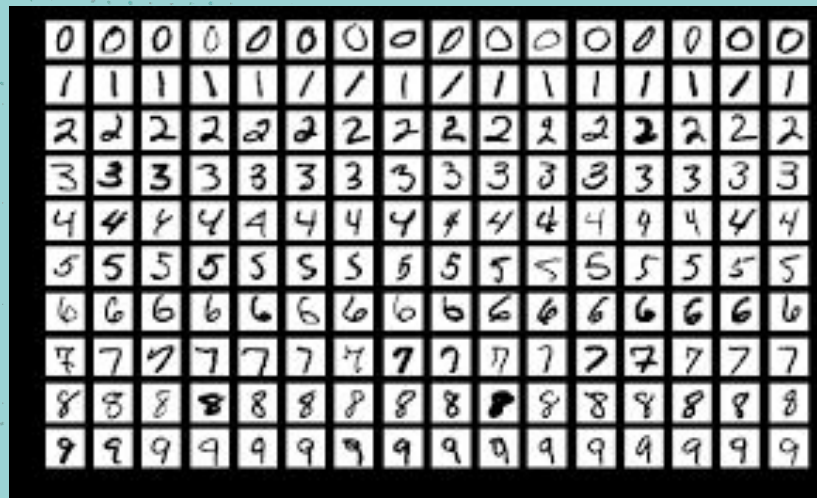


Machine Learning Non supervisé

Présentation des algorithmes
5IABD



Datasets



60 000 images

10 classes

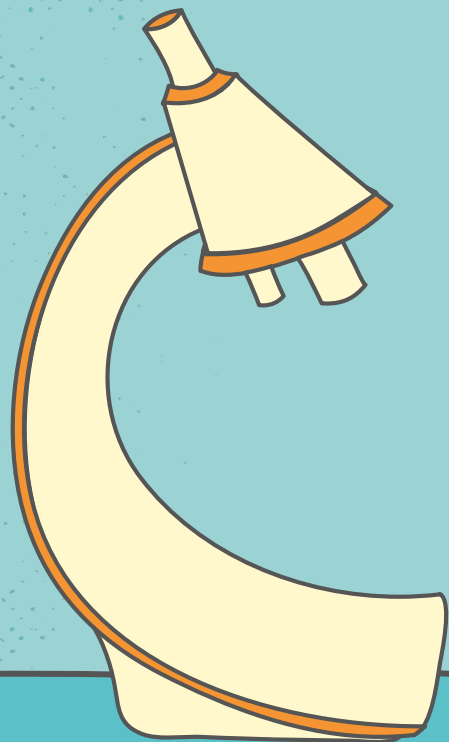


20 933 images

42 classes

Algos implémentés

	SUR MNIST DATASET	SUR SIMPSON DATASET
Kmeans	X	X
PCA	X	X
AutoEncoder	X	X
Kohonen	X	
VAE	X	X



