



Bienvenue!

**ÉCOLE D'HIVER FRANCOPHONE
EN APPRENTISSAGE PROFOND**

5 - 9 mars 2018



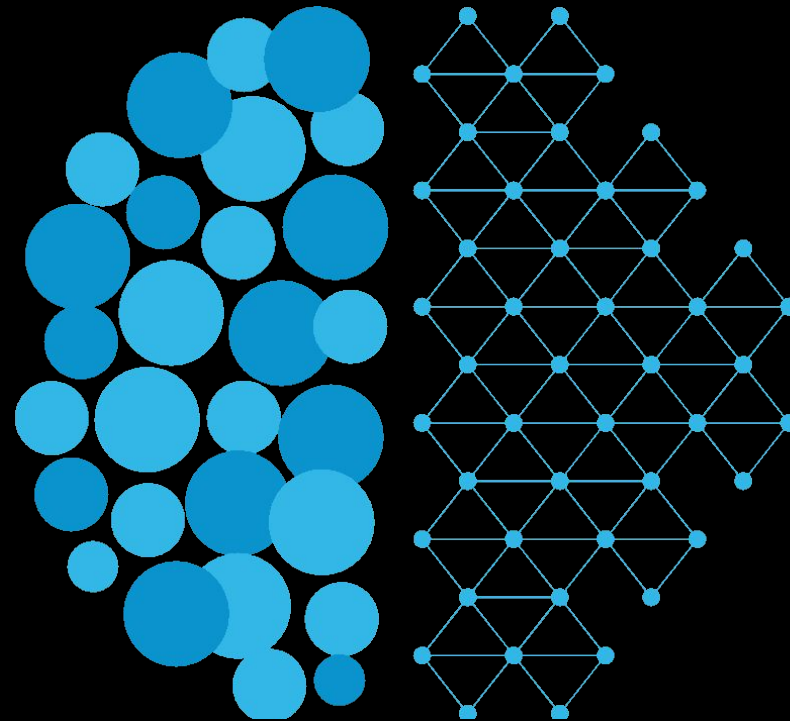
IVADO

HEC Montréal
Polytechnique Montréal
Université de Montréal



MILA

Institut
québécois
d'intelligence
artificielle

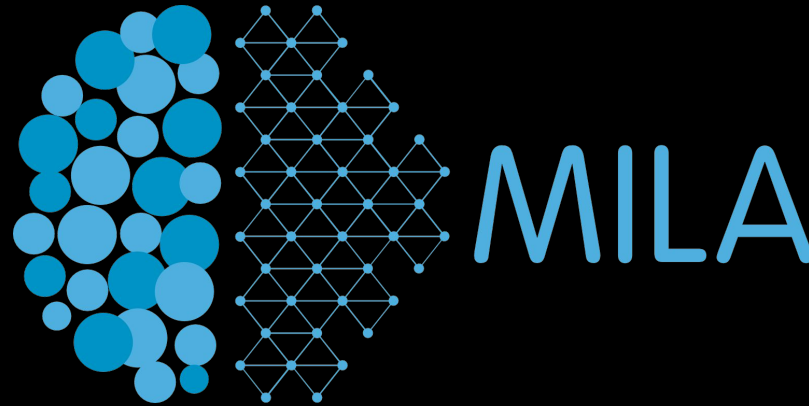


MILA

Université 
de Montréal



Institut
des algorithmes
d'apprentissage
de Montréal



Introduction

Apprentissage profond

Gaétan Marceau Caron
gaetan.marceau.caron@rd.mila.quebec

Apprentissage automatique

- Tâche
- Base d'apprentissage
- Modèle
- Algorithme d'apprentissage
- Évaluation du modèle

