

A2DI - TP2 clustering spectral

Benjamin DANGLOT

Pierre FALEZ

25 janvier 2016

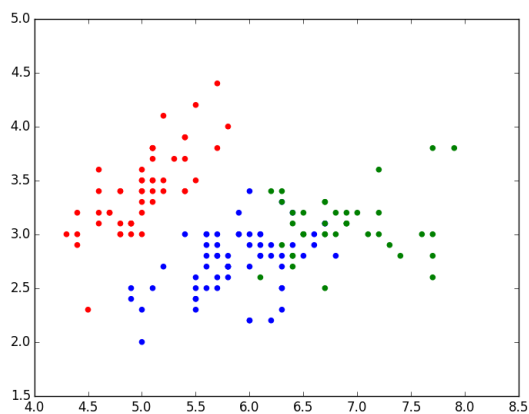
Notre méthodologie consiste à comparer le kmeans classique et le clustering spectral. Pour plus de précision, nous répétons l'algorithme du kmeans, qui est la seule partie non déterministe, 100 fois et nous moyennons le nombre d'erreurs par classe.

Nous avons utilisé différents jeu de données, l'iris, deux demis-lunes et deux cercles imbriqués généré à l'aide de *sklearn*.

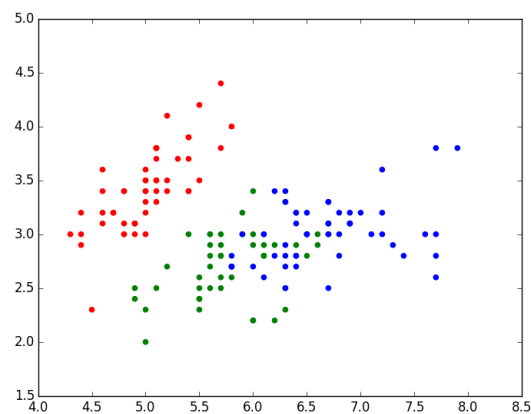
1 Iris

Erreur Kmeans			Erreur Spectrale		
Iris-setosa	Iris-versicolor	Iris-virginica	Iris-setosa	Iris-versicolor	Iris-virginica
0.	8.48	11.66	0.	7.63	7.67

Voici les plots des deux premières dimension du jeu de donnée iris.



(a) kmeans pour iris



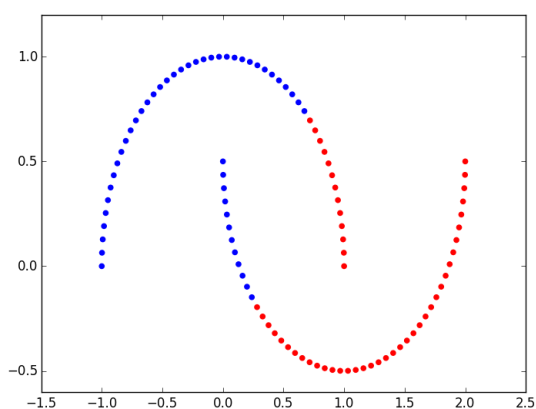
(b) spectral clustering pour iris

L'algorithme spectral permet de faire un peu moins d'erreur.

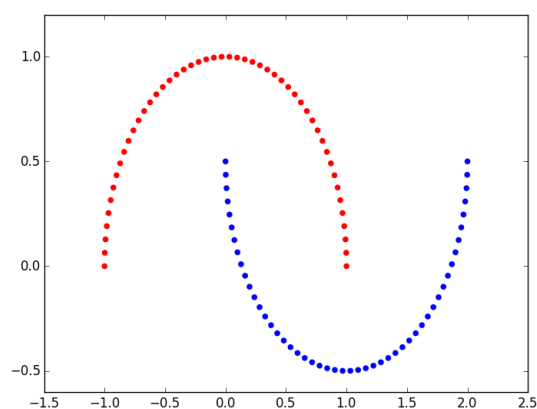
2 Demi-lune

Nous avons essayé l'algorithme sur différent niveau de bruit sur les données :