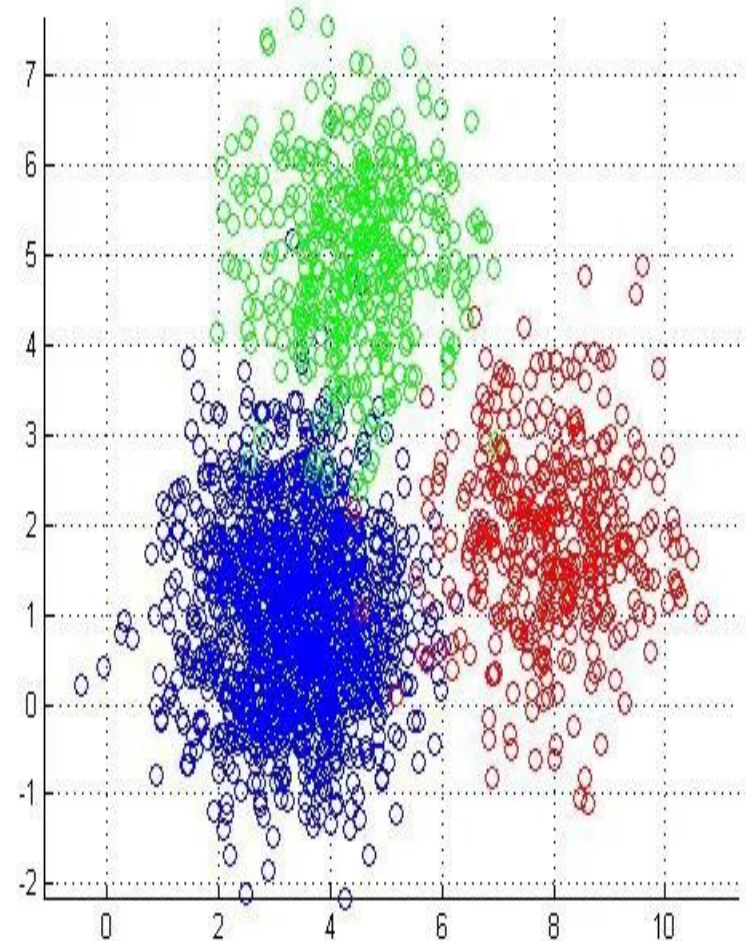
01

K-means

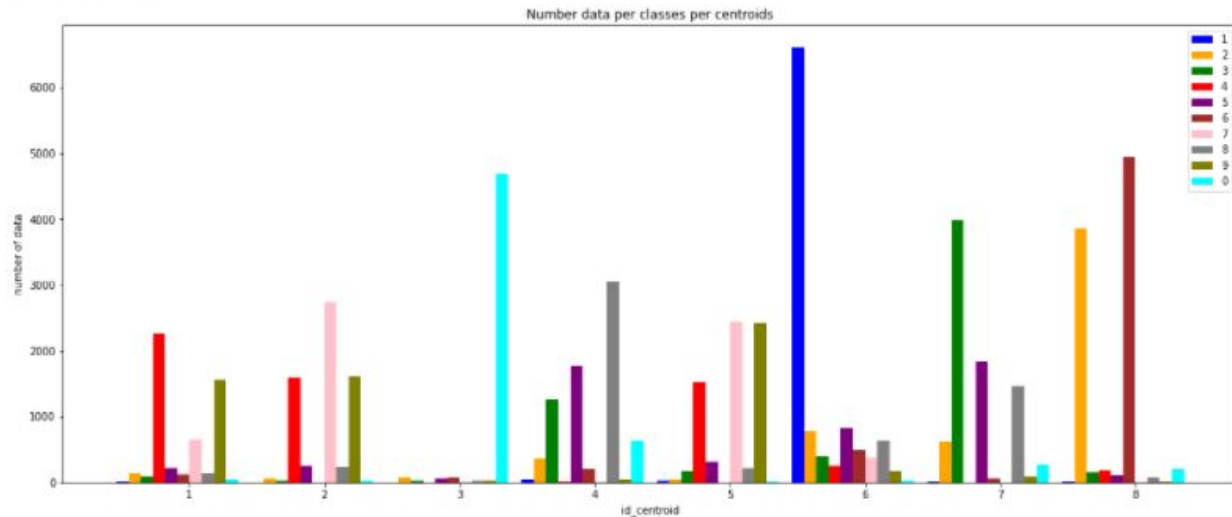
Partitionnement des données et
résolution

Qu'est ce que c'est ?

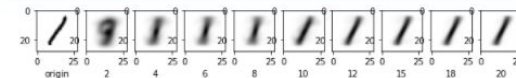
- Détection de "clusters" de données
- Représentants
- Distance Euclidienne



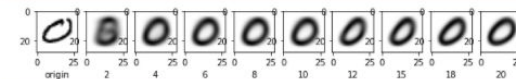
Résultats sur le dataset Mnist



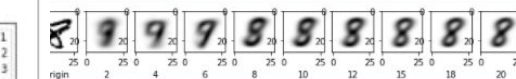
```
15it [00:00, 7913.78it/s]
18it [00:00, 4821.04it/s]
20it [00:00, 7235.30it/s]
```



```
21it [00:00, 498.52it/s]
41it [00:00, 1176.52it/s]
61it [00:00, 2443.28it/s]
81it [00:00, 1952.43it/s]
101it [00:00, 4022.16it/s]
121it [00:00, 2479.64it/s]
151it [00:00, 7053.20it/s]
181it [00:00, 2332.18it/s]
201it [00:00, 5721.72it/s]
```

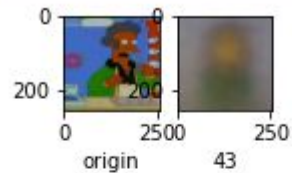


```
21it [00:00, 414.74it/s]
41it [00:00, 4054.43it/s]
61it [00:00, 2926.26it/s]
81it [00:00, 1939.45it/s]
101it [00:00, 2533.40it/s]
121it [00:00, 2115.84it/s]
151it [00:00, 3240.85it/s]
181it [00:00, 3827.89it/s]
201it [00:00, 3666.67it/s]
```



Résultats sur le dataset simpson

43it [00:00, 1433.32it/s]

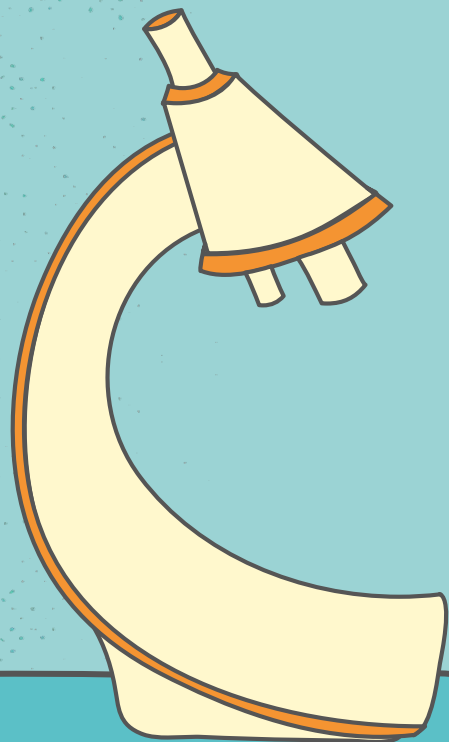


43it [00:00, 1343.79it/s]



43it [00:00, 1323.07it/s]



