

# Deep Learning par la Pratique

Reccurent Neural Networks et Autoencodeurs

---

GIRAUD François-Marie



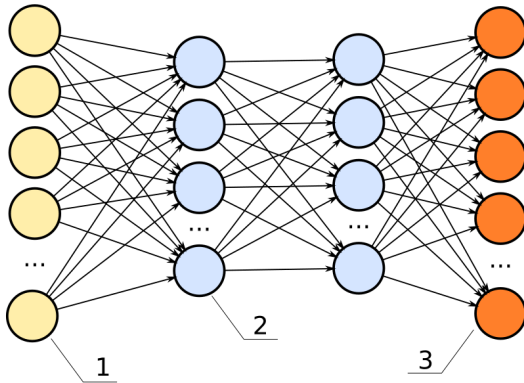
<https://www.orsys.fr/>

# Deep Learning par la Pratique

Cours 3 : Recurrent Neural Networks et Autoencodeurs

---

# Rappels



# Deep Learning par la Pratique

Les Données Séquentielles

---

## **Données ancrées dans le temps**

Séquences de données à dimension fixe

# Les Données Séquentielles

Soit  $X = (x_i)_{1 \leq i \leq k}$  un ensemble de  $k$  exemples :

où :

$x_i = (x_i^1, \dots, x_i^{n_i})$  avec  $x_i$  une séquence de  $n_i$  frames

Pour des séries d'entiers par exemple :

$$x_1 = (1, 3, 5, 2, 8)$$

$$x_2 = (7, 3)$$

$$x_3 = (4, 0, 9, 1)$$

$$x_4 = (\dots)$$

Pour des séries de vecteurs :

$$x_1 = ([2, 5], [9, 8], [3, 6])$$

$$x_2 = ([1, 1], [3, 4], [5, 4], [3, 2], [8, 1])$$

$$x_3 = (\dots)$$

## Texte

1 : lorem ipsum dolor sit amet, consectetur adipiscing elit, ...

2 : Great minds discuss ideas ; average minds discuss events ; small minds discuss people.

3 : Hier, mon voisin a mangé une pomme et sa femme une poire.

4 : ...

## A.D.N

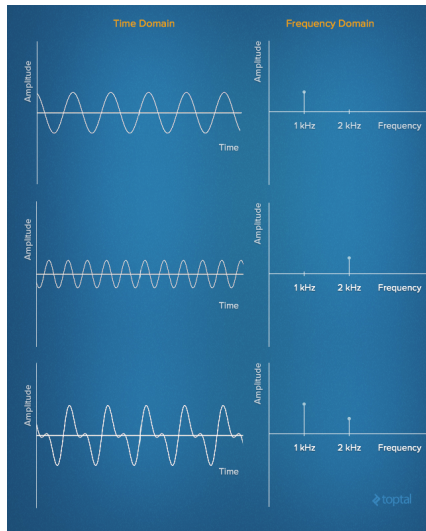
1 : ATGCGATCTATCGCTAGCCGCGCTATACGCA

2 : GATTATAGCTAGCTCGCGCTATATCGCTAGCTAGCTAGCTAGC

3 : ...



## Son



## Video



- Météo (time, Lat., Long.,  $T^\circ$ , P,  $X_{vent}$ ,  $Y_{vent}$ ,  $Q_{pluie}$ )
- Santé (time,  $T^\circ$ , Pouls, ...)
- Economie (time, dollar, euro, livre, yen, BTC, riz ...)
- Comportements Clients (time, action-event, ...)
- ...

# Deep Learning par la Pratique

Les Modèles Récursifs

---

Des tâches variées :

- Prédiction d'une classe
- Prédiction de la suite d'une séquence
- Génération de séquence
- Découverte de Patterns

Prédiction d'une classe :

input :  $x_i = (1, 6, 8, 4)$

output : "good" or "bad"

- Son (chant d'oiseau, personne, genre musical, ...)
- Vidéos (film, documentaire, stand up , ...)
- ...