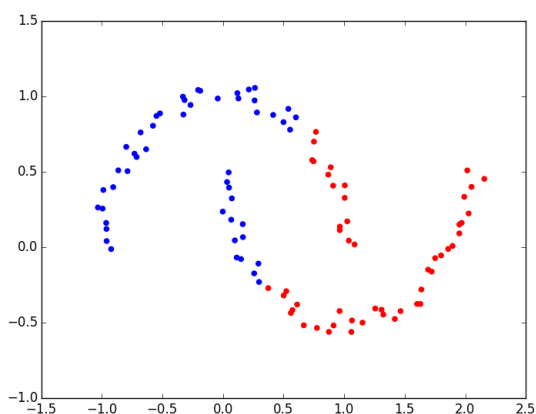
Niveau de bruit	Erreur Kmeans		Erreur Spectrale	
	A	B	A	B
0	12.33	13.33	0	1.5
0.05	13.55	13.34	0	2.5
0.1	12	13.97	0	0.5
0.2	16.32	12.38	0	49



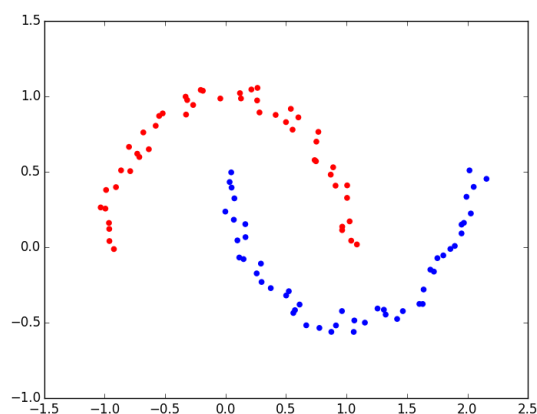
(a) kmeans sans bruit



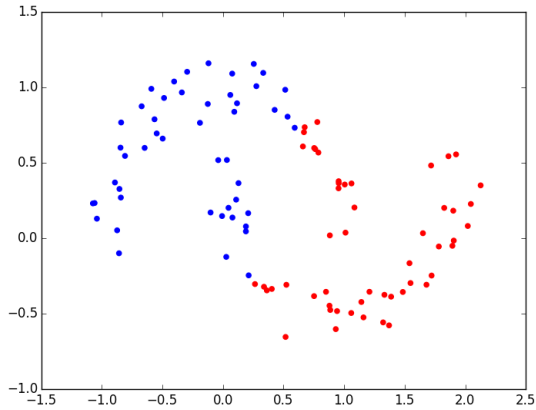
(b) spectral clustering sans bruit



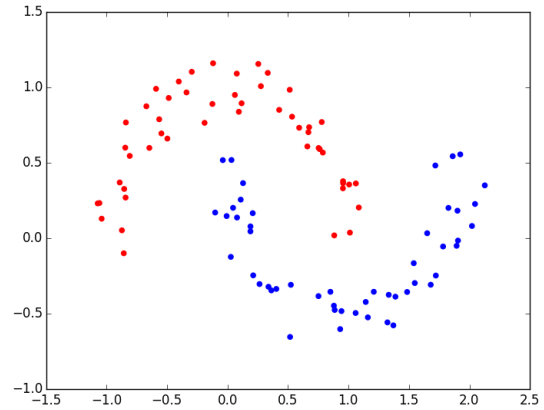
(c) kmeans avec 0.05 de bruit



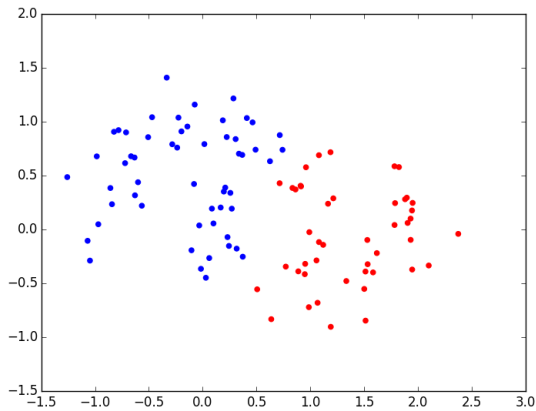
(d) spectral clustering avec 0.05 de bruit