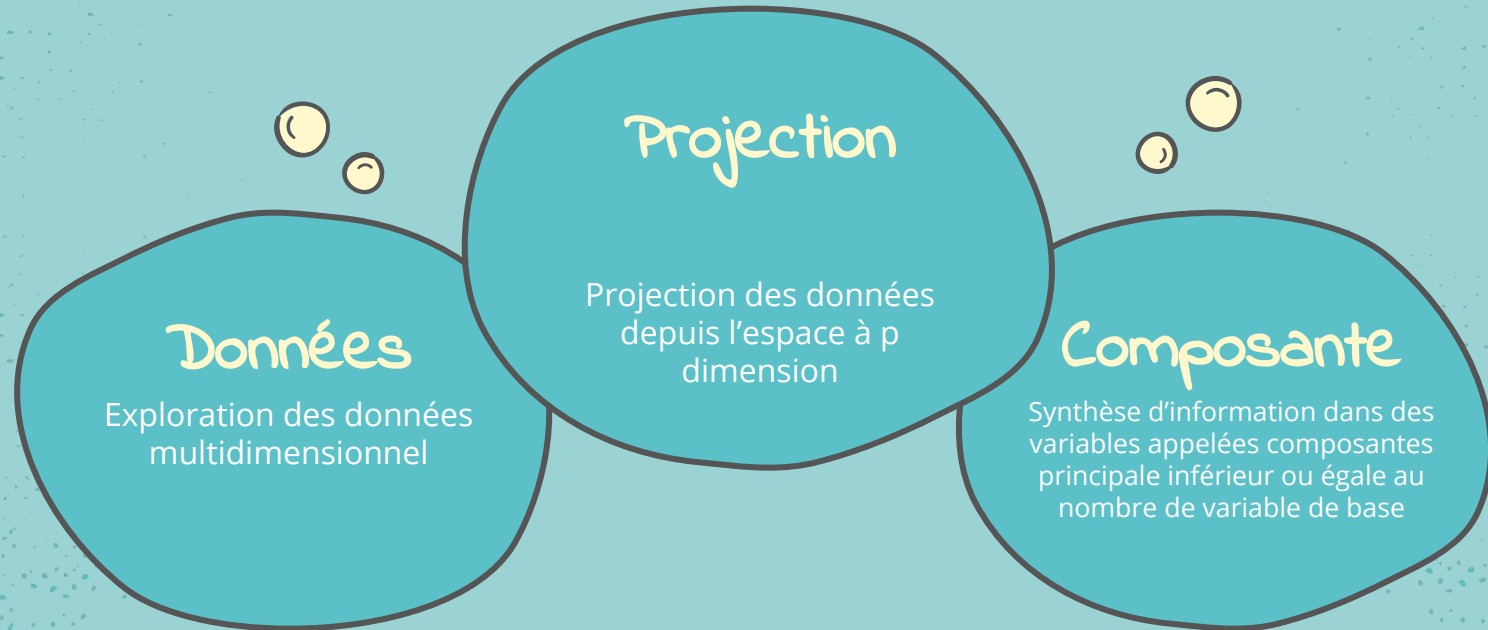


02

PCA

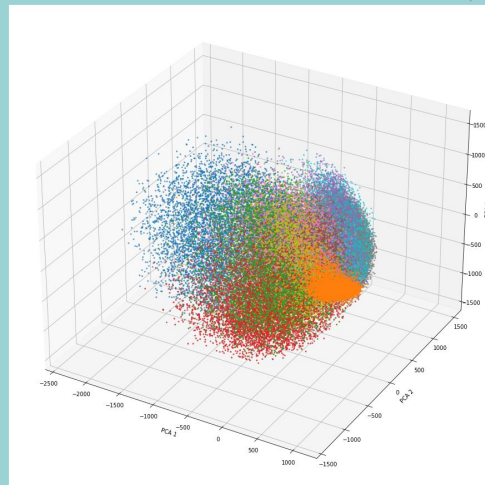
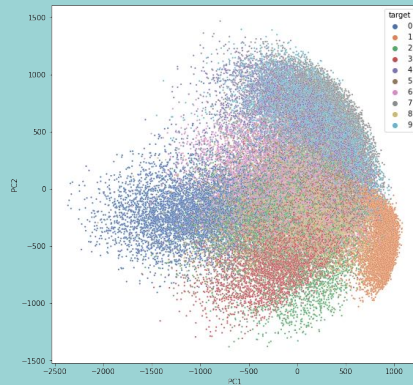
**Partitionnement des données et
résolution**

PCA Principal Component Analysis

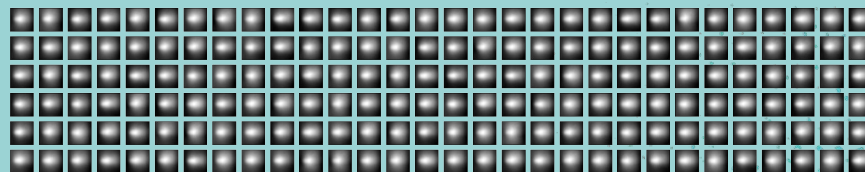
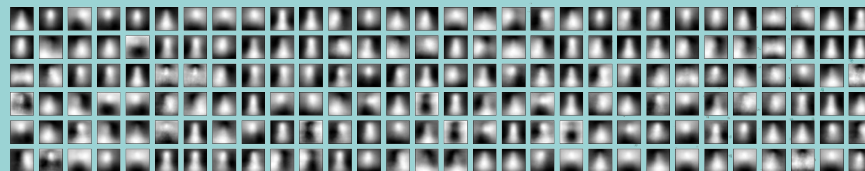
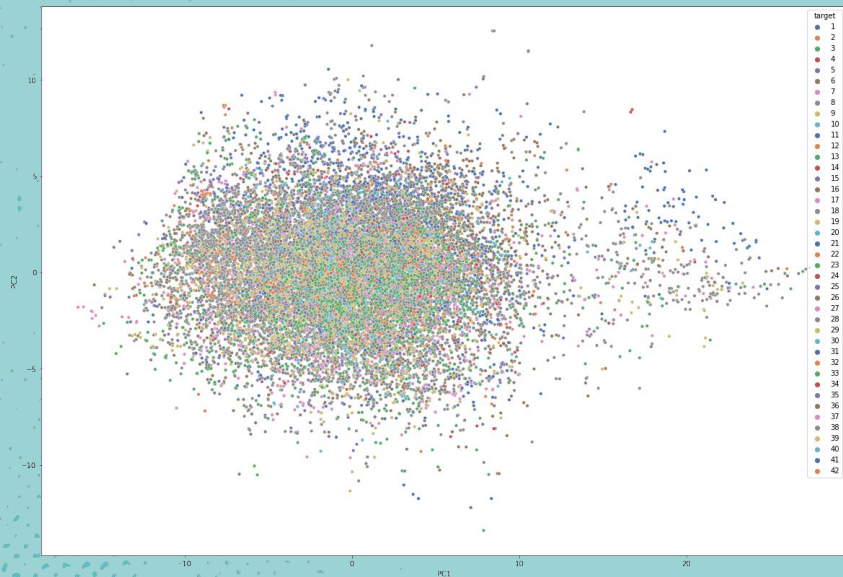