Certain modèles permettent de clusteriser (non-supervisée)

Prédiction de la suite d'une séquence :

input :  $x_i = (1, 2, 3, 4, 5)$

output : 6

- Données Économiques
- Comportement Clients
- Météo
- Modèle de Langage
- ...

Génération :

input :  $x_1 = (1, 3, 5)$ ,  $x_2 = (7, 9, 11)$  output :  $(5, 7, 9)$

- Génération de partitions
- Texte vers Voix
- Voix vers Texte
- ...



Découverte de Patterns :

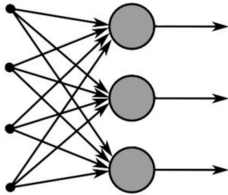
- Découverte de gènes
- Compression de signal
- Décrire et comprendre des phénomènes
- ...

# Deep Learning par la Pratique

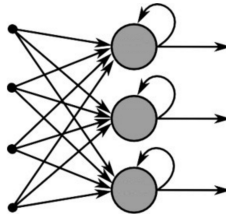
Réseaux de Neurones Récurrents

---

# Réseaux de Neurones Récurents

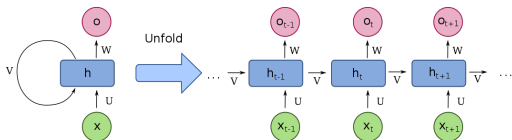


Feed-Forward Neural Network



Recurrent Neural Network

# Réseaux de Neurones Récurrents



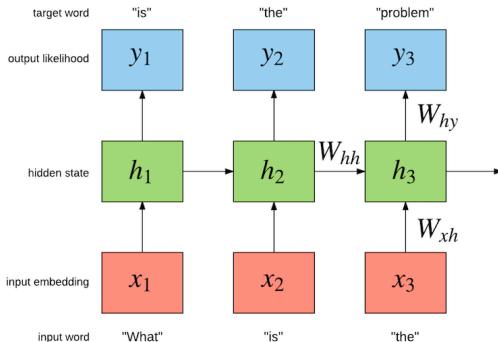
$$h_t = \sigma_h(U * x_t + V * h_{t-1} + b_h)$$

$$o_t = \sigma_o(W * h_t + b_o)$$

- $x_t$  : vecteur d'entrée
- $h_t$  : vecteur de la couche cachée
- $o_t$  : vecteur de sortie
- $U$ ,  $V$ ,  $W$ ,  $b_h$  et  $b_o$  : matrices et vecteurs (paramètres)
- $\sigma_h$  et  $\sigma_o$  : fonctions d'activation (ReLU)

# Réseaux de Neurones Récurrents

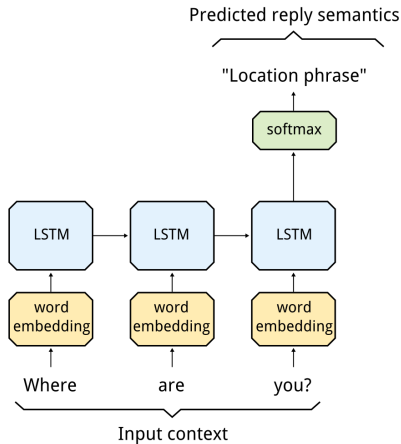
Prédiction de la suite d'une séquence (ex : modèle de langage)



