CR – Soutenance Projet 6

*Conseils Mentor Evaluateur Mohammed Sedki*

* Présentation d'une méthode (**k-means**) : pas nécessaire de mettre les liens web…

*Je vous fais confiance et ne remets pas en cause votre connaissance du k-means*

* Commentaires / ACP & k-means (à partir de 52'12s enregistrement vidéo)

Ce pourquoi je plaide beaucoup, personnellement, je suis un peu contre les k-means sur des axes issus de l'ACP.

(F1, F2, … Fn)

Explications :

* Les axes issus de l'ACP sont des axes conçus pour maximiser la variance, cad que l'on souhaite conserver le maximum de variance du jeu de données de départ.
* Sauf qu'il faut comprendre que les axes qui portent la variance ne sont pas forcement les axes qui portent l'information discriminante.

Exemple :

Pensons que vous avez 2 groupes que vous voulez discriminer, qui sont horizontaux suivant 2 lignes horizontales.

Sous forme de bandes horizontales, l'une au-dessus de l'autre.

On sait que l'axe qui maximise la variance est l'axe des abscisses (x) et l'axe qui discrimine est l'axe des ordonnées (y).

Dans ce cas, si vous prenez l'ACP cela va être pathologique car cela va nous dire qu'un seul axe nous suffit pour conserver toute la variance, alors qu'elle ne discrimine pas.

Par contre lorsque l'on a des DataSet de très grandes dimensions, on n'a pas le choix on fait l'ACP avant, mais on conservera TOUS les axes pour faire notre k-means, ainsi nous aurons l'information qui est conservée.

