

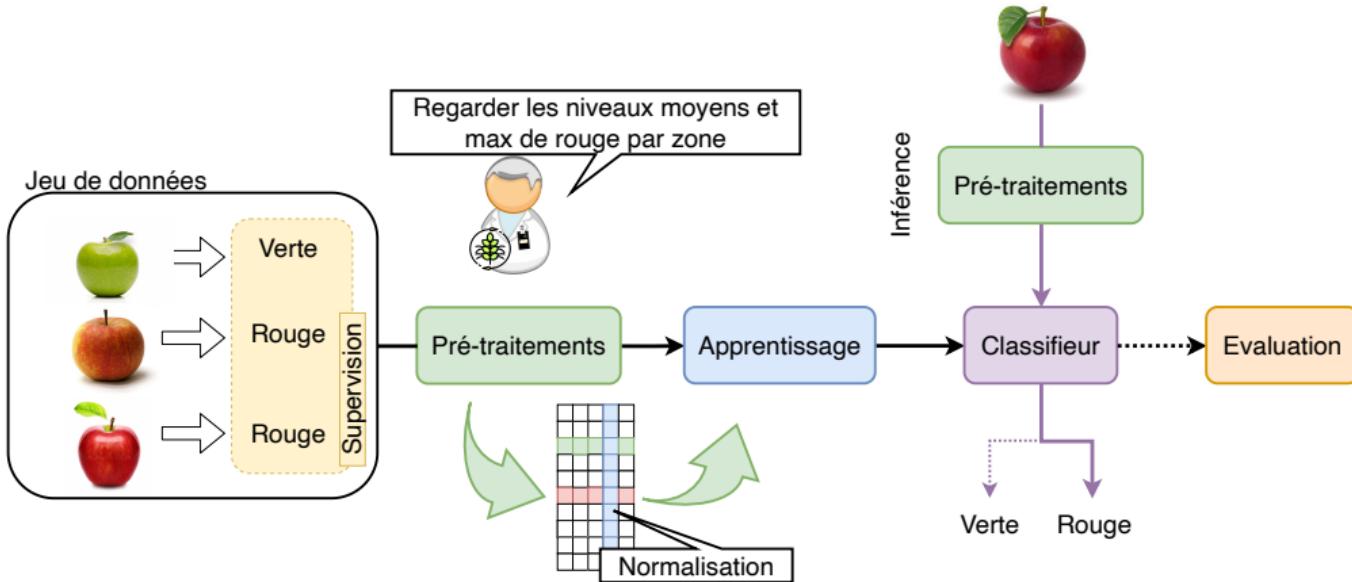
# APPRENTISSAGE DE PRÉSENTATION

Vincent Guigue,  
inspiré des supports de Nicolas Baskiotis & Benjamin Piwowarski



# INTRODUCTION

# Chaine de traitements



Facteur de performances: *features > modèle*

# Promesse du deep-learning

- Apprendre des représentations continues de concepts **discrets** (+métrique continue)
  - Sémantique des **mots**, profils **utilisateurs**
- Apprendre des représentations compactes de concepts **continus**
  - Représentation des images, extraction de motifs dans les signaux

Optimiser les caractéristiques extraites pour une applications cibles

*Deep learning ⇒ Representation Learning*

Pourquoi est ce que ça marche si bien sur des données complexes+sémantiques mais moins bien sur des données tabulaires?

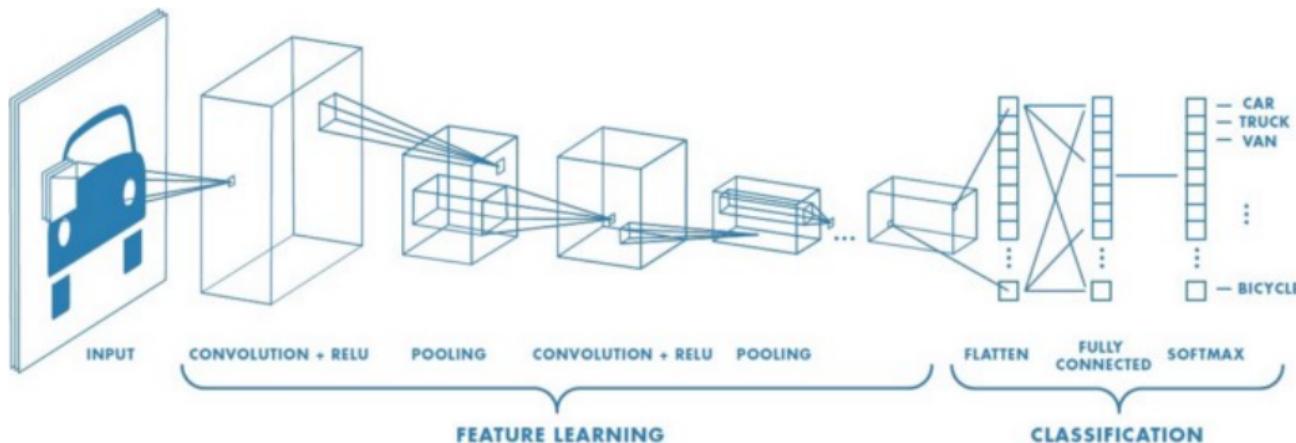


Grinsztajn et al., NeurIPS 22

Why do tree-based models still outperform deep learning on tabular data?

# Une frontière floue entre représentation et décision

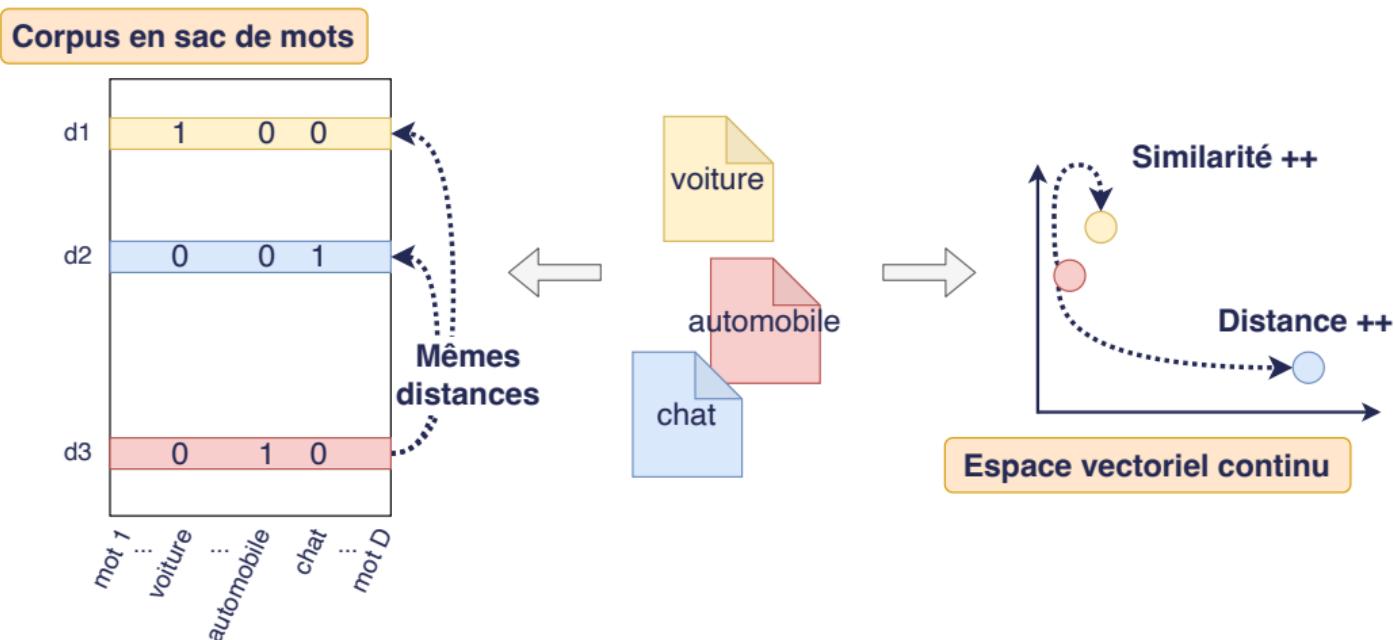
Exemple de la détection d'objets dans les images:



- Dernière couche de caractéristiques idéales =  
$$1 / \text{objet présent} + 0 / \text{objet absent}$$
- Souhaite-t-on vraiment une représentation qui décide?  
ECOC, notion de distance...