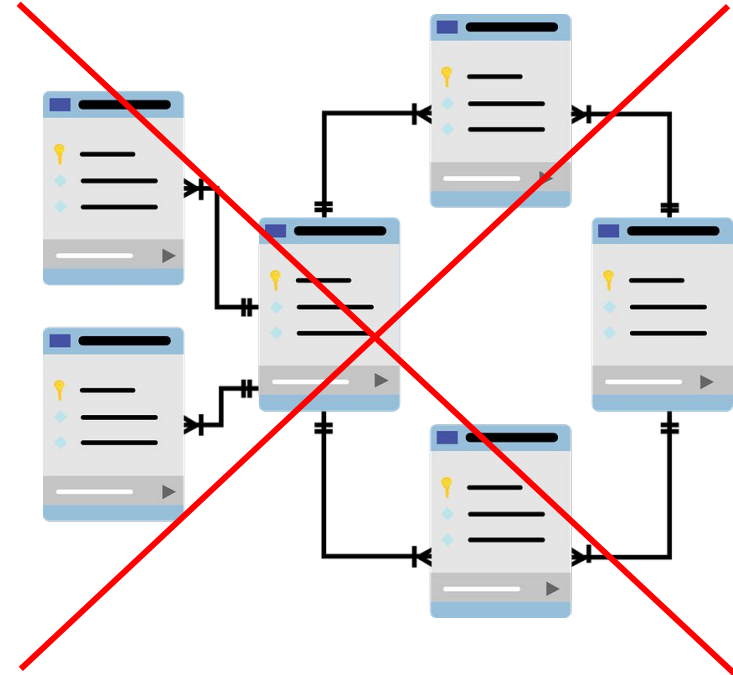
Définition de la tâche

Types d'apprentissage:

- **Supervisé (X, y)**
- **Non-supervisé (X)**
- Par renforcement (Actions, récompenses)

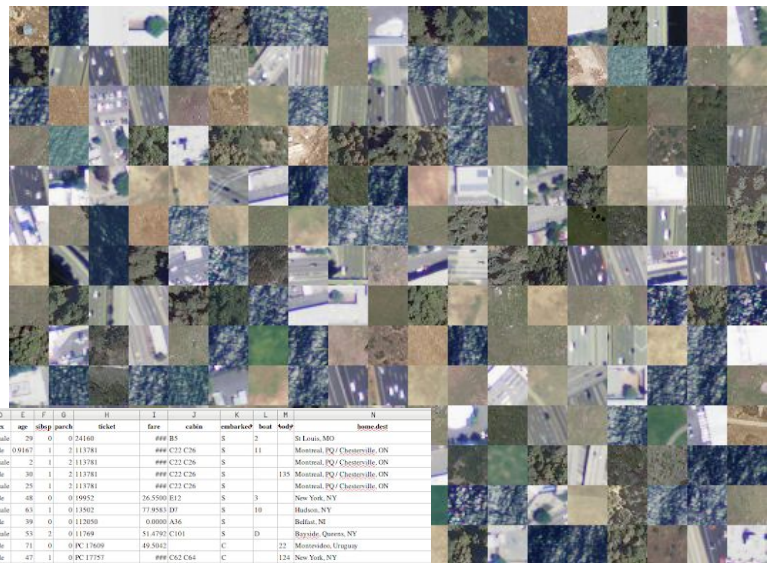
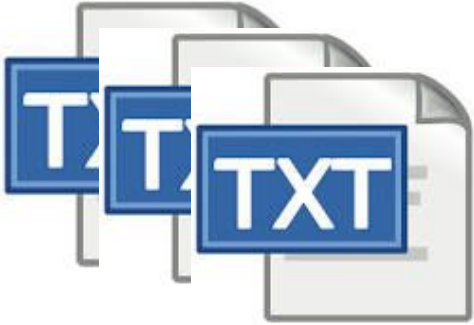
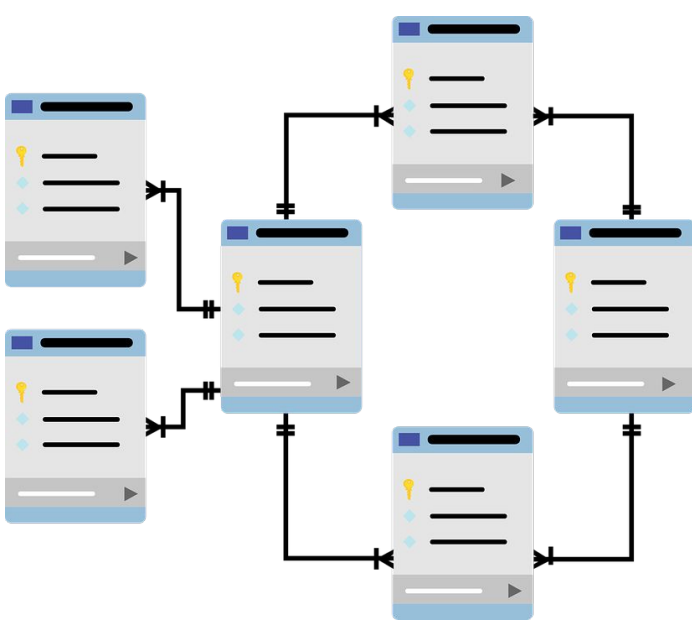
Base d'apprentissage

Une base de données
n'est pas une base
d'apprentissage!



