

# Fouille de données et medias sociaux

## TP5 : Visualisation et recommandation

Olivier Schwander

16 octobre 2017

### Exercice 1 *Réduction de dimension*

Jeux de données :

- Swiss Roll [http://dac.lip6.fr/master/wp-content/uploads/2017/09/swiss\\_roll.csv](http://dac.lip6.fr/master/wp-content/uploads/2017/09/swiss_roll.csv)
- Decathlon <http://dac.lip6.fr/master/wp-content/uploads/2017/09/decathlon.csv> (supprimer les variables qualitatives, et renverser les temps de façon à avoir une valeur élevée pour les meilleurs temps)
- MNIST

#### Question 1

Implémentez le modèle t-SNE.

#### Question 2

Construisez des visualisation (en 2D et en 3D) des jeux de données proposés (ou autres). Comparez avec d'autres méthodes de réduction de dimension, linéaires et non-linéaires (voir <http://scikit-learn.org/stable/modules/classes.html#module-sklearn.manifold>).

#### Question 3

Pour Isomap et LLE, manipulez le paramètre contrôlant la taille du voisinage pour la construction du graphe.

### Exercice 2 *Recommandation*

Jeu de données MovieLens <http://files.grouplens.org/datasets/movielens/> :

- 100k <http://files.grouplens.org/datasets/movielens/ml-100k.zip>
- 1M <http://files.grouplens.org/datasets/movielens/ml-1m.zip>

#### Question 1

Construisez une visualisation avec t-SNE (on peut utiliser `metric=precomputed` pour donner une matrice de distance au lieu des coordonnées des points en entrée de la fonction `sklearn`).

### Question 2

Implémentez un modèle de collaborative filtering avec :

- descente de gradient stochastique
- distance L2
- régularisation L2
- sans puis avec biais

### Question 3

Évaluez vos modèles sur MovieLens.