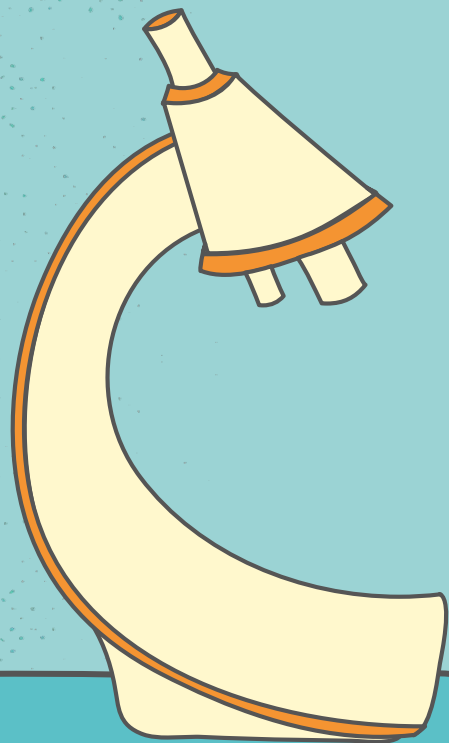
PCA MNIST TEST

5 0 4 1 9 2 1 3 1 4 3 5 3 6 1 7 2 8 6 9 4 0 9 1 1 2 4 3 2 7
3 8 6 9 0 5 6 0 7 6 1 8 7 9 3 9 8 5 9 3 3 0 7 4 9 8 0 9 4 1
4 4 6 0 4 5 6 1 0 0 1 7 1 6 3 0 2 1 1 7 8 0 2 6 7 8 3 9 0 4
6 7 4 6 8 0 7 8 3 1 5 7 1 7 1 6 3 0 2 9 3 1 1 0 4 9 2 0 0
2 0 2 7 1 8 6 4 1 6 3 4 5 9 1 3 3 9 5 4 7 7 4 2 8 5 8 6 9 3
4 6 1 9 9 6 0 3 7 2 8 2 9 4 4 6 4 9 7 0 9 2 7 5 1 5 9 1 9 3



PCA SIMPSONS TEST





03



Auto Encoders

Partitionnement des données et
résolution

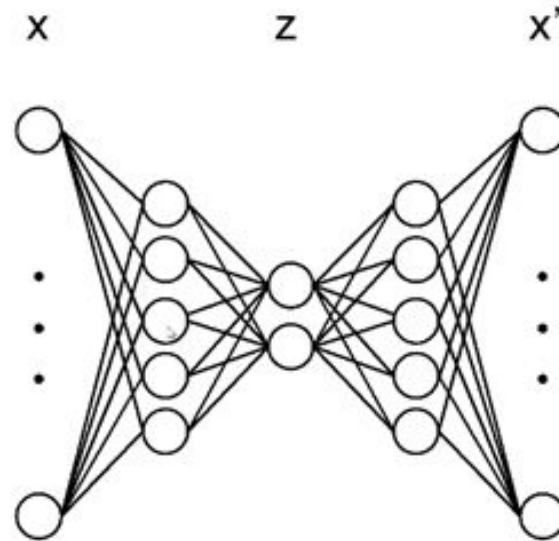
Qu'est-ce que c'est ?

Auto Encoders

- 2 modules :
 - Encoder
 - $z = \text{encode}(x)$
 - Decoder
 - $x' = \text{decode}(z)$