# Représentation d'éléments discrets ou continus?

Machine Learning = difficile de représenter des éléments discrets:  
**variables catégorielles, mots, utilisateur**



**Objectif:** introduire la notion de distance  $\neq$  orthogonalité entre les catégories

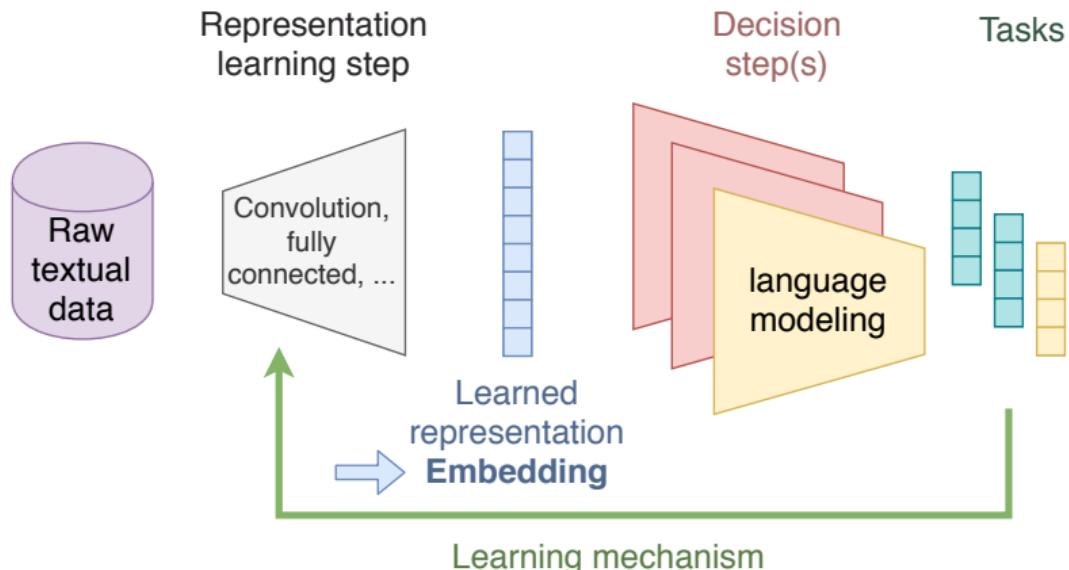
# Representation learning & transfer

## 1 Apprendre la représentation

sur n'importe quel critère

## 2 Fine-tuner = apprendre un peu

Sur la tâche finale / sur peu de données etc...

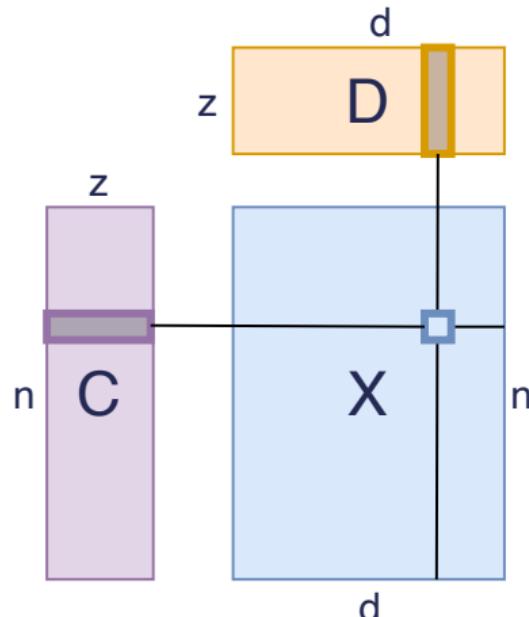


# ARCHITECTURES APPRENTISSAGE REPRÉSENTATION DE



# Apprentissage de représentation

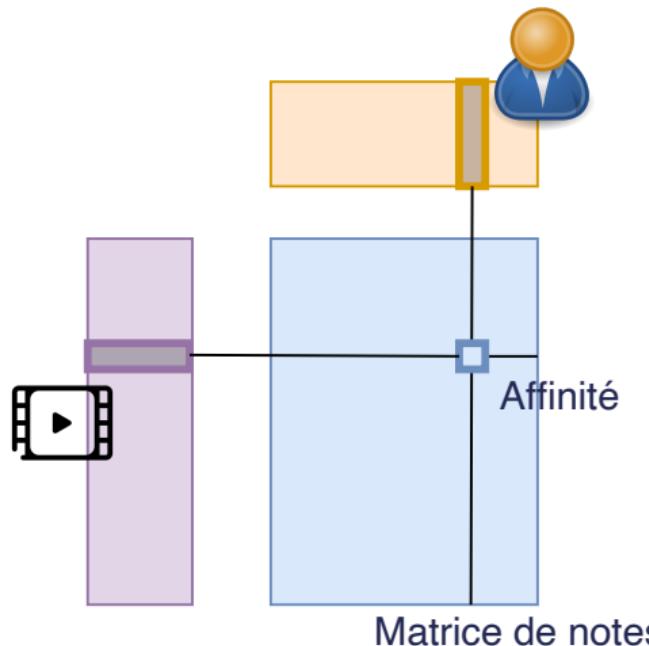
- Domaines : recommandation, représentation de graphes, classification multi-labels multi-classes, séparation sources audio, radar
- Appris en **end-to-end**, **pré-appris** et/ou **fine-tuné**
- Parfait pour la supervision est indirecte : similarité par triplet, contexte, ...





# Apprentissage de représentation

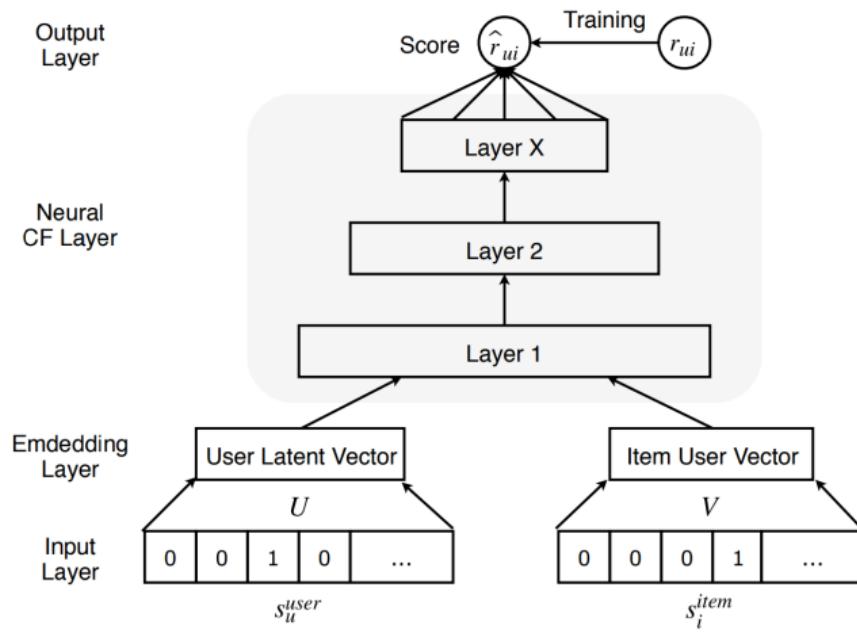
- Domaines : recommandation, représentation de graphes, classification multi-labels multi-classes, séparation sources audio, radar
- Appris en **end-to-end**, **pré-appris** et/ou **fine-tuné**
- Parfait pour la supervision est indirecte : similarité par triplet, contexte, ...