Base d'apprentissage

(X, y)



Extraction de données



1	A	B	C	D	E	F	G	H	I	J	K	L	M	N
survived	class	name	sex	age	sibsp	parth	ticket	fare	cabin	embarked	boat	body	home.dest	
1	1	Allen, Miss. Elisabeth Walton	female	29	0	1	02160	##	B3	S			St Louis, MO	
2	1	Allen, Master. Hudson Trevor	male	0/167	1	1	2113781	##	C22 C26	S	11		Montreal, PQ / Chesapeake, ON	
3	1	Allen, Miss. Helen Louisa	female	2	1	1	2113781	##	C22 C26	S			Montreal, PQ / Chesapeake, ON	
4	1	Allen, Mr. Hudson Joshua Compston	male	30	1	1	2113781	##	C22 C26	S			135 Montreal, PQ / Chesapeake, ON	
5	1	Allen, Mrs. Hudson (F. Rose) Walter Doolan	female	25	1	1	2113781	##	C22 C26	S			Montreal, PQ / Chesapeake, ON	
6	1	Anderson, Mr. Harry	male	48	0	0	05952	26.5500	E12	S	3		New York, NY	
7	1	Andrews, Miss. Katerina Theodosia	female	63	1	0	013502	77.9500	D7	S	10		Boston, NY	
8	1	Andrew, Mr. Thomas Jr	male	59	0	0	0112009	10.0000	A38	S			Boston, MI	
9	1	Appleton, Mrs. Edward Dale (Charlotte Lamson)	female	53	2	0	0112049	51.4700	C101	S	D		Royalton, Queens, NY	
10	1	Atterberg, Mr. Ramon	male	71	0	0	0PC 17609	49.5000		C			Montevideo, Uruguay	
11	1	Axel, Col. John Jacob	male	47	1	0	0PC 17757	##	C62 C64	C			124 New York, NY	
12	1	Axel, Mrs. John Jacob (Mariahanna Telinde Power)	female	18	1	0	0PC 17757	##	C62 C64	C			New York, NY	
13	1	Ashford, Miss. Loretta Pauline	female	26	0	0	0PC 17437	69.3000	B33	C			Paris, France	
14	1	Baer, Miss. Ellen "Nellie"	female	26	0	0	05827	78.8500		S	6			
15	1	Barkworth, Mr. Algerton Henry Wilson	male	80	0	0	027042	30.0000	A23	S	B		Utah, York	
16	1	Barnston, Mr. John D	male	40	0	0	0PC 17318	35.8200		S			New York, NY	
17	1	Barron, Mr. Quigg Edmund	male	24	0	0	0PC 17558	##	B38 B60	C			Montreal, PQ	
18	1	Barron, Mrs. James Oliver (DeLondras) Chappell	female	50	0	0	0PC 17558	##	B38 B60	C	6		Montreal, PQ	
19	1	Bazant, Miss. Alberta	female	32	0	0	011813	76.2917	D15	C	8			
20	1	Beatty, Mr. Thomas	male	36	0	0	013050	75.2417	C4	C	A		Wilmington, MN	
21	1	Beckwith, Mr. Richard Leonard	male	37	1	1	111781	52.5542	D05	S	8		New York, NY	
22	1	Beckwith, Mrs. Richard Leonard (Sallie Maryann)	female	47	1	1	111781	52.5542	D05	S	5		New York, NY	
23	1	Belo, Mr. Karl Howell	male	26	0	0	011368	30.0000	C148	C	5		New York, NY	
24	1	Biles, Miss. Beatrice	female	42	0	0	0PC 17757	##		C	4			
25	1	Bird, Miss. Ellen	female	29	0	0	0PC 17483	##	C97	S	8			
26	1	Bircham, Mr. Jakob	male	25	0	0	013905	26.0000		C			148 San Francisco, CA	
27	1	Bishop, Mr. Dickinson H	male	29	1	0	011967	91.0700	B46	C	7		Dorchester, MI	
28	1	Bishop, Mrs. Dickinson (Helen Walton)	female	19	1	0	011967	91.0700	B49	C	7		Dorchester, MI	
29	1	Bissone, Mrs. Anadia	female	35	0	0	0PC 17780	##	C99	S	8			
30	1	Bjornstrom Suominen, Mr. Maaret Erika	female	28	0	0	010564	26.5500	C52	S	D		Stockholm, Sweden / Washington, DC	
31	1	Blackwell, Mr. Stephen Wyatt	male	45	0	0	0113781	35.5000	F	S			Thames, NJ	
32	1	Black, Mr. Henry	male	40	0	0	0112377	31.0000	A11	C	7		Old Bridge, NJ	
33	1	Bonnell, Miss. Caroline	female	30	0	0	06928	##	C7	S	8		Yonkers, OH	
34	1	Bonnell, Miss. Caroline	female	58	0	0	0113781	26.5500	C103	S	8		Bethlehem, England / Cleveland, Ohio	
35	1	Boothby, Mr. John James	male	42	0	0	0110489	26.5500	D02	S			London / Winnipeg, MB	
36	1	Boothby, Mrs. Grace Scott	female	45	0	0	0PC 17608	##		C	4		Croqueton, NY	
37	1	Bowman, Miss. Elsie Edith	female	22	0	1	113505	55.0000	E33	S	A		St Leonards-on-Sea, England / Ohio	
38	1	Bradley, Mr. George (George Arthur Briggs?)	male	0	0	0	011427	26.5500		S	9		Los Angeles, CA	
39	1	Brady, Mr. John Burton	male	41	0	0	0113054	30.5000	A21	S			Perrygo, WA	
40	1	Brady, Mr. Emil	male	48	0	0	0PC 17783	50.4000	B10	C	5		288 Omaha, NE	
41	1	Brady, Dr. Arthur Jackson	male	0	0	0	0112379	39.6000		C			Philadelphia, PA	
42	1	Brown, Mrs. James Joseph (Margaret Edith)	female	44	0	0	0PC 17610	77.2000	B4	C	6		Denver, CO	
43	1	Brown, Mrs. John Munny (Caroline Lane Lamson)	female	59	2	0	011789	51.4700	C101	S	D		Boston, MA	
44	1	Brown, Mrs. William Robert (Emma Eliza West)	female	60	0	0	011813	26.2917	D15	C	8		Philadelphia, PA	
45	1	Brown, Miss. Elizabeth Margaret	female	41	0	0	06966	##	E40	C	3			
46	1	Burt, Major. Archibald Wellingham	male	45	0	0	0113050	26.5500	B38	S			Washington, DC	
47	1	Carron, Mr. Alexander	male	40	0	0	0113794	31.0000		S				
48	1	Catherine, Mr. Edward Pennington	male	42	0	0	0PC 17476	26.2917	E24	S	5		New York, NY	
49	1	Caulson, Mrs. Edward (Helen Charlotte) Blackwood	female	63	0	0	0PC 17606	77.4200		C	6		Washington, DC	