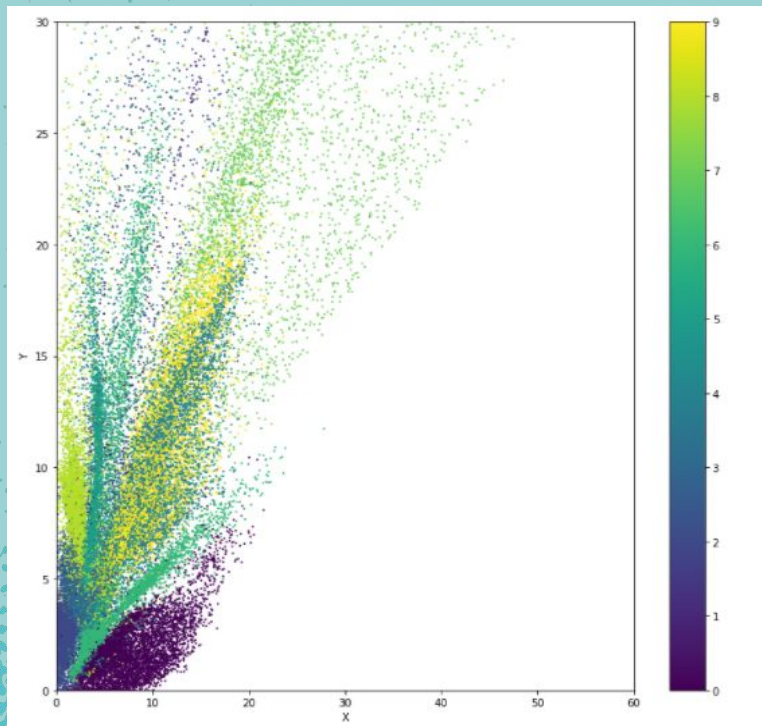
z = espace latent

x' = reconstruction

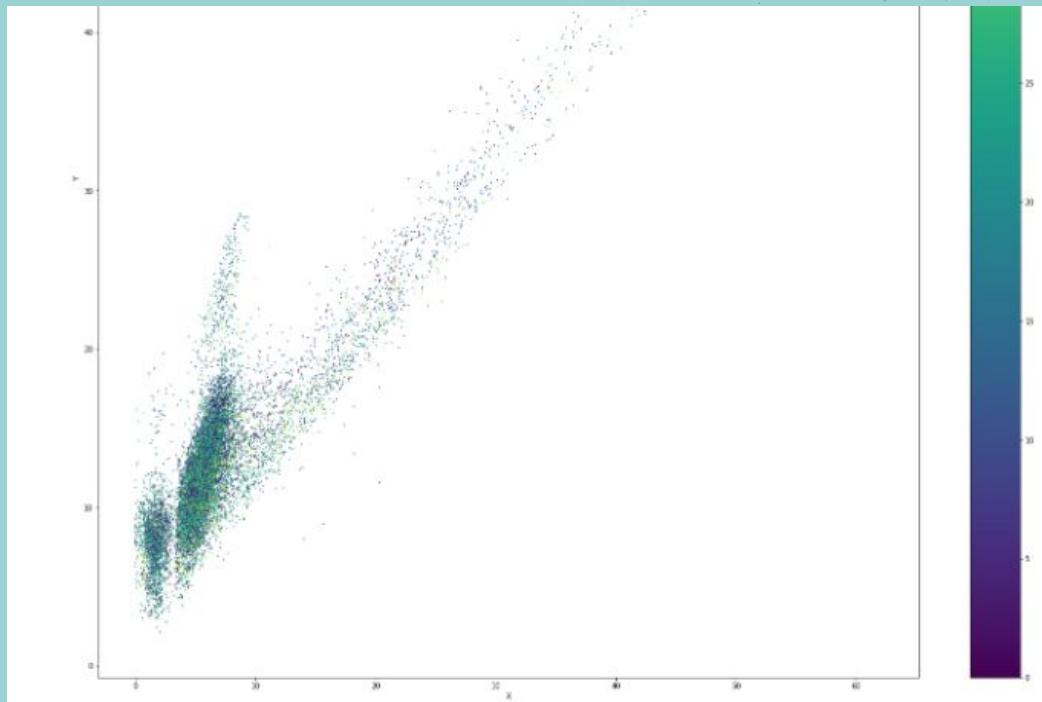


Affichage des données (2D)