Peut être utilisé pour générer de nouvelles séquences

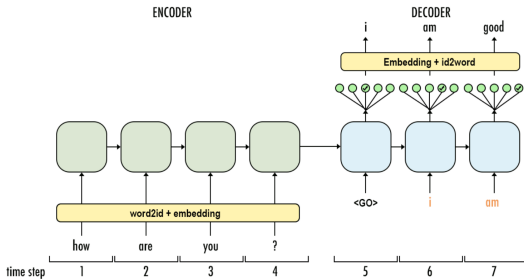
# Réseaux de Neurones Récurrents

Prédiction d'une classe



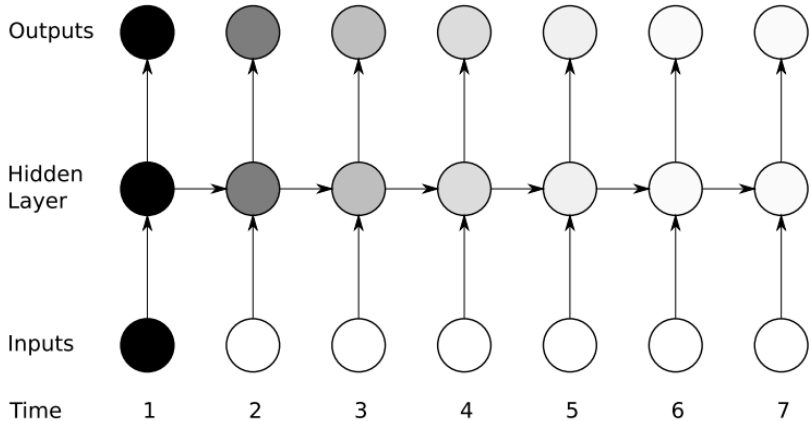
# Réseaux de Neurones Récurrents

Génération d'une séquence (seq2seq)



# Réseaux de Neurones Récurents

Le problème du gradient qui disparaît (vanishing gradient)



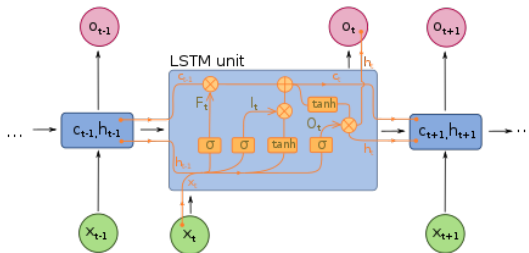


# Deep Learning par la Pratique

Long-Short Term Memory

---

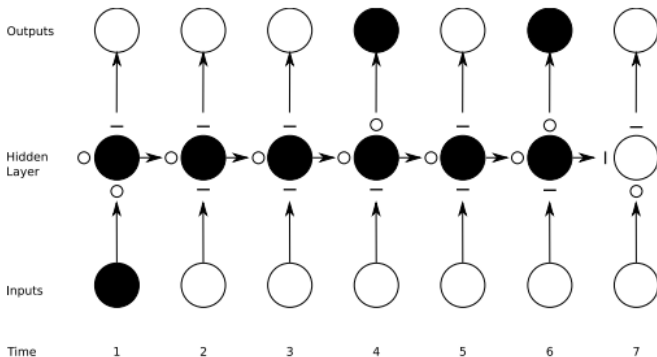
# Long-Short Term Memory



- $F_t = \sigma(W_F * x_t + U_F * h_{t-1} + b_F)$  (forget gate)
- $I_t = \sigma(W_I * x_t + U_I * h_{t-1} + b_I)$  (input gate)
- $O_t = \sigma(W_O * x_t + U_O * h_{t-1} + b_O)$  (output gate)
- $c_t = F_t \odot c_{t-1} + I_t \odot \tanh(W_c * x_t + U_c * h_{t-1} + b_c)$
- $h_t = O_t \odot \tanh(c_t)$
- $o_t = f(W_o * h_t + b_o)$

# Long-Short Term Memory

vanishing gradient “résolu” (ou presque)



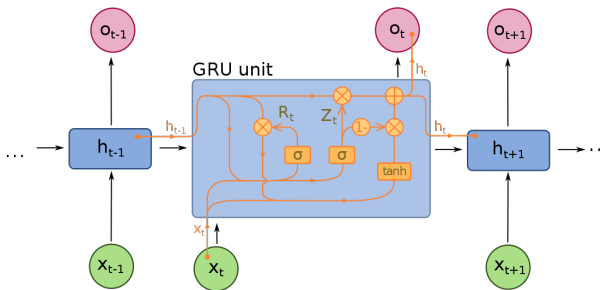
# Deep Learning par la Pratique

D'autres variantes...