# Apprentissage de représentation

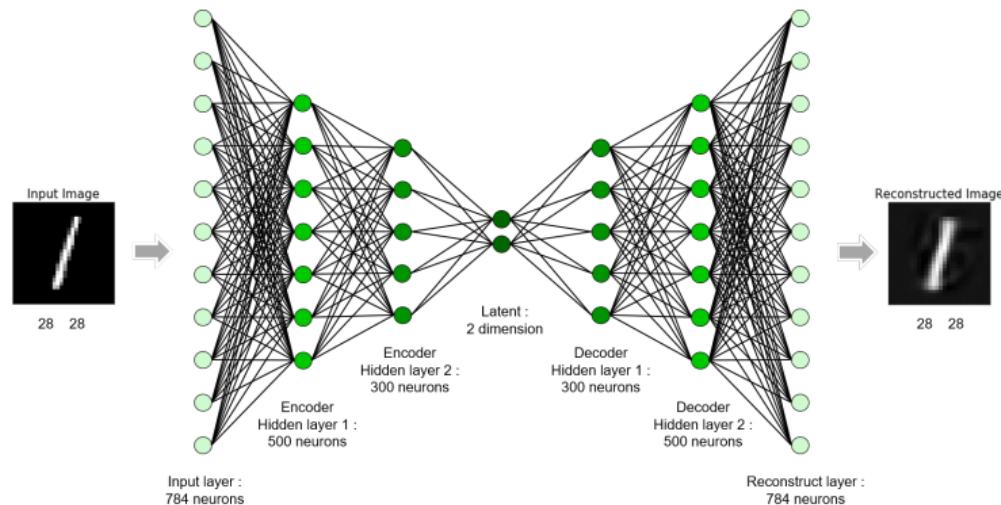
- Domaines : recommandation, représentation de graphes, classification multi-labels multi-classes, séparation sources audio, radar
- Appris en **end-to-end**, **pré-appris** et/ou **fine-tuné**
- Parfait pour la supervision est indirecte : similarité par triplet, contexte, ...





# Apprentissage de représentation

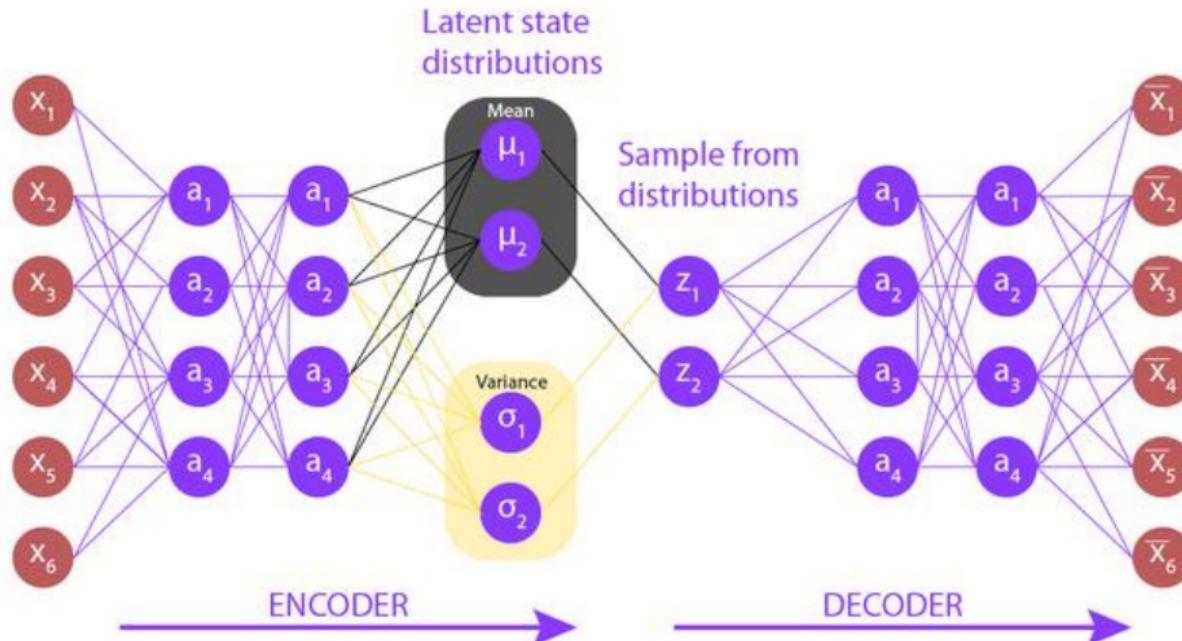
- Domaines : recommandation, représentation de graphes, classification multi-labels multi-classes, séparation sources audio, radar
- Appris en **end-to-end**, **pré-appris** et/ou **fine-tuné**
- Parfait pour la supervision est indirecte : similarité par triplet, contexte, ...
- Possibilité d'utiliser des auto-encoders





# Structurer l'espace appris

- Variational AutoEncoder
- Reparameterization Trick

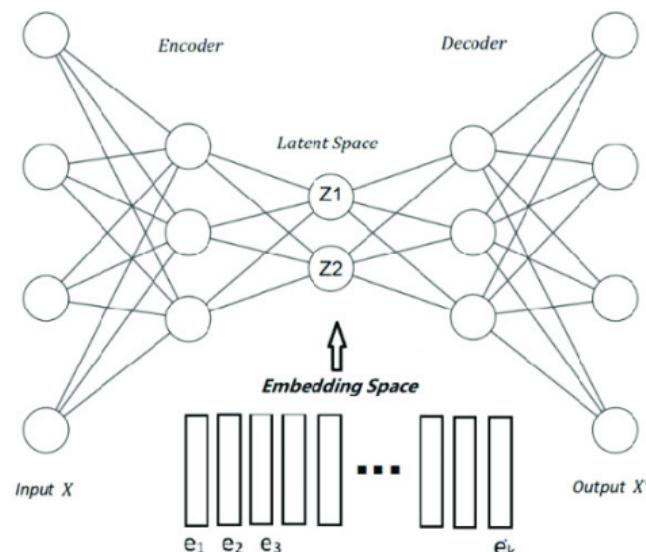




# Représentation & explicabilité

Hypothèse: expliquer un processus = décomposer sur un dictionnaire *réduit* & *intelligible*

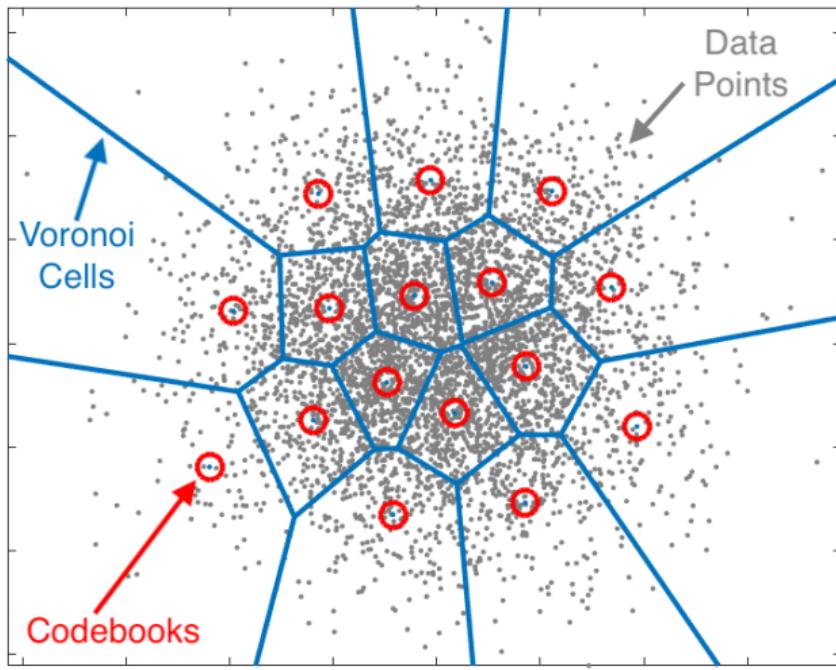
Input  
Discret/Continu → Repr. latente  
Continu ⇒ distance → dictionnaire  
Discret





# Deep LVQ

- LVQ = Learning vector quantization
- Idée ~ plus proche voisin
- + deep learning = apprendre la représentation des supports



# APPLICATIONS (EFFICACES)



# Deep learning et données textuelles

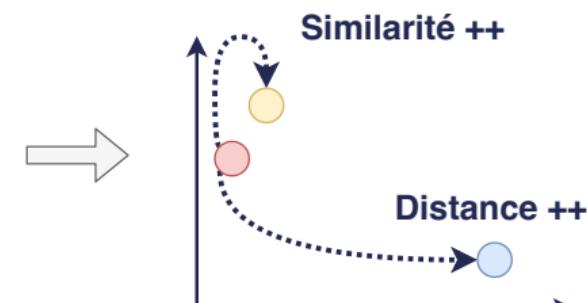
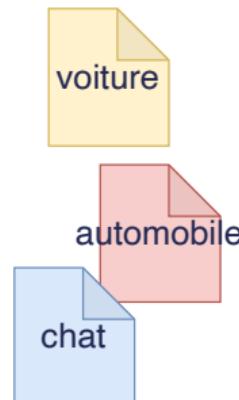
## 1. Des sacs de mots aux représentations vectorielles

[2008, 2013, 2016]