MNIST

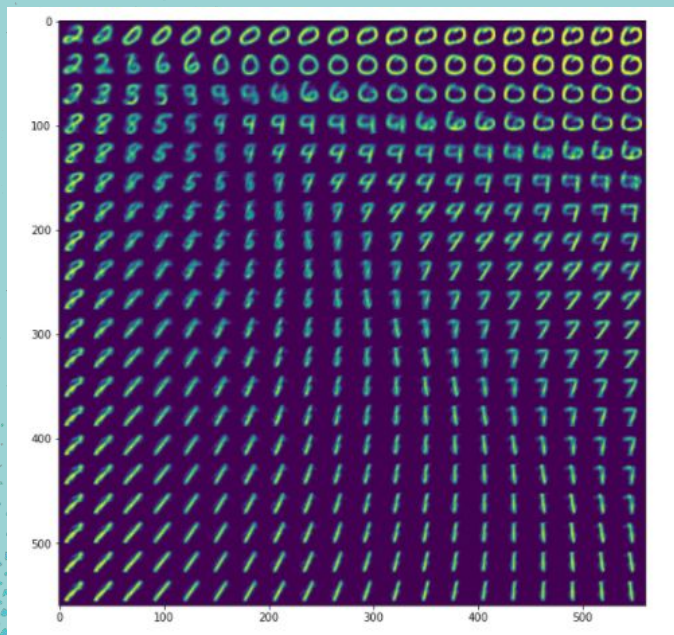


SIMPSONS

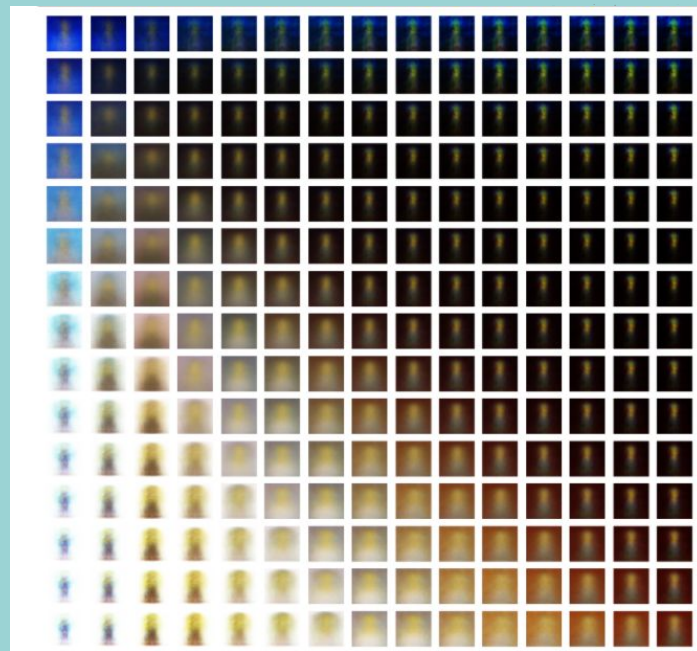


Génération d'images

○ MNIST



○ SIMPSONS

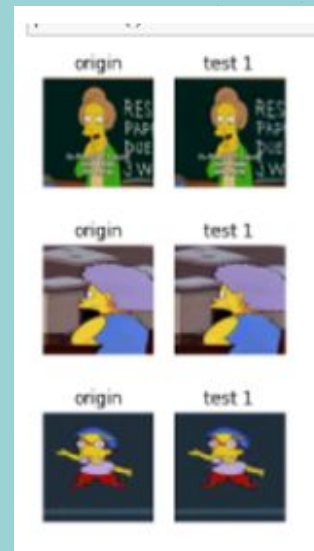
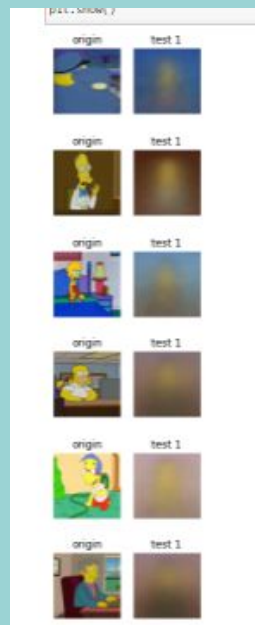
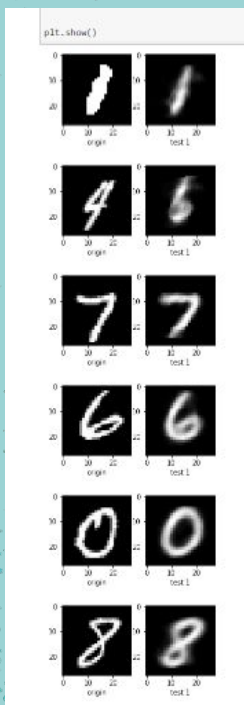


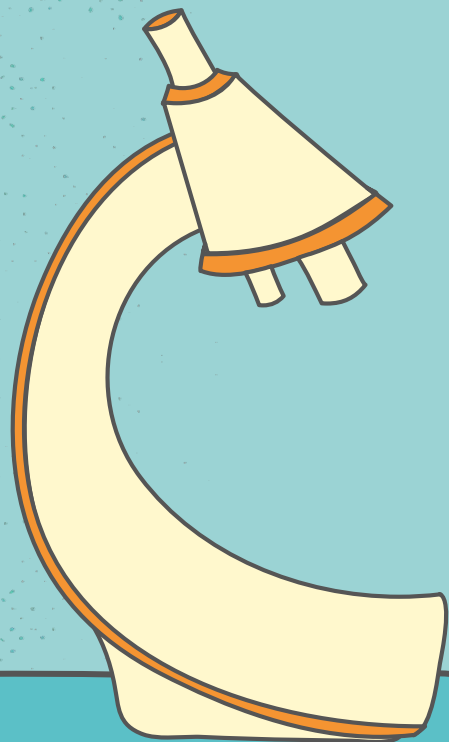
Compression / Decompression

MNIST

SIMPSONS

Meilleur modèle





04



Kohonen Maps

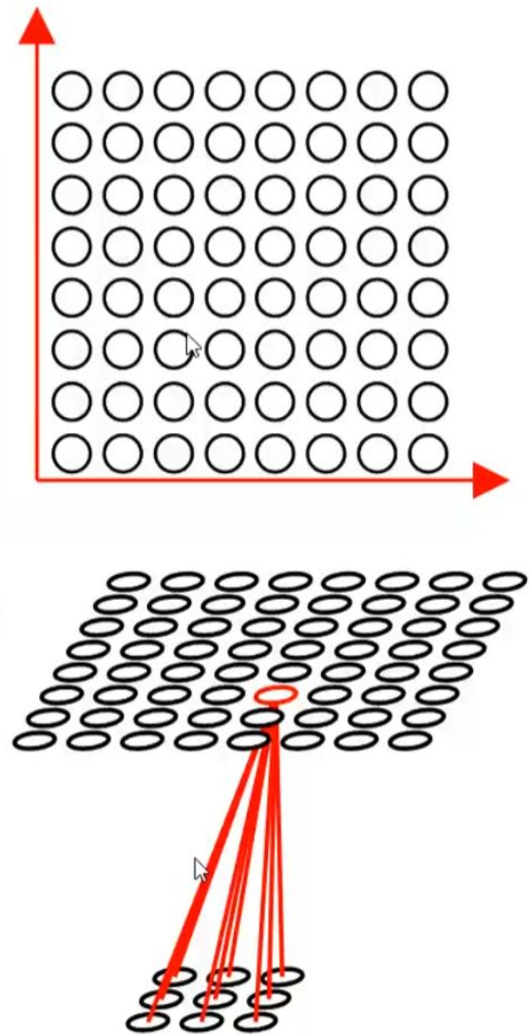
Partitionnement des données et
résolution

Qu'est ce que c'est ?