---

# D'autres variantes

## Gated Recurrent Unit



$Z_t = \sigma(W_Z * x_t + U_Z * h_{t-1} + b_Z)$  (update gate  $\approx$  input et forget gates)

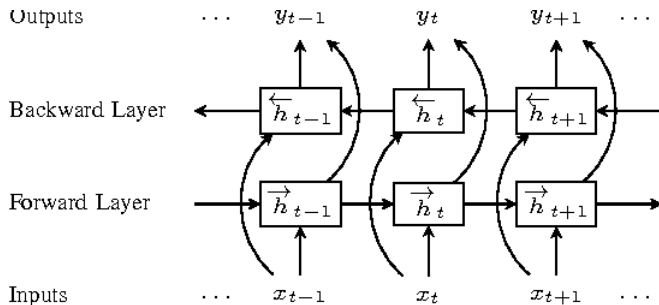
$R_t = \sigma(W_R * x_t + U_R * h_{t-1} + b_R)$  (reset gate  $\approx$  output gate)

$h_t = Z_t \circ h_{t-1} + (1 - Z_t) \circ \tanh(W_h * x_t + U_h(R_t \circ h_{t-1}) + b_h)$

Moins de paramètres que LSTM, aussi efficace dans beaucoup de tâches

## D'autres variantes

### RNN bi-directionnels



Permet de “gérer” des séquences plus longues

Dans des réseaux récurrents profonds, on peut ajouter des récurrences entre les différentes couches...

## D'autres variantes





# Deep Learning par la Pratique

Données à dimension variable : Traitement du langage

---

## Classification :

- thème/genre (gutenberg.org : 57k livres)
- auteur (gutenberg.org : 57k livres)
- sentiment (Kaggle movie review : 222k commentaires rotten tomatoes)
- reconnaissance d'entités nommées (Kaggle Annotated Corpus for NER : 1.3M tags)
- ...

Compréhension :

- Question/réponses (SQUAD : 150k questions)
- Traduction (europarl : 450k phrases alignées)
- ...

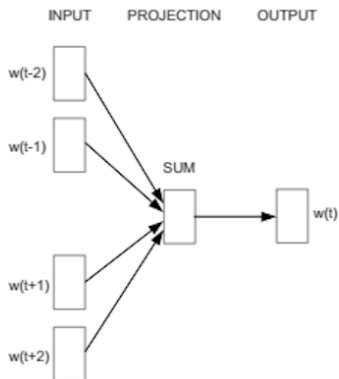
# Traitement du langage : word embeddings

mot = indice dans un dictionnaire (dimension  $> 30000$ )

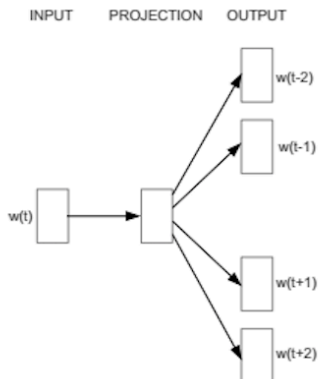
mot = vecteur “sémantique” (dimension  $< 1000$ )

- word2vec
- CBOW/Skip-Gram
- Thought vector (pour des phrases ou même des documents entiers)
- ...

