

# TP 10 : ACP, MDS, LLE, ISOMAP, SNE et d'autres méthodes de réduction de la dimensionnalité

Stéphane Canu

Décembre 2017, ASI, INSA Rouen

Le but du TP est d'étudier et de comparer différentes méthodes de réduction de la dimensionnalité

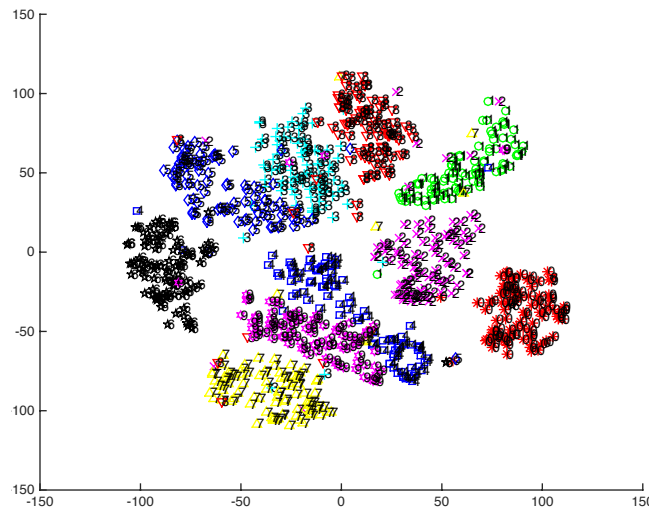


FIGURE 1 – Résultat du TP10

## Ex. 1 — MDS

### 1. Génération des données du problème.

a) Générez les données du problème en utilisant les données MNIST

```
load Donnees_Mnist/mnist-app
load Donnees_Mnist/mnist-test
```

```
b) na = 100;
Xr = [];
yr = [];
for i=1:10
    ii = find(Ya == i-1);
    ind(i,:) = ii(1:na);
    Xr = [Xr ; Xa(ii(1:na),:)]';
    yr = [yr ; Ya(ii(1:na))];
end
```

### 2. ACP

a) Effectuez une analyse en composantes principales des données MNIST. Pensez-vous qu'il faille centrer et/ou réduire les données ?

```
Xac = Xa - ones(length(Xa),1)*mean(Xa);
[U,S,V] = svd(Xac,0);
```

b) Visualisez les données en deux dimensions en utilisant un code couleur/forme différent pour chaque classe

```
c = ['*r'; 'og'; 'xm'; '+c'; 'sb'; 'bd'; 'kp'; 'y^'; 'rv'; 'hm'];
for i=1:10
    plot(U(ind(i,:),1),U(ind(i,:),2),c(i,:))
    text(U(ind(i,:),1),U(ind(i,:),2),num2str(i-1))
end
```

### 3. MDS

- a) Effectuez une MDS des données MNIST

```
D = dist(Xr');  
Um = cmdscale(D,2);
```

- b) Visualisez les données en deux dimensions

### 4. Projection de Sammon

- a) Effectuez une MDS avec une distance de Sammon<sup>1</sup>

```
Um = mdscale(D,2,'criterion','sammon');
```

- b) Visualisez les données en deux dimensions

### 5. ISOMAP

- a) Effectuez une projection suivant le critère ISOMAP<sup>2</sup>

```
Ui = isomap(D,2);
```

- b) Visualisez les données en deux dimensions

### 6. LLE

- a) Effectuez une projection suivant le critère LLE<sup>3</sup>

```
rng default % for reproducibility  
d = 2;  
k = 10; % 1 by class ?  
Ull=lle(Xr,k,d);  
Ul(indp,:) = Ull';
```

- b) Visualisez les données en deux dimensions

### 7. Local MDS Feature Learning

- a) Effectuez une projection suivant le critère LLE<sup>4</sup>

```
d = 2;  
[Ud, total_cost]=MDS_training(D,d,10,0,1);
```

- b) Visualisez les données en deux dimensions

### 8. t-SNE - t-Distributed Stochastic Neighbor Embedding (t-SNE)

- a) Effectuez une projection suivant la méthode tSNE<sup>5</sup>

```
Ul = tsne(Xr,[]);
```

- b) Visualisez les données en deux dimensions

### 9. A vous de jouer :

- Que se passe-t-il lorsque l'on fait varier les paramètres des différentes méthodes ?
- Combien de temps cela prendrait-il de calculer la projection t-SNE sur 10 000 exemples ?
- Proposez une autre méthode de réduction de la dimensionnalité
- Quelle est à votre avis la meilleure méthode de projection, et pourquoi

---

<sup>1</sup><https://github.com/areslp/matlab/blob/master/drtoolbox/techniques/sammon.m>

<sup>2</sup>[https://github.com/gpeyre/matlab-toolboxes/tree/master/toolbox\\_dimreduc](https://github.com/gpeyre/matlab-toolboxes/tree/master/toolbox_dimreduc)

<sup>3</sup><https://cs.nyu.edu/~roweis/lle/code.html>

<sup>4</sup><https://sites.google.com/site/simpmatrix/>

<sup>5</sup><https://lvdmaaten.github.io/tsne/>