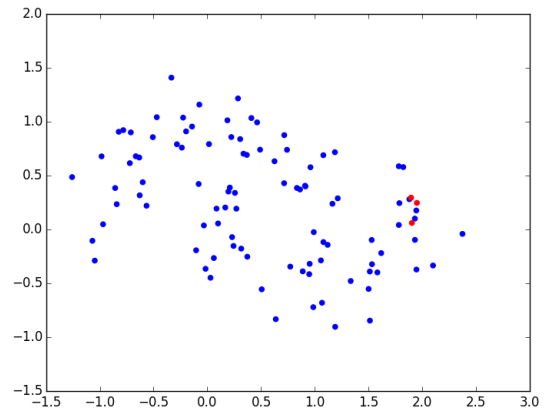
(a) kmeans avec 0.1 de bruit



(b) spectral clustering avec 0.1 de bruit



(c) kmeans avec 0.2 de bruit

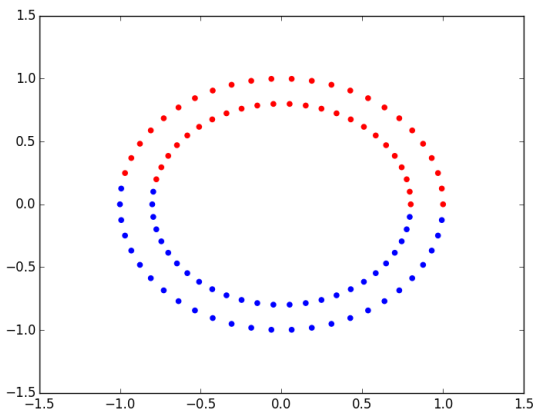


(d) spectral clustering avec 0.2 de bruit

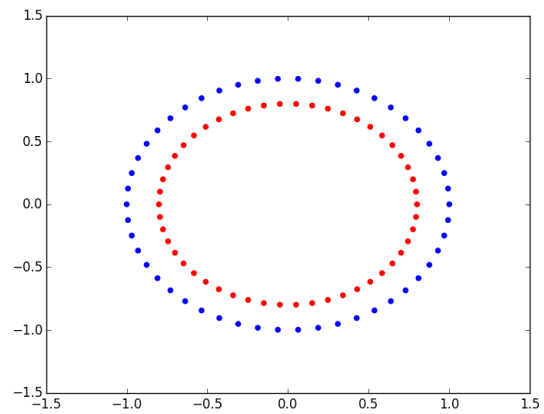
On peut constater que le clustering spectral fonctionne beaucoup mieux, car le kmean classique ne produit que des sous-ensembles connexes. Cependant, on peut voir qu'avec trop de bruit l'algorithme ne fonctionne plus. Il faut également bien ajuster le paramètre σ , dans notre cas $\sigma = 0.2$ a obtenu les meilleurs performances.

3 Cercles imbriqués

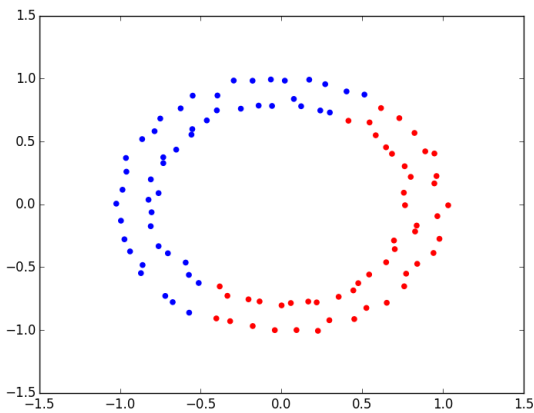
Niveau de bruit	Erreur Kmeans		Erreur Spectrale	
	A	B	A	B
0	0.	50.	0.	1.5
0.025	0.	50.	0.	47



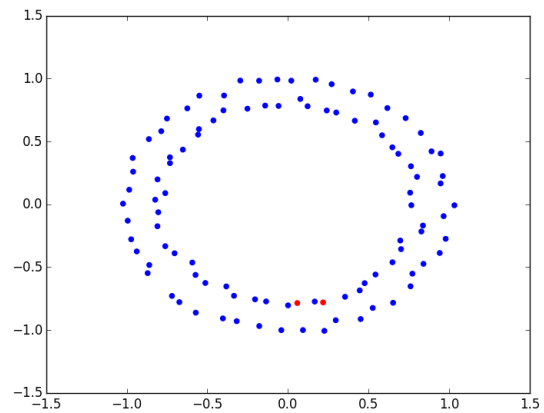
(a) kmeans sans bruit



(b) spectral clustering sans bruit



(c) kmeans avec 0.025 de bruit



(d) spectral clustering avec 0.025 de bruit

On peut voir que le spectral clustering arrive à bien classifier sans bruit, mais avec un tout petit bruit il ne fonctionne plus (nous avons essayé jusqu'à 10^{-5}). Comme pour les demis-lunes, l'algorithme est très sensible au σ . (ici $\sigma = 0.09$).