# Traitement du langage : word embeddings

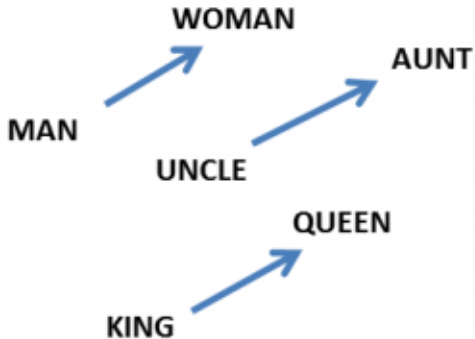


**CBOW**



**Skip-gram**

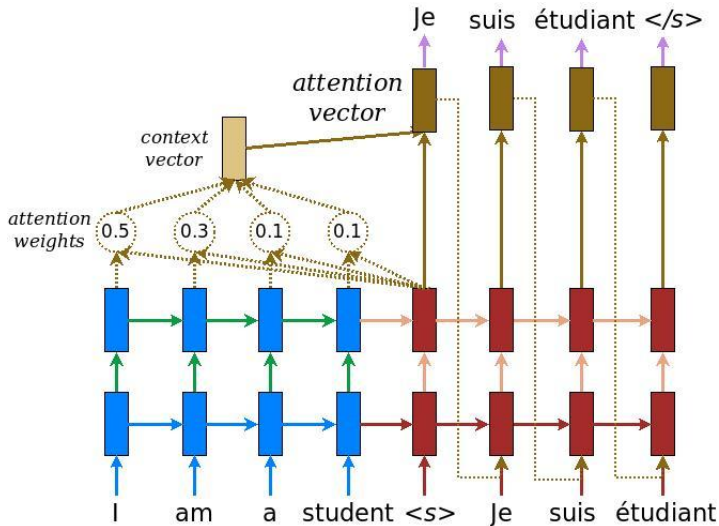
# Traitement du langage : word embeddings



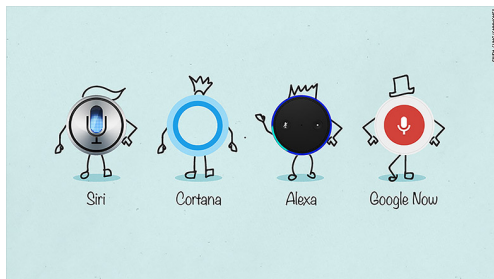
Repository d'embeddings

Visualisation de l'espace word2vec

# Traitement du langage : Modèle à attention



## Transcription et synthèse de la parole

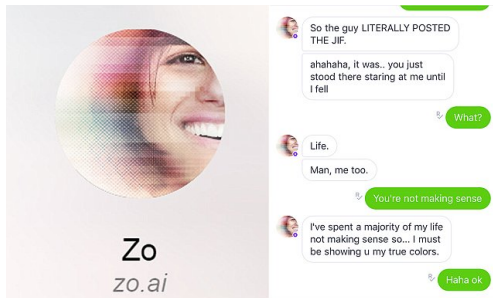




## Identification du locuteur



## Chatbot



## Moteur de recherche

YAHOO!



## Traitement du langage : Champs d'application

## Extraction de données



## Analyse de sentiments



## Résumé

<b>Input: Article 1st sentence</b>	<b>Model-written headline</b>
metro-goldwyn-mayer reported a third-quarter net loss of dlrs 16 million due mainly to the effect of accounting rules adopted this year	mgm reports 16 million net loss on higher revenue
starting from july 1, the island province of hainan in southern china will implement strict market access control on all incoming livestock and animal products to prevent the possible spread of epidemic diseases	hainan to curb spread of diseases
australian wine exports hit a record 52.1 million liters worth 260 million dollars (143 million us) in september, the government statistics office reported on monday	australian wine exports hit record high in september

## Traduction

Traduire **français** (langue identifiée) ▾

Bonjour,

Nos filles souhaitent faire un échange ensemble dans le cadre du programme Brigitte Sauzay. Avant de passer au dossier administratif, nous souhaitons vous poser quelques questions pour s'assurer que nous envisagions cet échange entre nos filles de la même façon.

Le professeur d'allemand de Gaïa a gentiment accepté de traduire nos questions afin d'éviter des incompréhensions à cause de problèmes de traduction.