### Corpus en sac de mots

d1	1	0	0
d2	0	0	1
d3	0	1	0
mot <sub>1</sub>	⋮	⋮	⋮
voiture	⋮	⋮	⋮
automobile	⋮	⋮	⋮
chat	⋮	⋮	⋮
mot <sub>D</sub>	⋮	⋮	⋮

Mêmes distances



Espace vectoriel continu



# Deep learning et données textuelles

## 1. Des sacs de mots aux représentations vectorielles

[2008, 2013, 2016]

### Comment apprendre un tel espace?

he curtains open and the moon shining in on the barely  
 ars and the cold , close moon " . And neither of the w  
 rough the night with the moon shining so brightly , it  
 made in the light of the moon . It all boils down , wr  
 surely under a crescent moon , thrilled by ice-white  
 sun , the seasons of the moon ? Home , alone , Jay pla  
 m is dazzling snow , the moon has risen full and cold  
 un and the temple of the moon , driving out of the hug  
 in the dark and now the moon rises , full and amber a  
 bird on the shape of the moon over the trees in front  
 But I could n't see the moon or the stars , only the  
 rning , with a sliver of moon hanging among the stars  
 they love the sun , the moon and the stars . None of  
 the light of an enormous moon . Theplash of flowing w  
 man 's first step on the moon ; various exhibits , aer  
 the inevitable piece of moon rock . Housing The Airsh  
 oud obscured part of the moon . The Allied guns behind



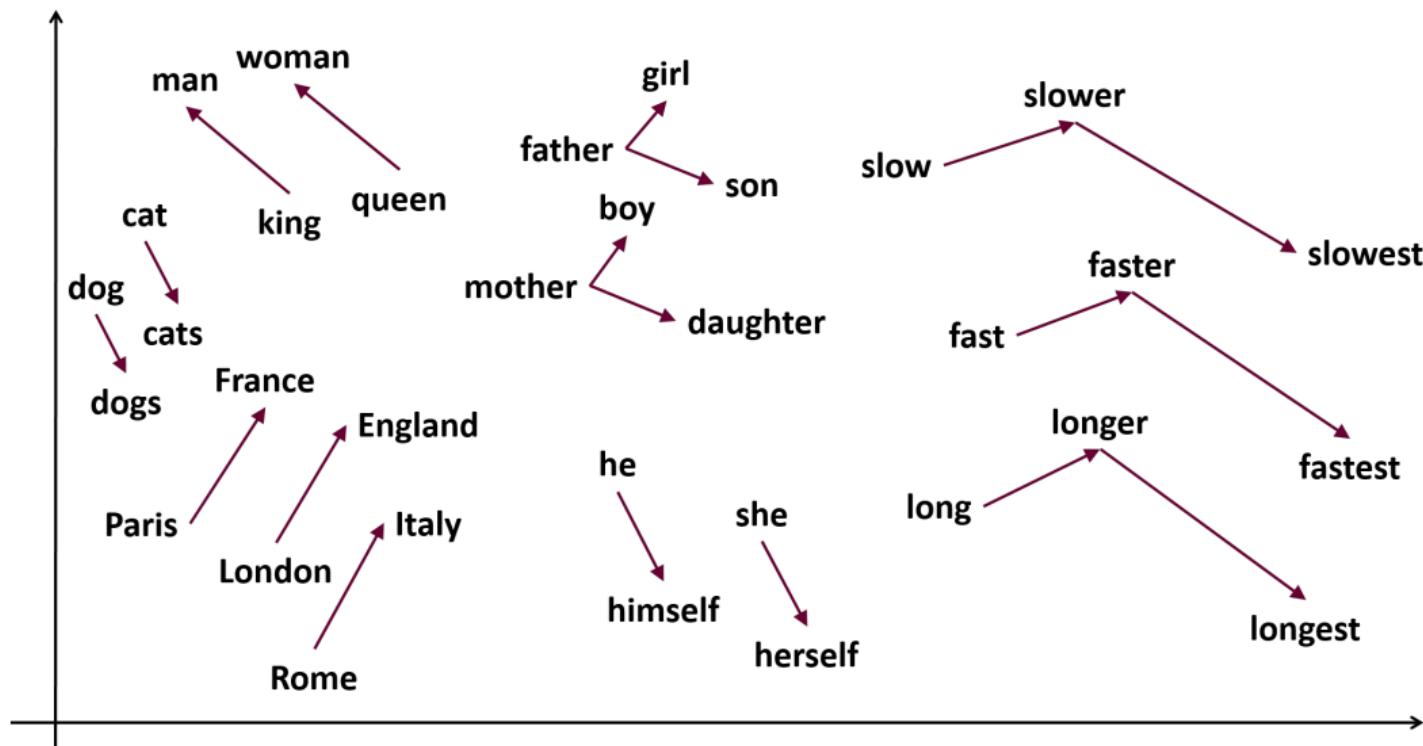
Observations massives ⇒  
 succession de petits  
 mouvements



# Deep learning et données textuelles

## 1. Des sacs de mots aux représentations vectorielles

[2008, 2013, 2016]



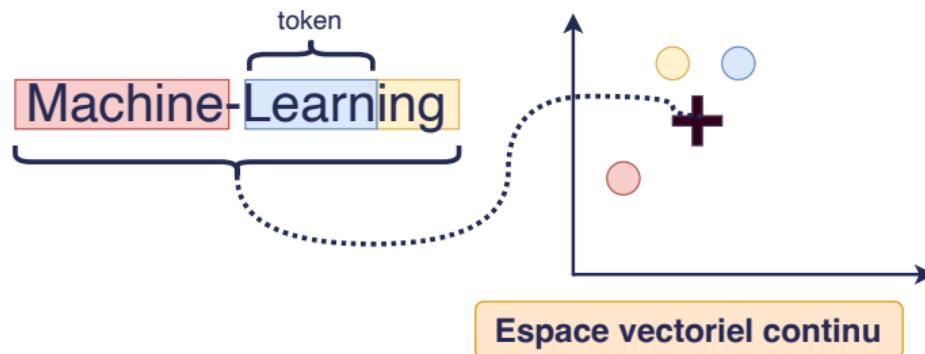
# Deep learning et données textuelles

## 1. Des sacs de mots aux représentations vectorielles

[2008, 2013, 2016]

### Des mots aux tokens

Décomposition en groupes  
de lettres fréquentes



- Représentation des mots inconnus
- Résistance aux fautes d'orthographe



# Deep learning et données textuelles

## 2. Agrégation des représentations de mots

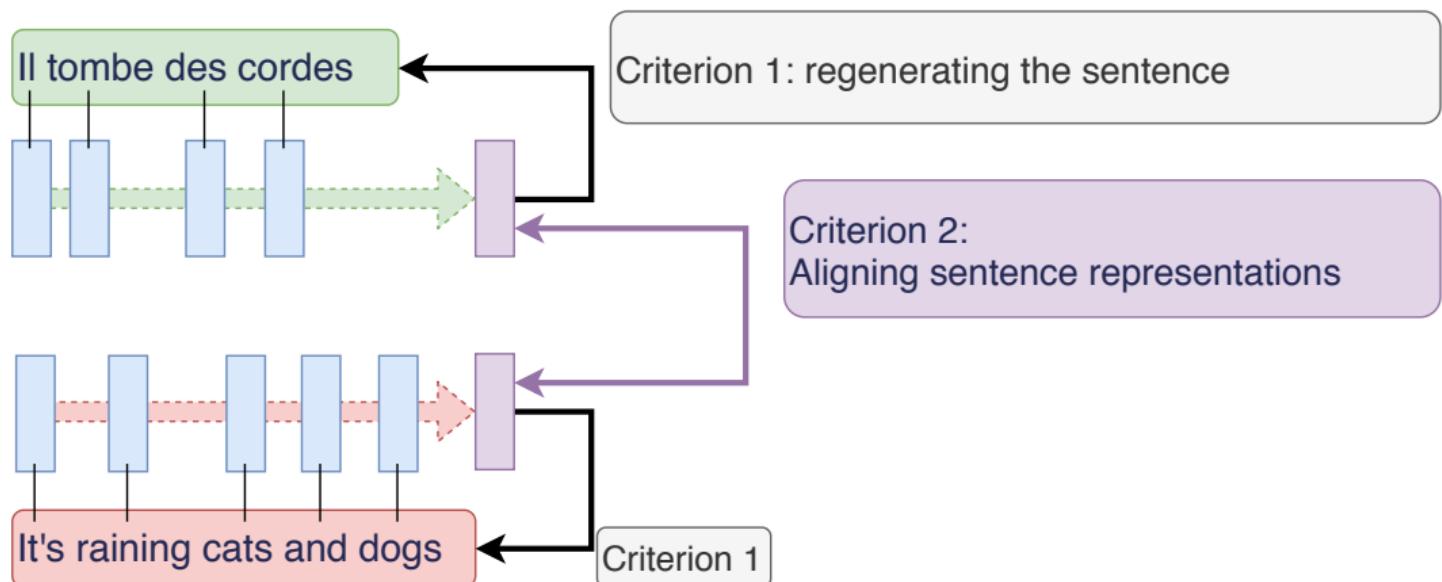
[2014, 2017]



