

CES Data Scientist : Apprentissage non supervisé

16-17 juillet 2015

Slim ESSID, Olivier CAPPE, Manuel MOUSSALAM

16 juillet 2015

Objectifs pédagogiques

A l'issue de ce cours, les étudiants seront capables de :

- O1** reconnaître un problème non supervisé
- O2** sélectionner une méthode de classification non supervisée pour aborder un problème particulier
- O3** mettre en œuvre différents classificateurs non supervisés et évaluer leurs résultats
- O4** analyser des données nouvelles, sans connaissance a priori à l'aide de classificateurs non supervisés
- O5** expliquer les données observées, à l'aide de modèles à variables latentes
- O6** réutiliser de façon autonome ces savoirs dans un projet personnel

Validation

40% QCM sur des savoirs (*sur Moodle*)

60% pour les savoir faire : rapport individuel d'activité pratique à rendre dans le mois qui suit :

- synthèse des connaissances acquises dans ce module (détaillées) ;
- discussion critique de leur mise en œuvre sur le projet personnel (démonstration d'une réflexion personnelle face à des problèmes concrets) ;
- 1 question que l'on aurait aimé approfondir (pour prendre du recul).

Activités (14h)

Jour 1 (16 juillet)

- | | |
|---|-----------------------------|
| 09 :00-10 :30 Cours : <i>clustering</i> (S. Essid)
<i>Approches hiérarchiques, K-moyennes, modèles de mélanges (GMM)</i> | Objectifs : O1-O2-O3 |
| 10 :45-12 :15 TP : <i>K-means</i> et GMM (S. Essid) | Objectifs : O2-O3-O4 |
| 13 :30-15 :00 Cours : détection de nouveauté/anomalies (O. Cappé)
<i>Tests d'hypothèses, distance de Mahalanobis, détection de ruptures</i> | Objectifs : O3 |
| 15 :15-17 :45 Cours : méthodes de décomposition de matrices (S. Essid)
<i>PCA, NMF, ICA</i> | Objectifs : O2-O4 |

Jour 2 (17 juillet)

09 :00-11 :45 TP : PCA, NMF, ICA (S. Essid)

Objectifs : **O2-O5**

13 :00-14 :30 Intervention d'un industriel (M. Moussallam)

Objectifs : **O3**

Apprentissage non-supervisé chez Deezer

14 :45-17 :45 Aide au démarrage d'un projet personnel (S. Essid)

Objectifs : **O6**

Définition d'un problème pratique sur lequel les étudiants devront fournir un travail personnel au-delà des 2 jours de cours