Concernant les dates, Emma nous a dit que vous souhaitiez que Gaïa soit repartie avant le 19 juin car vous partiez en vacances.

Traduire en **anglais** ▾

Hello,

Our daughters would like to do an exchange together as part of the Brigitte Sauzay program. Before moving on to the administrative file, we wanted to ask you a few questions to ensure that we consider this exchange between our daughters in the same way.

Gaïa's German teacher kindly agreed to translate our questions in order to avoid misunderstandings due to translation problems.

Regarding the dates, Emma told us that you wanted Gaïa to leave before June 19 because you were going on vacation.



## Language Technology

making good progress

mostly solved

### Spam detection

Let's go to Agra!

Buy VIAGRA ...



### Part-of-speech (POS) tagging

ADJ ADJ NOUN VERB ADV

Colorless green ideas sleep furiously.

### Named entity recognition (NER)

PERSON ORG LOC

Einstein met with UN officials in Princeton

### Sentiment analysis

Best roast chicken in San Francisco!



The waiter ignored us for 20 minutes.



### Coreference resolution

Carter told Mubarak he shouldn't run again.

### Word sense disambiguation (WSD)

I need new batteries for my *mouse*.



### Parsing

I can see Alcatraz from the window!

### Machine translation (MT)

第13届上海国际电影节开幕...



The 13<sup>th</sup> Shanghai International Film Festival...

### Information extraction (IE)

You're invited to our dinner party, Friday May 27 at 8:30



still really hard

### Question answering (QA)

Q. How effective is ibuprofen in reducing fever in patients with acute febrile illness?

### Paraphrase

XYZ acquired ABC yesterday

ABC has been taken over by XYZ

### Summarization

The Dow Jones is up

The S&P500 jumped

Housing prices rose

Economy is good

### Dialog

Where is Citizen Kane playing in SF?

Castro Theatre at 7:30. Do you want a ticket?

≈ 2015



# Traitement du langage : Champs d'application

## mostly solved

### Spam detection

Let's go to Agra!



Buy VIAGRA ...



### Part-of-speech (POS) tagging

ADJ ADJ NOUN VERB ADV

Colorless green ideas sleep furiously.

### Named entity recognition (NER)

PERSON ORG LOC

Einstein met with UN officials in Princeton

### Word sense disambiguation (WSD)

I need new batteries for my *mouse*.



### Machine translation (MT)

第13届上海国际电影节开幕...



The 13<sup>th</sup> Shanghai International Film Festival...

### Information extraction (IE)

You're invited to our dinner party, Friday May 27 at 8:30



## making good progress

### Sentiment analysis

Best roast chicken in San Francisco!



The waiter ignored us for 20 minutes.



### Coreference resolution

Carter told Mubarak he shouldn't run again.

### Parsing

I can see Alcatraz from the window!

### Question answering (QA)

Q. How effective is ibuprofen in reducing fever in patients with acute febrile illness?

### Paraphrase

XYZ acquired ABC yesterday

ABC has been taken over by XYZ

### Summarization

The Dow Jones is up

The S&P500 jumped

Housing prices rose

Economy is good

## still really hard

### Dialog

Where is Citizen Kane playing in SF?



Castro Theatre at 7:30. Do you want a ticket?



# Deep Learning par la Pratique

RNN avec Keras

---

# RNN avec Keras

```
1 model = tf.keras.Sequential()
2 model.add(layers.Embedding(input_dim=1000, output_dim=64))
3 model.add(layers.SimpleRNN(128))
4 model.add(layers.Dense(10, activation='softmax'))
5 model.summary()
```

Layer (type)	Output Shape	Param #
embedding_1 (Embedding)	(None, None, 64)	64000
simple_rnn (SimpleRNN)	(None, 128)	24704
dense_1 (Dense)	(None, 10)	1290
Total params: 89,994		
Trainable params: 89,994		
Non-trainable params: 0		