Base d'apprentissage

$x \in \mathbb{R}^4$

Fisher's Iris Data

Y

Sepal Length	Sepal Width	Petal Length	Petal Width	Species
5.1	3.5	1.4	0.2	<i>setosa</i>
4.9	3.0	1.4	0.2	<i>setosa</i>
4.7	3.2	1.3	0.2	<i>setosa</i>
4.6	3.1	1.5	0.2	<i>setosa</i>
5.0	3.6	1.4	0.2	<i>setosa</i>
5.4	3.9	1.7	0.4	<i>setosa</i>
4.6	3.4	1.4	0.3	<i>setosa</i>
etc...				
5.7	2.8	4.5	1.3	<i>versicolor</i>
6.3	3.3	4.7	1.6	<i>versicolor</i>
4.9	2.4	3.3	1.0	<i>versicolor</i>
6.6	2.9	4.6	1.3	<i>versicolor</i>
5.2	2.7	3.9	1.4	<i>versicolor</i>
5.0	2.0	3.5	1.0	<i>versicolor</i>
etc...				
7.7	3.0	6.1	2.3	<i>virginica</i>
6.3	3.4	5.6	2.4	<i>virginica</i>
6.4	3.1	5.5	1.8	<i>virginica</i>
6.0	3.0	4.8	1.8	<i>virginica</i>
6.9	3.1	5.4	2.1	<i>virginica</i>

