien → mise à jour des points → itération jusqu'à convergence). - **Visualisation** du noyau gaussien 2-D montre la forme du voisinage considéré. - **Comparaison** avec K-Means : Mean Shift détermine automatiquement le nombre de clusters, alors que K-Means impose un nombre fixe et des clusters sphériques de taille similaire. --- **4**. Points remarquables / statistiques - **Largeur de bande typique** : ' $h \approx$ moyenne des distances entre voisins proches'. - **Complexité** : $O(n^2)$ (temps) – limitation pour grands jeux de données. - **Comparaison** avec K-Means : - **Mean Shift** : non-paramétrique, détecte des formes complexes, robuste aux outliers. - **K-Means** : nécessite le nombre de clusters, forme sphérique, sensible aux outliers. --- **5**. Objectif global du document Fournir une **introduction complète** au **Mean Shift** : expliquer son principe de **déplacement de densité**, illustrer l'impact du



paramètre de largeur de bande, présenter les **applications** (vision par ordinateur, suivi d'objets, segmentation d'images, analyse spatiale), détailler **avantages et limites**, et proposer **stratégies d'optimisation** (sélection automatique de la bande, noyaux adaptatifs, structures de données). Le tout est illustré par des visualisations, des formules mathématiques et un exemple d'implémentation Python. --- **Résumé final** : Le **Mean Shift** est un algorithme de clustering non-paramétrique qui, grâce à un déplacement itératif vers les zones de densité maximale, détecte automatiquement le nombre et la forme des clusters, tout en étant robuste aux valeurs aberrantes. Son principal défi réside dans le choix de la largeur de bande et la complexité computationnelle, mais des stratégies d'optimisation existent pour le rendre applicable à des ensembles de données de taille moyenne. --- *Document synthétisé à partir des diapositives fournies.*