# LSTM avec Keras

```
1 model = tf.keras.Sequential()
2 model.add(layers.Embedding(input_dim=1000, output_dim=64))
3 model.add(layers.LSTM(128))
4 model.add(layers.Dense(10, activation='softmax'))
5 model.summary()
```

Layer (type)	Output Shape	Param #
embedding_2 (Embedding)	(None, None, 64)	64000
lstm_1 (LSTM)	(None, 128)	98816
dense_2 (Dense)	(None, 10)	1290
Total params: 164,106		
Trainable params: 164,106		
Non-trainable params: 0		

# GRU avec Keras

```
1 model = tf.keras.Sequential()
2 model.add(layers.Embedding(input_dim=1000, output_dim=64))
3 model.add(layers.GRU(128))
4 model.add(layers.Dense(10, activation='softmax'))
5 model.summary()
```

Layer (type)	Output Shape	Param #
embedding_3 (Embedding)	(None, None, 64)	64000
gru (GRU)	(None, 128)	74112
dense_3 (Dense)	(None, 10)	1290
Total params: 139,402		
Trainable params: 139,402		
Non-trainable params: 0		

# Deep Encoder avec Keras

```
1 model = tf.keras.Sequential()
2 model.add(layers.Embedding(input_dim=1000, output_dim=64))
3 model.add(layers.GRU(256, return_sequences=True))
4 model.add(layers.SimpleRNN(128))
5 model.add(layers.Dense(10, activation='softmax'))
```

Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, None, 64)	64000
gru (GRU)	(None, None, 256)	246528
simple_rnn (SimpleRNN)	(None, 128)	49280
dense (Dense)	(None, 10)	1290
Total params: 361,098		

# Deep Learning par la Pratique

Travaux Pratiques : RNN

---

[Keras - RNN - Voltaire Generator](#)



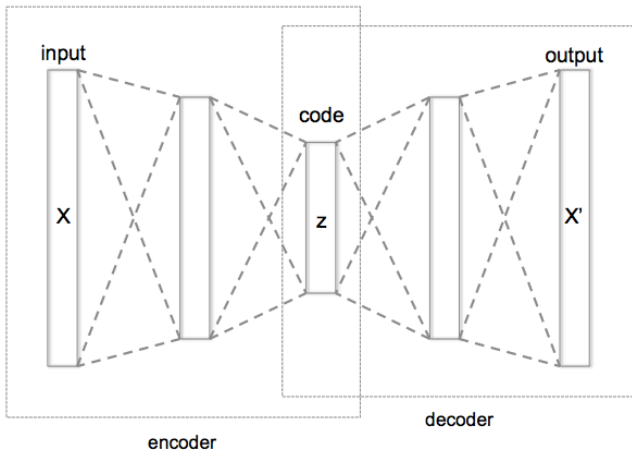
[Keras - seq2seq - Calculator](#)