n=150



Iris setosa



Iris versicolor



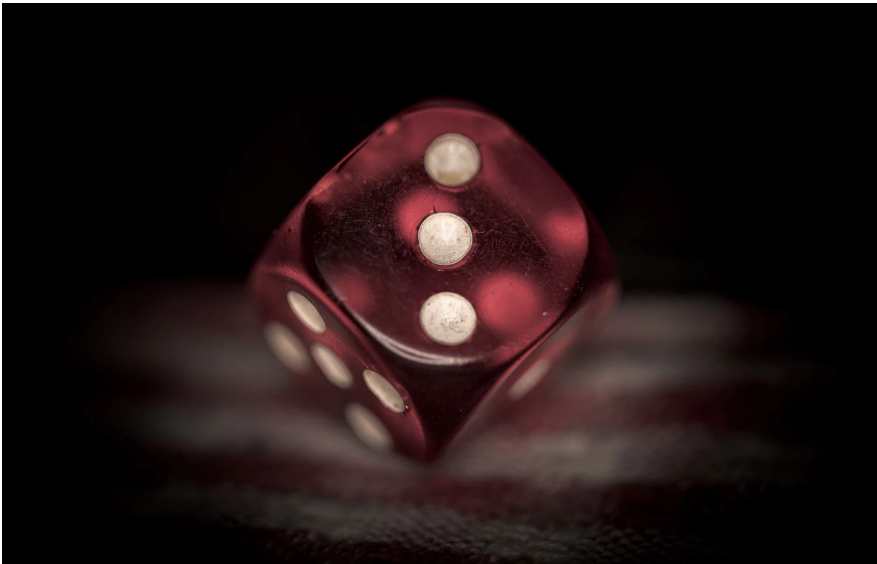
Iris virginica



Ronald Fisher
The use of multiple measurements in taxonomic problems (1936)

Statistique

Indépendant et identiquement distribué (iid)



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

Base d'apprentissage

Ensemble
d'entraînement



Déploiement



Nouvel
exemple



Non-stationnarité



Data augmentation



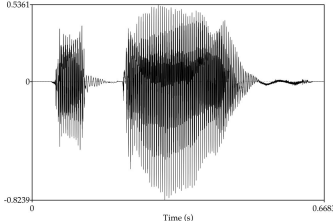

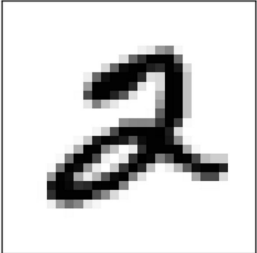

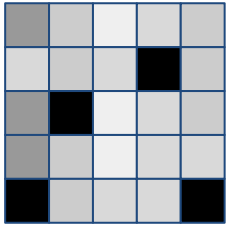
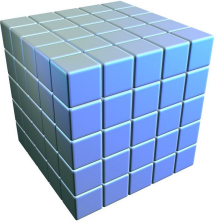


