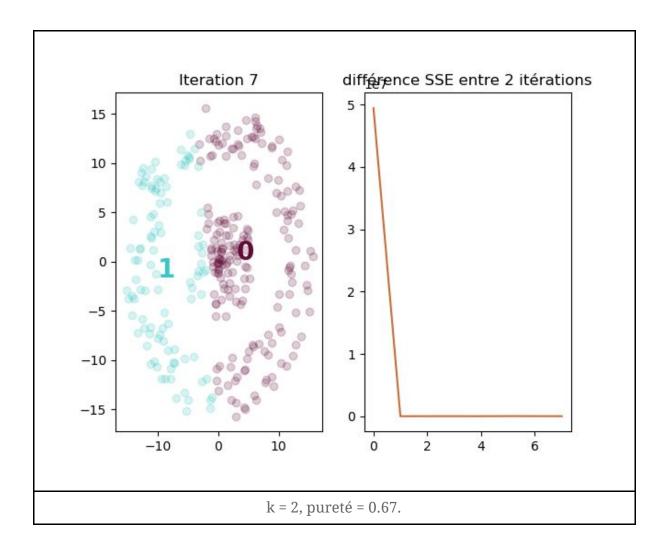
## Introduction à l'Apprentissage Automatique TP1

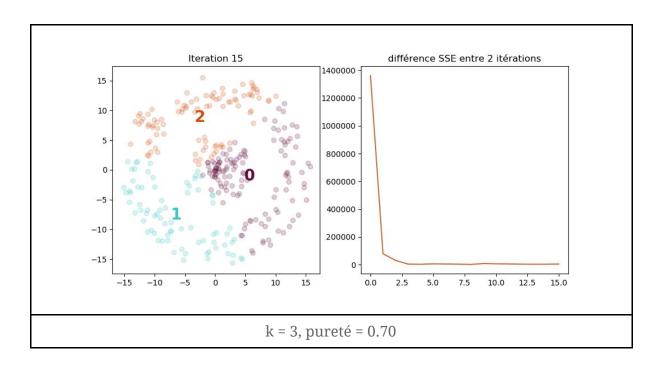
K-means

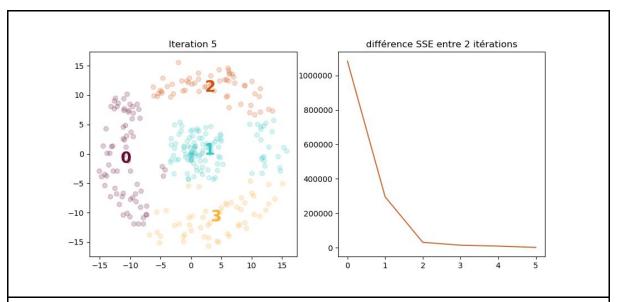
22/01/2019

Dataset 1:

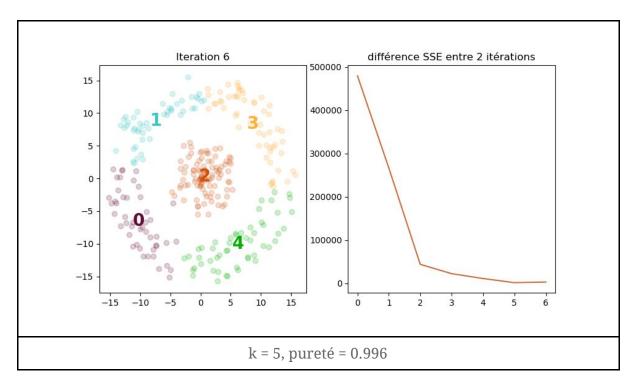
Résultats de K-means avec distance euclidienne.







k = 4, pureté = 0.91. Il peut être nécessaire de relancer plusieurs fois avant de tomber sur un résultat similaire, la pureté des solutions possible variant entre 0.75 et 0.90.

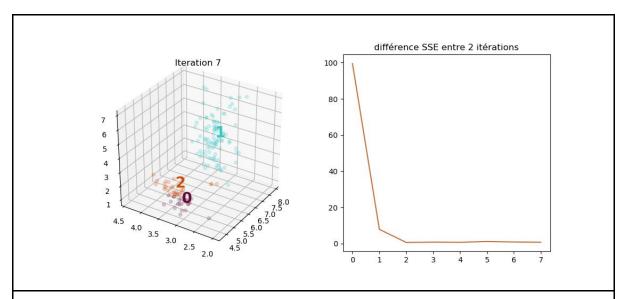


Pour k > 5 le résultat est similaire à k = 5, un centroïde pour le cluster du milieu, le reste se partageant le contour.

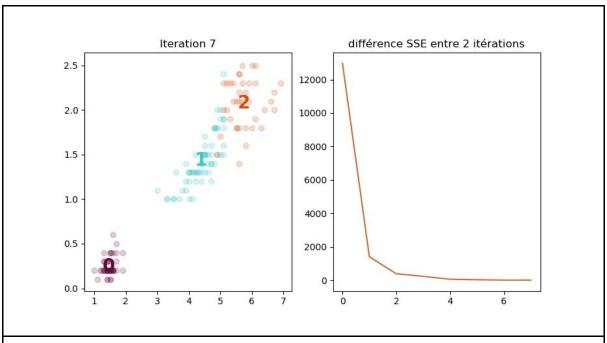
On observe ici clairement un problème avec cette version de K-means, on souhaiterait avoir deux cluster : un central et un extérieur, mais ce n'est pas possible. Il faudrait utiliser un algorithme différent, par exemple kernel k-means.

## **Dataset Iris:**

Résultats de K-means avec distance euclidienne. Nous savons que dans ce dataset k=3.



k = 3, pureté = 0.67. Un exemple d'affichage 3D, ici la classification est mauvaise, car 2 centres d'origine ont été choisi dans un cluster reculé.



k = 3, pureté = 0.89. Représentation selon les axes 2 et 3, plus intuitive que la représentation 3D.