# Deep learning et données textuelles

## 2. Agrégation des représentations de mots

[2014, 2017]

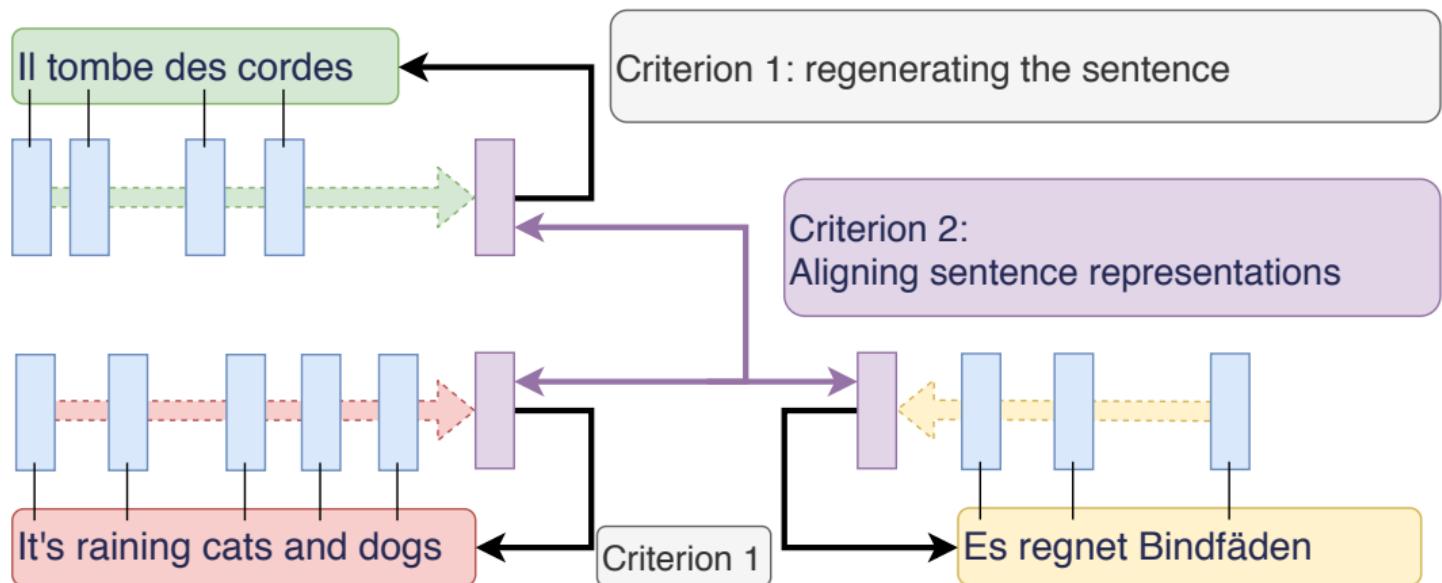




# Deep learning et données textuelles

## 2. Agrégation des représentations de mots

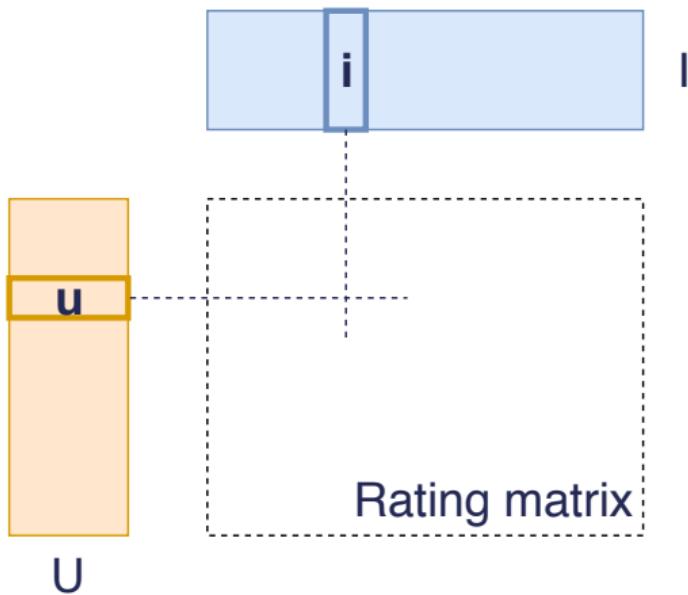
[2014, 2017]





# Recommandation

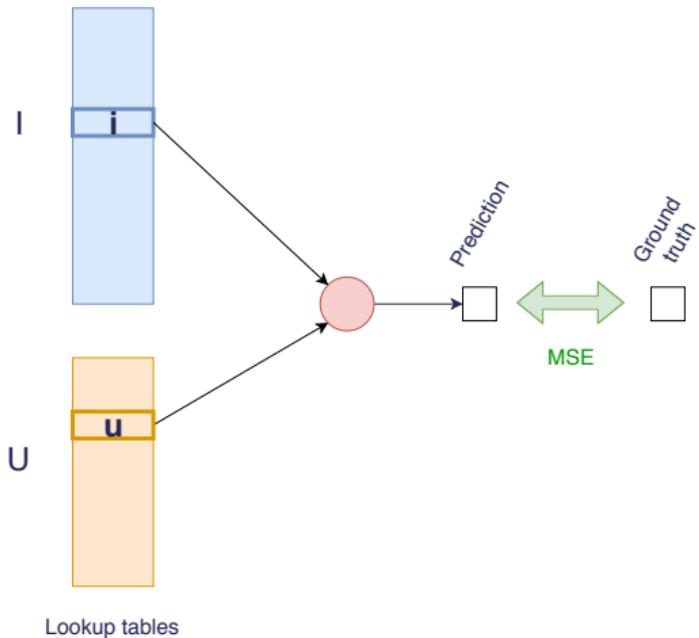
- 2001-2010 : Apprendre des profils à partir de trace = factorisation matricielle
- Déjà un modèle de deep-learning?



Profils se ressemblent  $\Rightarrow$  prédisent les mêmes notes

# Recommandation

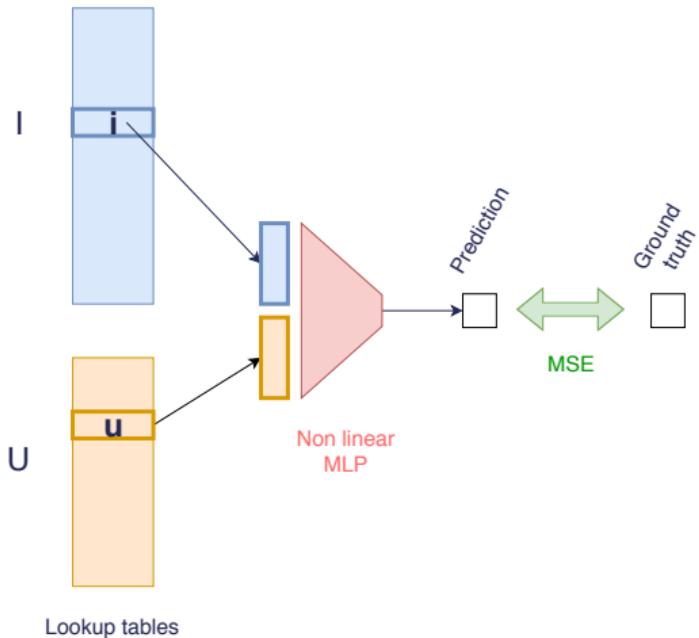
- 2001-2010 : Apprendre des profils à partir de trace = factorisation matricielle
- Déjà un modèle de deep-learning?