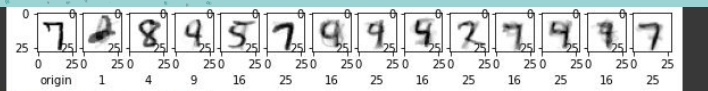
- Reprise des kmeans => ajout d'une contrainte
- Ajout d'un "coordinate vector"
- Sélection d'un exemple aléatoire
- Modification de tout les features vectors des représentants
- Distance Euclidienne
- Mise à jour proportionnelle

• Formule générique classique de mise à jour pour tous les feature vectors :

$$W_i = W_i + \alpha \times e^{\frac{-||C_i - C_k||}{2\gamma}} \times (S_j - W_i)$$



Résultats sur le dataset Mnist



```
1it [00:00, 637.14it/s]
```

```
4it [00:00, 960.23it/s]
```

```
9it [00:00, 10120.30it/s]
```

```
16it [00:00, 6024.68it/s]
```

25it [00:00, 7525.31it/s]

```
16it [00:00, 2254.09it/s]
```

```
25it [00:00, 10122.37it/s]
```

```
16it [00:00, 8101.02it/s]
```

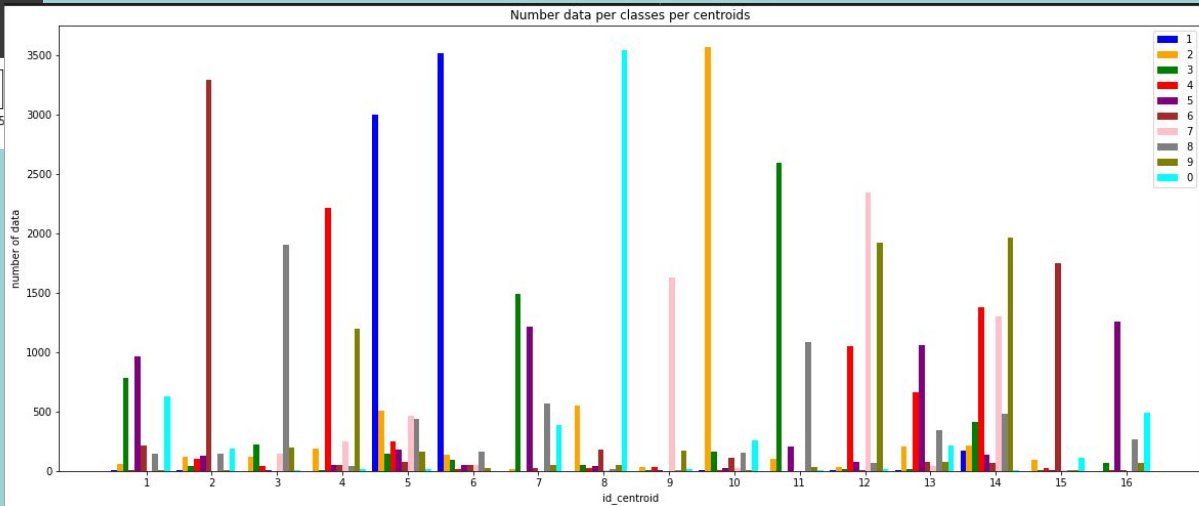
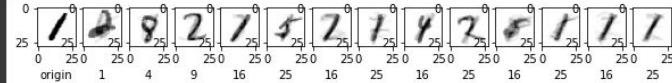
25it [00:00, 20313.37it/s]

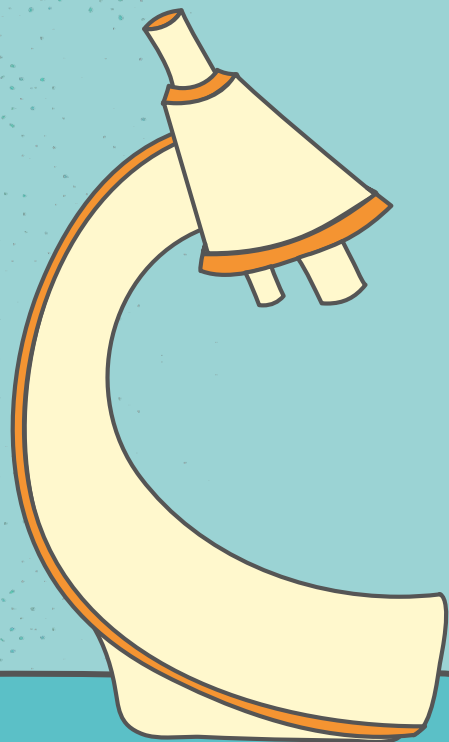
```
16it [00:00, 4995.82it/s]
```

```
25it [00:00, 36058.32it/s]
```

```
16it [00:00, 2102.28it/s]
```

25it [00:00, 3987.74it/s]





05

Variational Auto Encoders

Partitionnement des données et
résolution

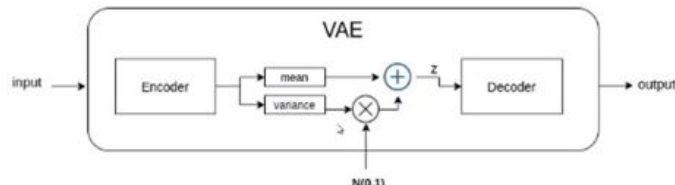
Qu'est-ce que c'est

Variational Auto Encoders

- On repart de l'auto encoder
- Idée :
 - L'encoder ne projette plus directement dans l'espace latent, mais produit les paramètres de plusieurs distributions de probabilités.
 - On choisit en général des Gaussiennes (cf. théorème central limite).
 - Pour générer, le générateur va échantillonner ('sampling') ces distributions de probabilités pour produire des valeurs dans l'espace latent et ensuite fonctionner comme précédemment.