# Deep Learning par la Pratique

Autoencodeurs

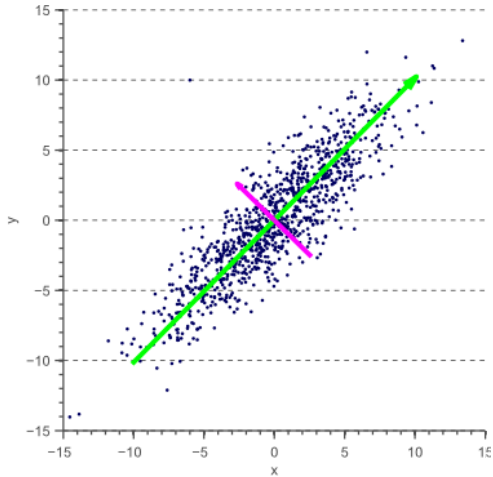
---

# Autoencodeurs



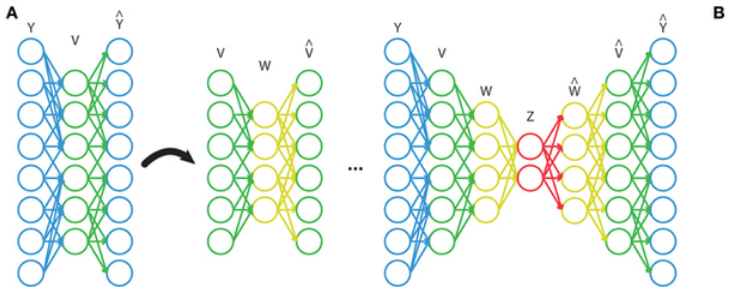
# Autoencodeurs

Activations linéaire et Loss MSE  $\approx$  ACP

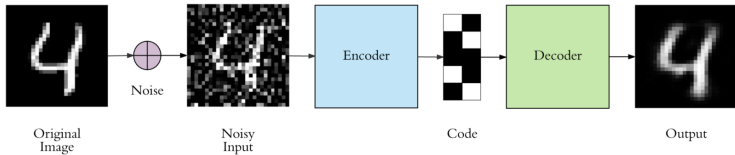


# Autoencodeurs

## Stacked Autoencoder (Empilés)

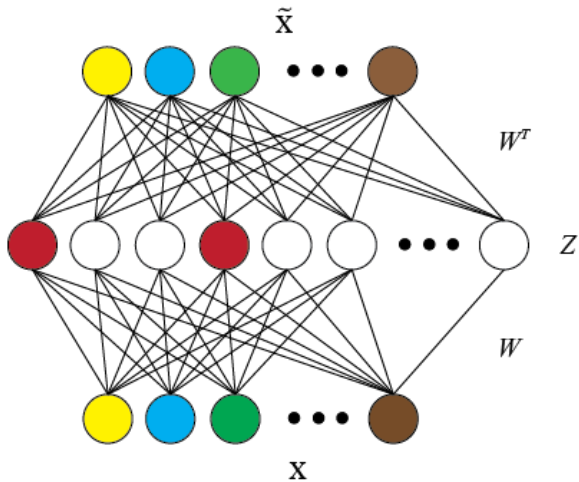


## Denoising Autoencoder (Débruiteurs)

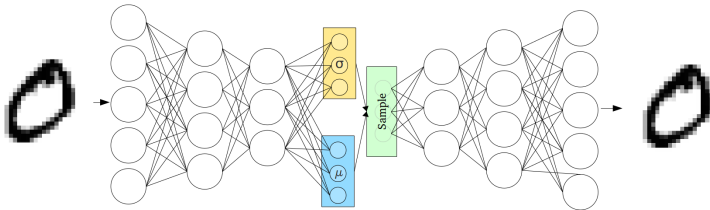


# Autoencodeurs

## Sparse Autoencoder (Épars)



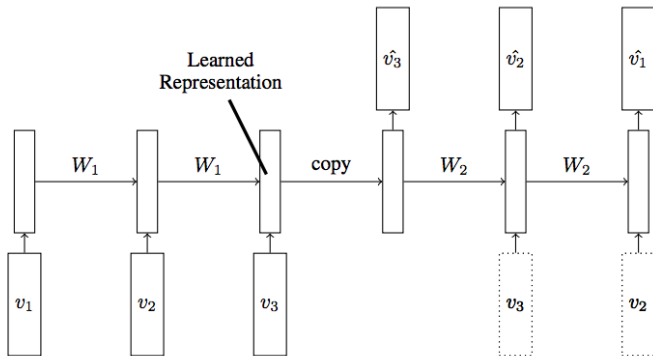
## Variational Autoencoder (Variationnels)





# Autoencoders

## LSTM Autoencoder



# Deep Learning par la Pratique

Travaux Pratiques : Autoencodeurs

---

## Keras - Autoencodeurs - Simple

## Keras - Autoencodeurs - Débruiteurs