Profils se ressemblent  $\Rightarrow$  prédisent les mêmes notes

# Recommandation

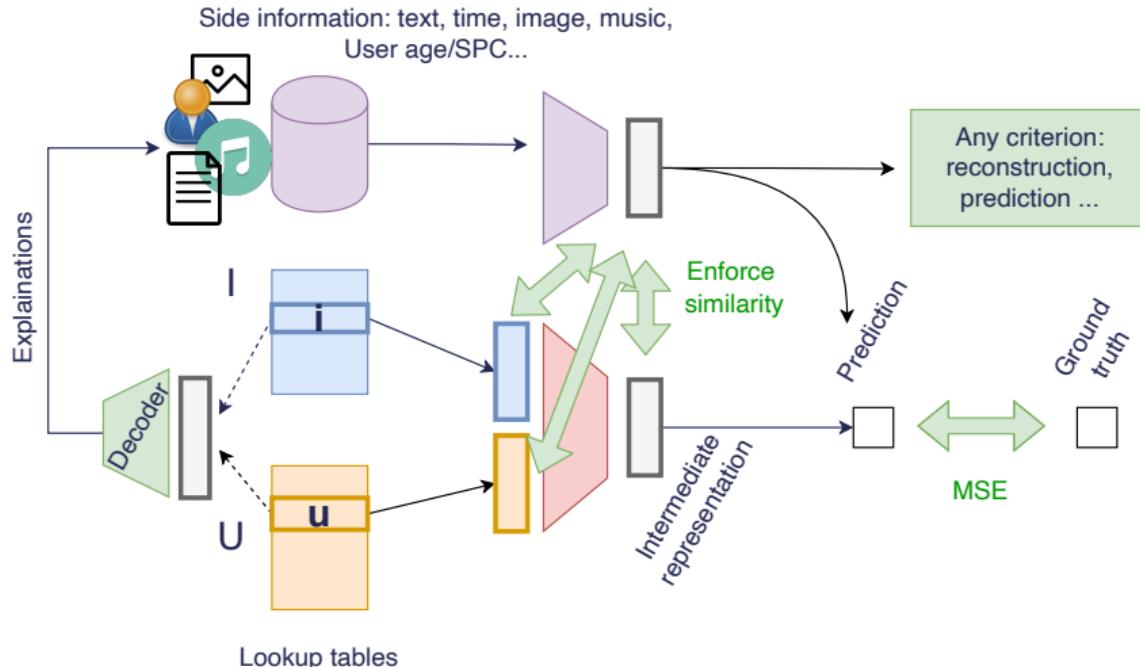
- 2001-2010 : Apprendre des profils à partir de trace = factorisation matricielle
- Déjà un modèle de deep-learning?



Profils se ressemblent  $\Rightarrow$  prédisent les mêmes notes



# Cas des systèmes de recommandation



- Richesse des profils = performances
- Expliquer le contenu des profils = xAI

[encodage ++]  
[décodage ++]

# Sortir de la boite noire

- Expliquer une décision (Instance Feature Attribution)
    - LIME, SHAP,...
  - Espace latent = boite noire...
    - 1 **Structurer** l'espace: VAE, GAN
    - 2 **Explorer** l'espace: counterfactual
    - 3 **Décoder** l'espace: approche générative
- Que faire?
- ⇒ + Prototypes
- ⇒ Montrer & générer des exemples
- ⇒ Parler à l'utilisateur

## Sortir de la boîte noire

- Expliquer une décision (Instance Feature Attribution)
    - LIME, SHAP,...

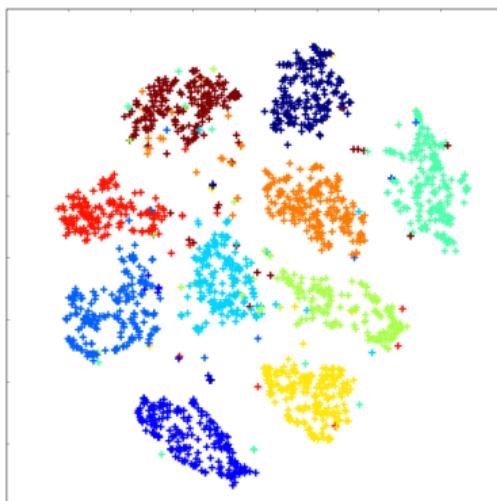
- ### ■ Espace latent = boite noire...

## Que faire?

⇒ + Prototypes

⇒ Montrer & générer des exemples

⇒ Parler à l'utilisateur



# Sortir de la boîte noire

- Expliquer une décision (Instance Feature Attribution)
  - LIME, SHAP, ...

- Espace latent = boîte noire...

1 Structurer l'espace: VAE, GAN

2 Explorer l'espace: counterfactual

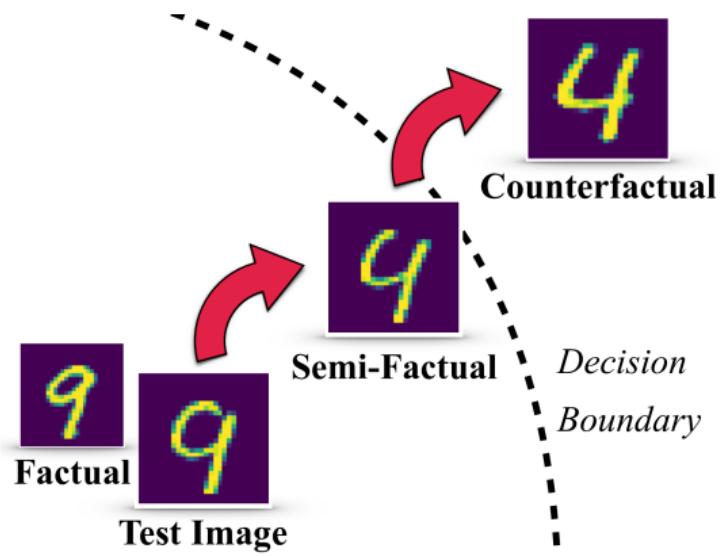
3 Décoder l'espace: approche générative

Que faire?

⇒ + Prototypes

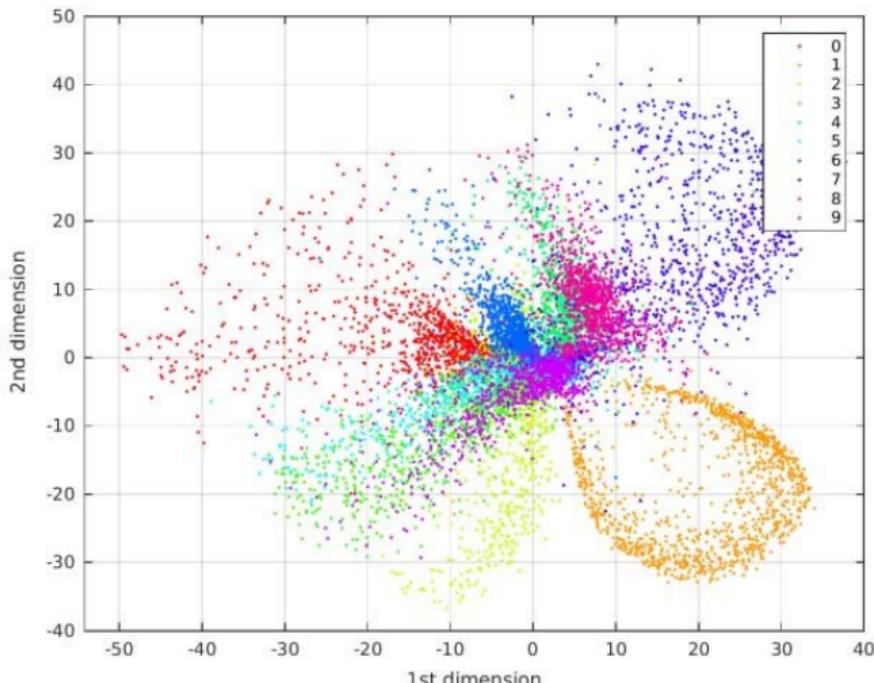
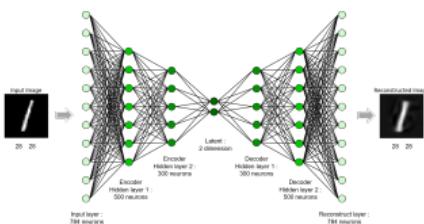
⇒ Montrer & générer des exemples