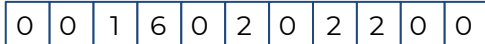
Symétries



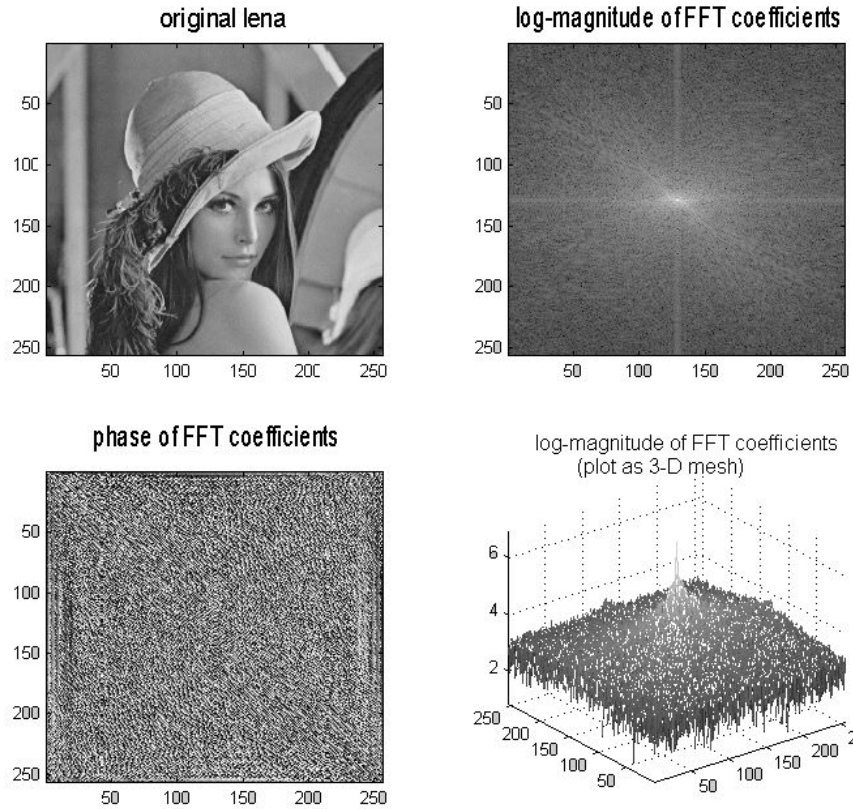
1 exemple

4 exemples

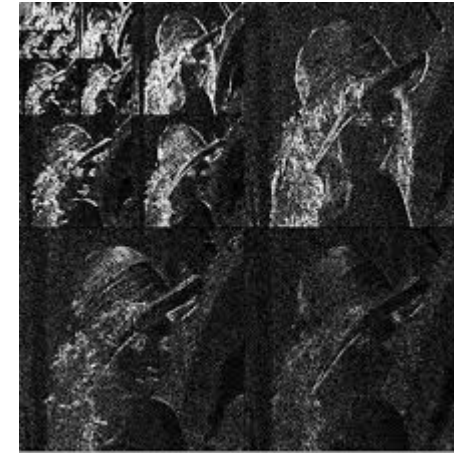
Représentation de l'entrée

	Représentation naturelle	Représentation à l'entrée du modèle
Son		 Séquence de valeurs réelles
Image		 1D  2D  3D
Catégoriel	<code>\x48\x65\x6c\x6c\x6f</code> (UTF8)	 Vecteur One-hot
Document	<code>\x48\x65\x6c\x6c\x6f...</code> (UTF8)	 Séquence de vecteurs One-hot  Bag-of-words Ou tf-idf

La meilleure représentation?



Wavelet decomposition



<https://www.ee.columbia.edu/~lx/courses/dip/imgfun/fft2.html>

Preprocessing

- Normaliser les données X et les données y

- Standardisation $\tilde{x} = \frac{x - \mu}{\sigma}$

- Normalisation $\tilde{x} = \frac{x - x_{min}}{x_{max} - x_{min}}$

Base d'apprentissage

Décomposition en trois ensembles:

- ensemble **d'entraînement**
- ensemble de **validation**
- ensemble de **test**



Apprentissage supervisé

Ingrédients:

- Base d'apprentissage
- **Modèle**
- Algorithme d'apprentissage

Modèle

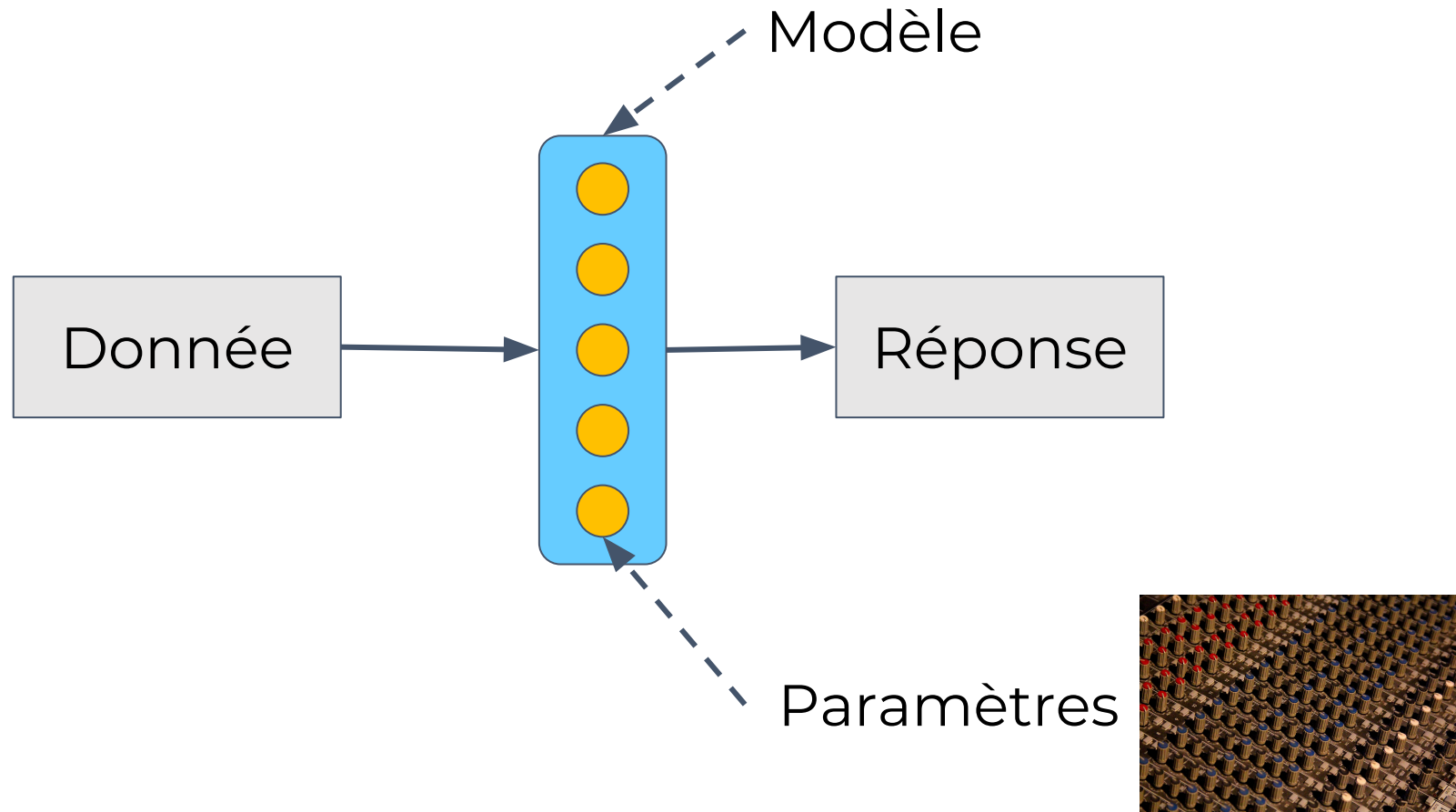
Interprétation **informatique**

- Programme: $f(X) \rightarrow y$