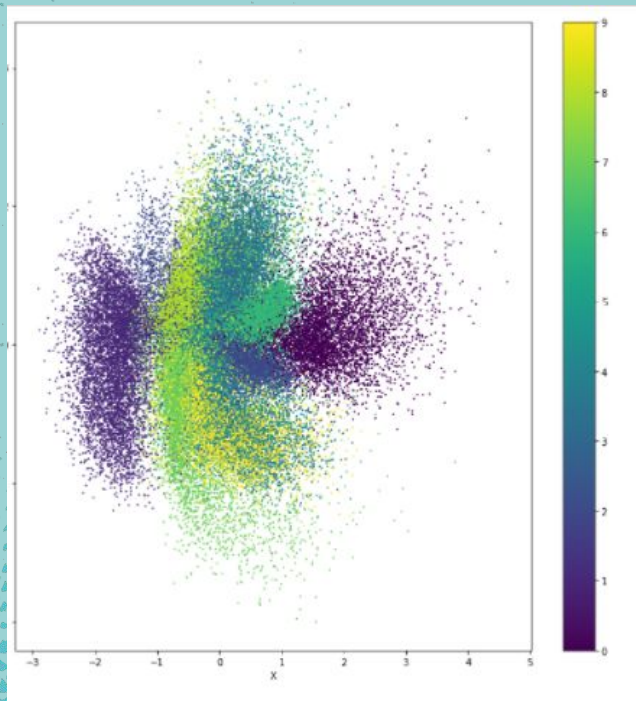
Solution

- Le 'Reparameterization Trick' :
 - Au lieu d'échantillonner la distribution d'espérance μ et d'écart type σ , on va échantillonner la distribution Normale standard : $\mathcal{N}(0,1)$ puis multiplier le résultat par σ et ajouter μ .



Affichage des données (2D)

MNIST



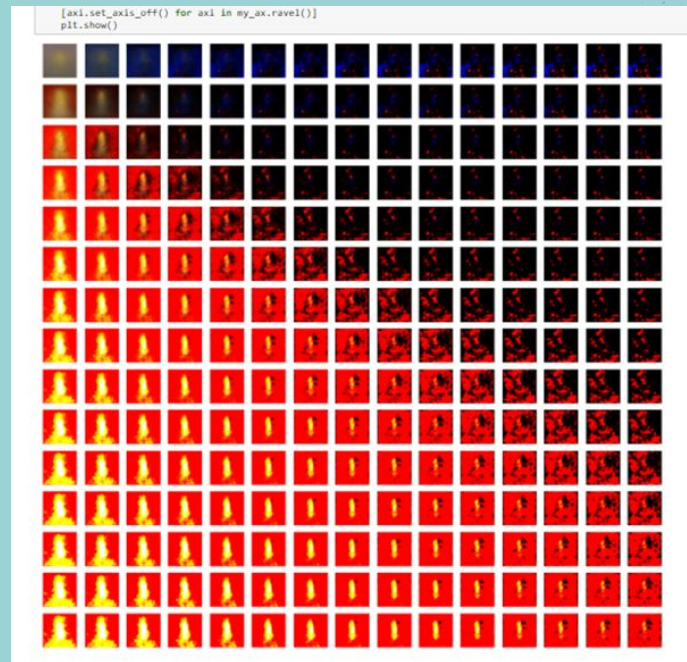
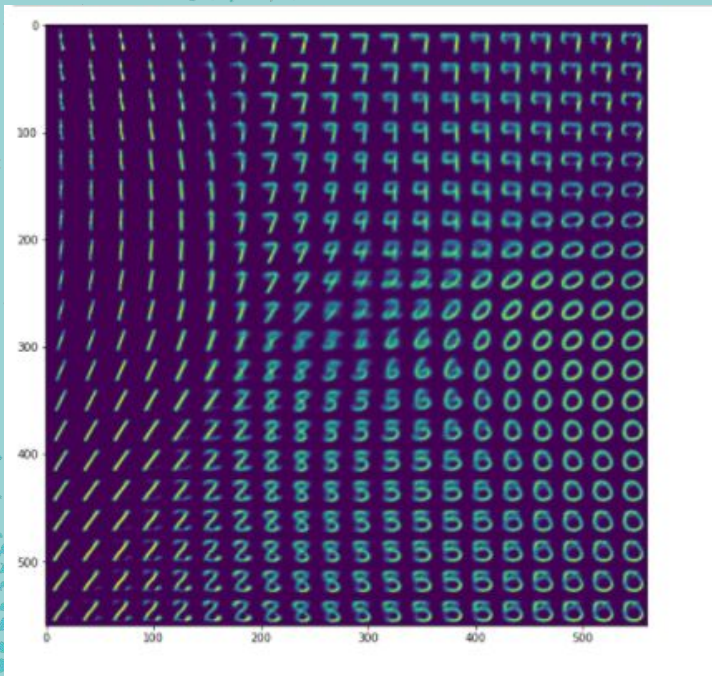
SIMPSONS



Génération d'images

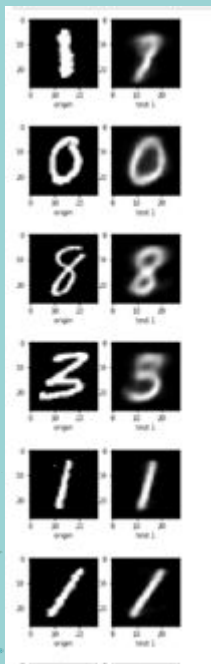
MNIST

SIMPSONS

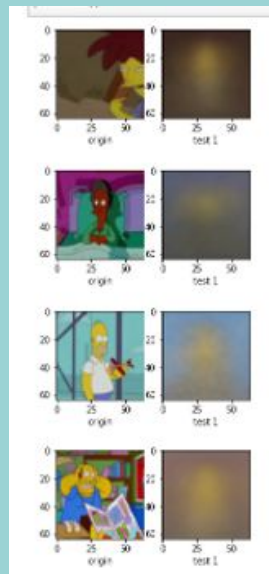


Compression / Decompression

MNIST



SIMPSONS



Meilleur modèle