⇒ Parler à l'utilisateur



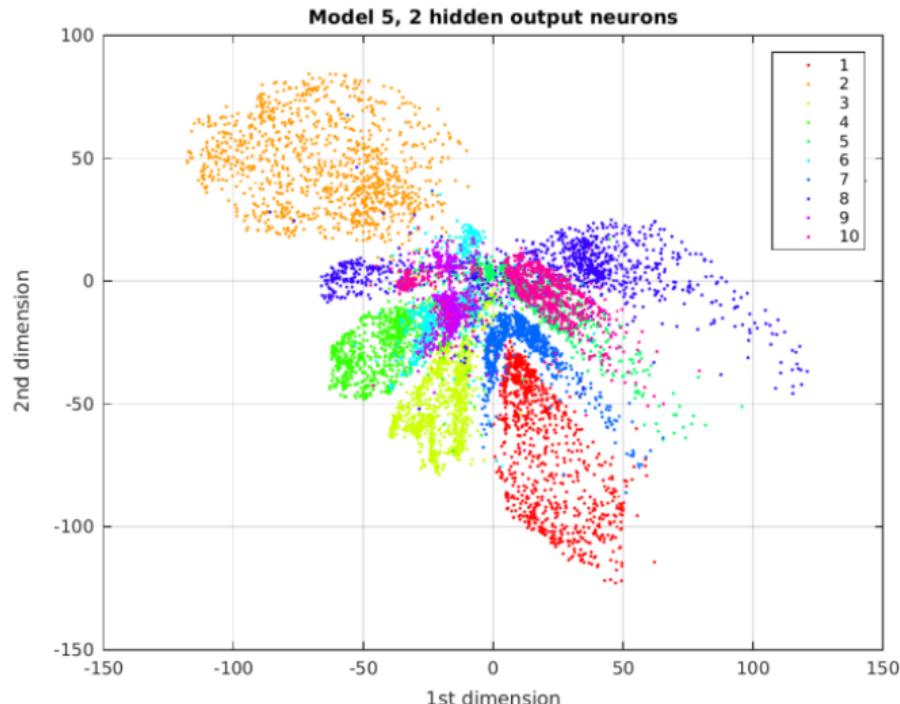
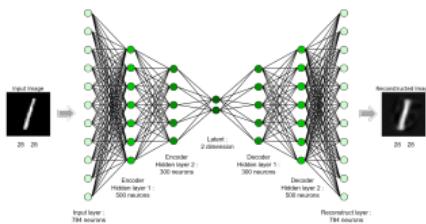
# Visualisation de données

- Auto-encodeur  $\Rightarrow$  2D = visualisation de MNIST
- Passage au VAE (toujours non supervisé)



# Visualisation de données

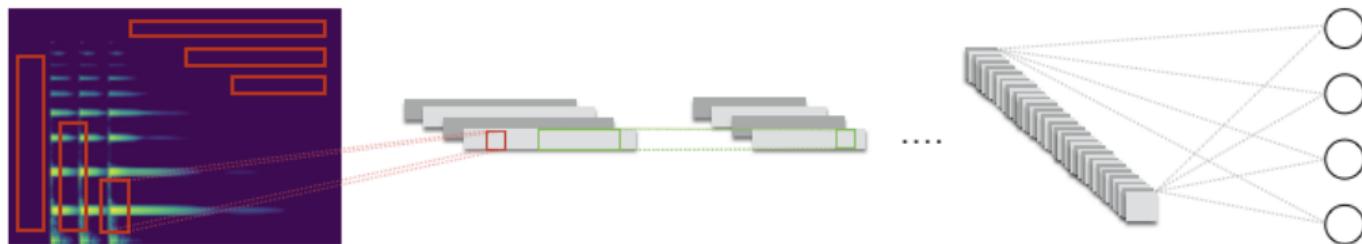
- Auto-encodeur  $\Rightarrow$  2D = visualisation de MNIST
- Passage au VAE (toujours non supervisé)





# MusiCNN

- Convolution sur la transformée temps-fréquences
- Apprentissage d'embedding de musique
  - Pour la classification / recommandation
  - Pour la génération



# Décomposition de signaux

VQ-VAE-GAN : Reconstruire une entrée...

- A partir d'un ensemble de motifs finis
- En la rendant indiscernable de son originaire !

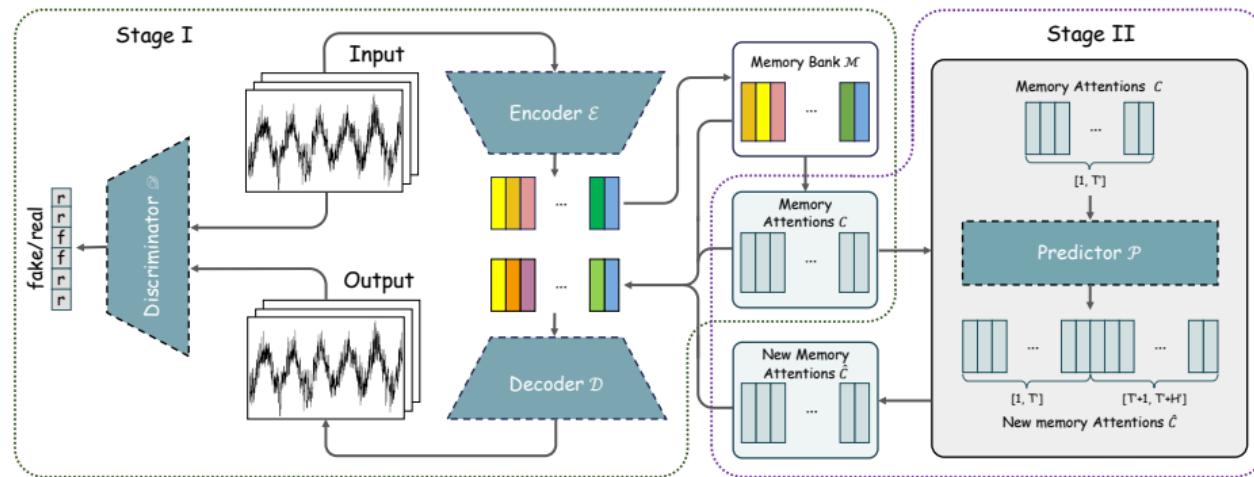


Figure 1: Overview of MATS. The left half part shows the first stage components and the right half part show the second stage components.

# OUTILS CONNEXES POUR LE TP



# Représentations en pytorch

Représentation = vecteur + gradient activé

Hypothèse: dans un problème de recommandation, je modélise les notes par 3 scalaires

- composante globale, composante utilisateur, composante item

$$r_{ui} = \mu + b_u + b_i$$

- Initialiser les *représentations* à apprendre

---

```
1 mu = torch.tensor([3.5], requires_grad=True) # activate gradient to be able to
2 bu = [torch.tensor([0.1], requires_grad=True) for _ in range(num_users)]
3 bi = [torch.tensor([0.1], requires_grad=True) for _ in range(num_items)]
```

---

- Utiliser ces grandeurs pour la prédiction
- Ajouter une fonction de coût (e.g. MSE)
- Retro-propager l'erreur



# Module de représentation

Le module existe en pytorch, il est idéal pour représenter les structures discrètes.

---

```
1 latent_size = 10
2 nb_users = 100
3 users = torch.nn.Embedding(nb_users, latent_size) # random init
4
5 # get representation of user 5:
6 print("User 5:", users(torch.tensor(5)))
7 # WARNING: call for a tensor (not an int)
```

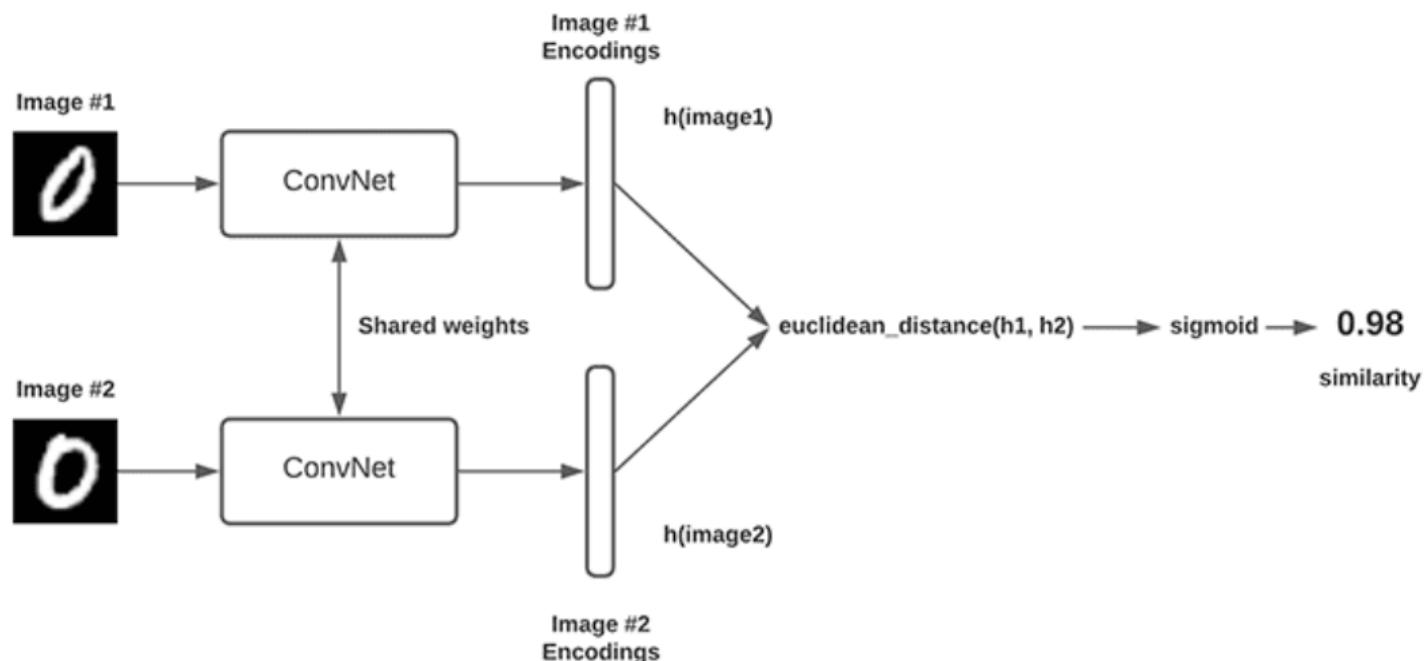
---



# Triplet loss

- Rapprocher ce qui se ressemble... (+ Eloigner autre chose)

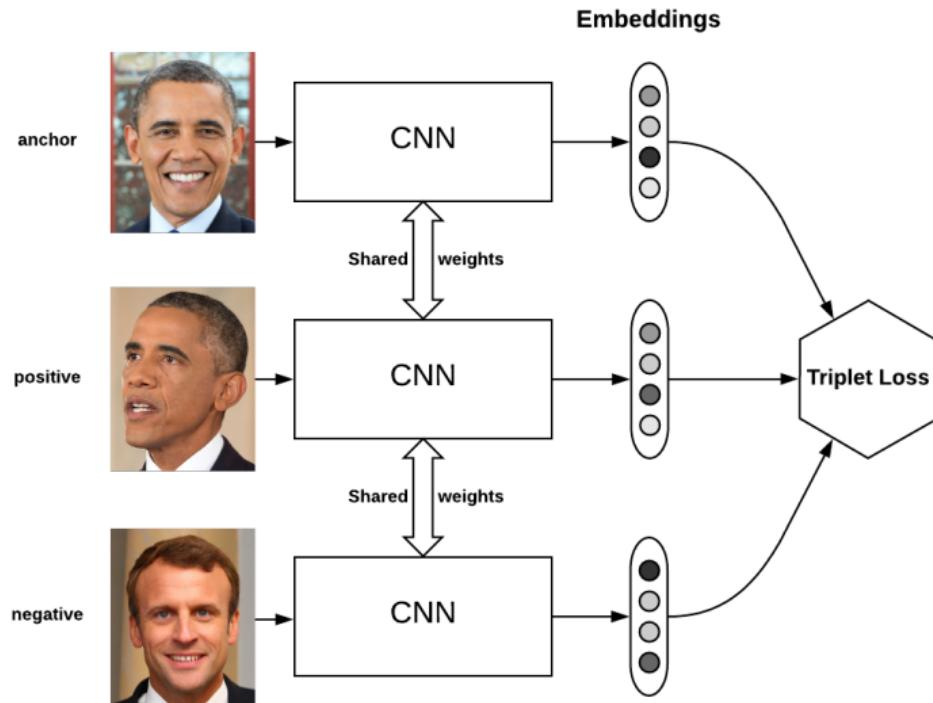
A l'ancienne: Siamese network





# Triplet loss

- Rapprocher ce qui se ressemble... (+ Eloigner autre chose)



# CONCLUSION



# Evolution scientifique & pratique

80-90' Partage d'algorithmes et de données

2000' Partage d'implémentation de classifieur (SVM light, liblinear)

2010' Données massives + Plateforme de ML

& partage de modèle appris (Torch, W2V, ...)

2020' Généralisation de github, tensorflow, pytorch

+ interfaces d'accès e.g. HuggingFace

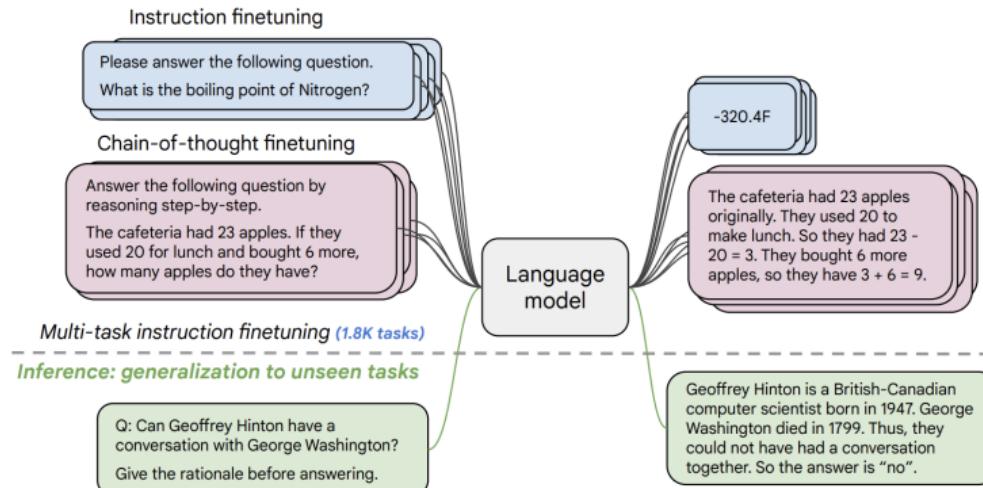
- Le plus grand levier de performances sur les 10 dernières années
- Très efficace sur les données sémantiques disponibles massivement
- Des modèles + complexes ET + accessibles

En texte, en image... Mais pas encore sur la classification de signaux

# Modèle pré-appris/ Multi-tâches / Tranfert / Fine-tuning

## Que faire maintenant?

- Plus de données (GPT2, 3, etc...),
- Plus de modalité (Grounding, visual language model)
- Plus de transfert (nouvelles tâches)

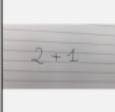




# Modèle pré-appris/ Multi-tâches / Tranfert / Fine-tuning

## Que faire maintenant?

- Plus de données (GPT2, 3, etc...),
- Plus de modalité (Grounding, visual language model)
- Plus de transfert (nouvelles tâches)

Input Prompt				Completion
	This is a chinchilla. They are mainly found in Chile.		This is a shiba. They are very popular in Japan.	 This is → a flamingo. They are found in the Caribbean and South America.
	What is the title of this painting? Answer: The Hallucinogenic Toreador.		Where is this painting displayed? Answer: Louvres Museum, Paris.	 What is the name of the city where this was painted? Answer: → Arles.
	Output: "Underground"		Output: "Congress"	 Output: → "Soulomes"
	$2+1=3$		$5+6=11$	 → $3 \times 6 = 18$

# Quelle perspectives pour notre recherche

- 1 + de données, + de complexité
  - Difficile de jouer contre les GAFAM
- 2 Applications originales / implémentation de contraintes métiers
  - développement de modèles
- 3 Recherche de niches à fort potentiel
  - nouvelles problématiques: data-to-text, recommandation explicable
- 4 Recherche de propriétés théoriques
  - optimisation, prior sur l'espace latent (Nesterov, VAE, ...)
  - complémentarité intéressante mais complexe entre théorie et pratique

## Niveaux d'usage des architectures deep:

- **Utilisateur** de modèles pré-entraînés
- **Affineur** sur de petit jeu de données, domaine applicatifs diverses
- **Entraîneur** sur des données jouets (démonstration de concepts) / données massives