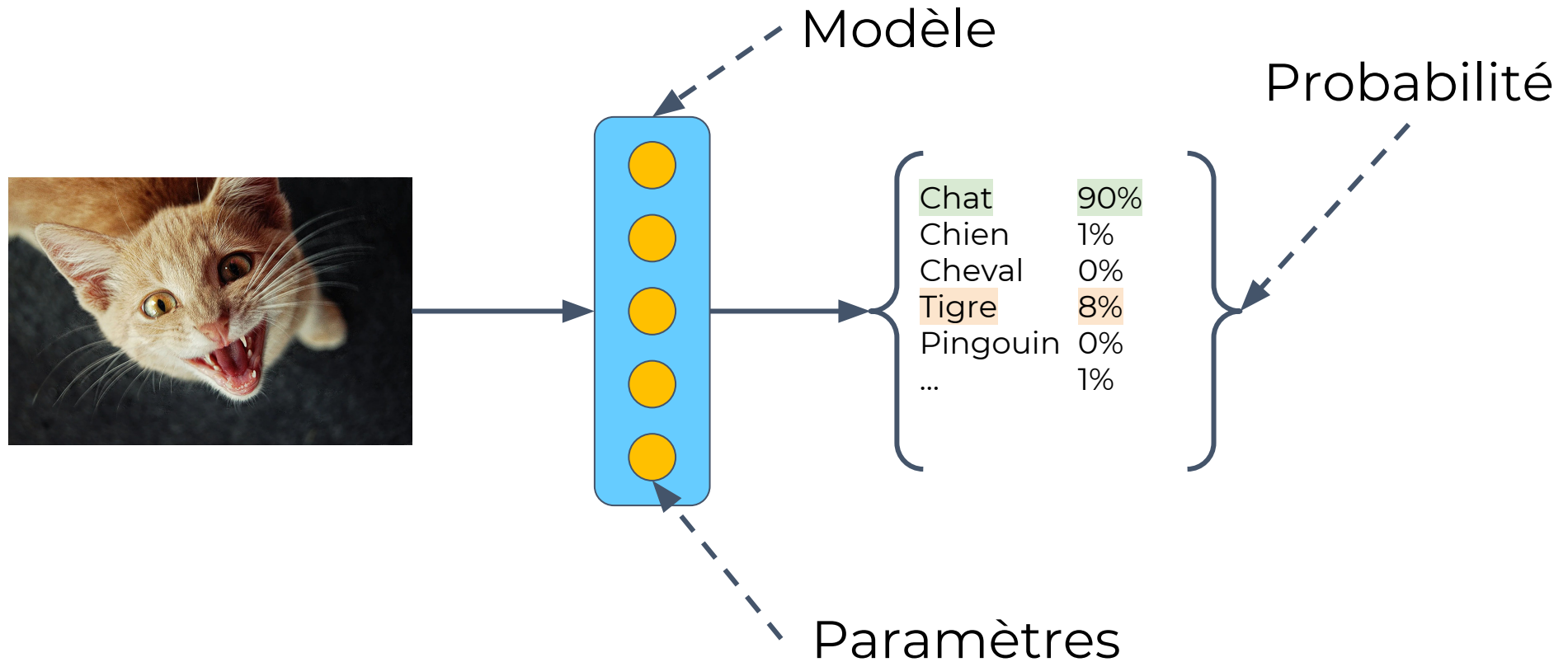
Interprétation **probabiliste**

- Distribution de probabilité: $p(y|X)$

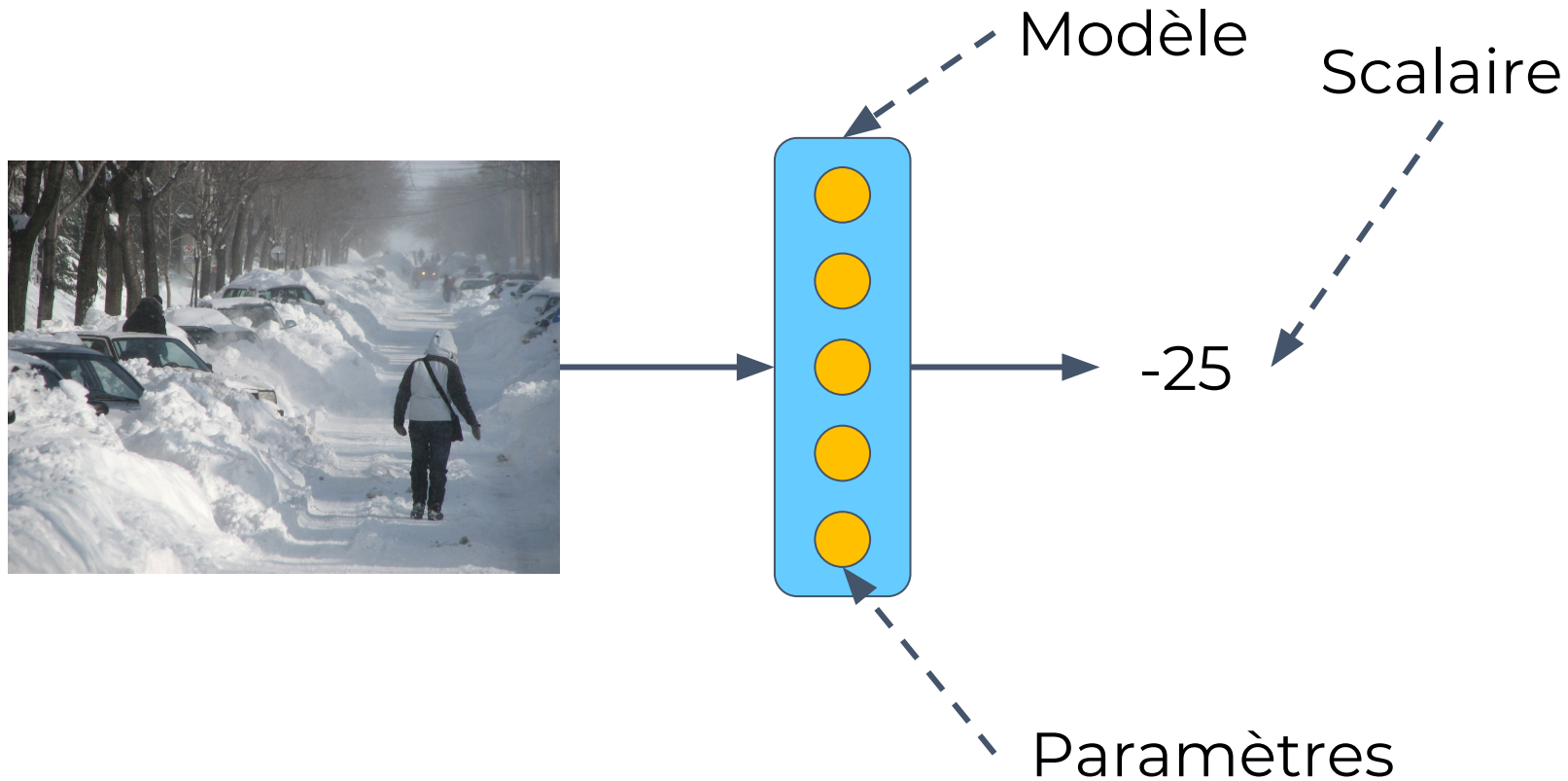
Définition du modèle



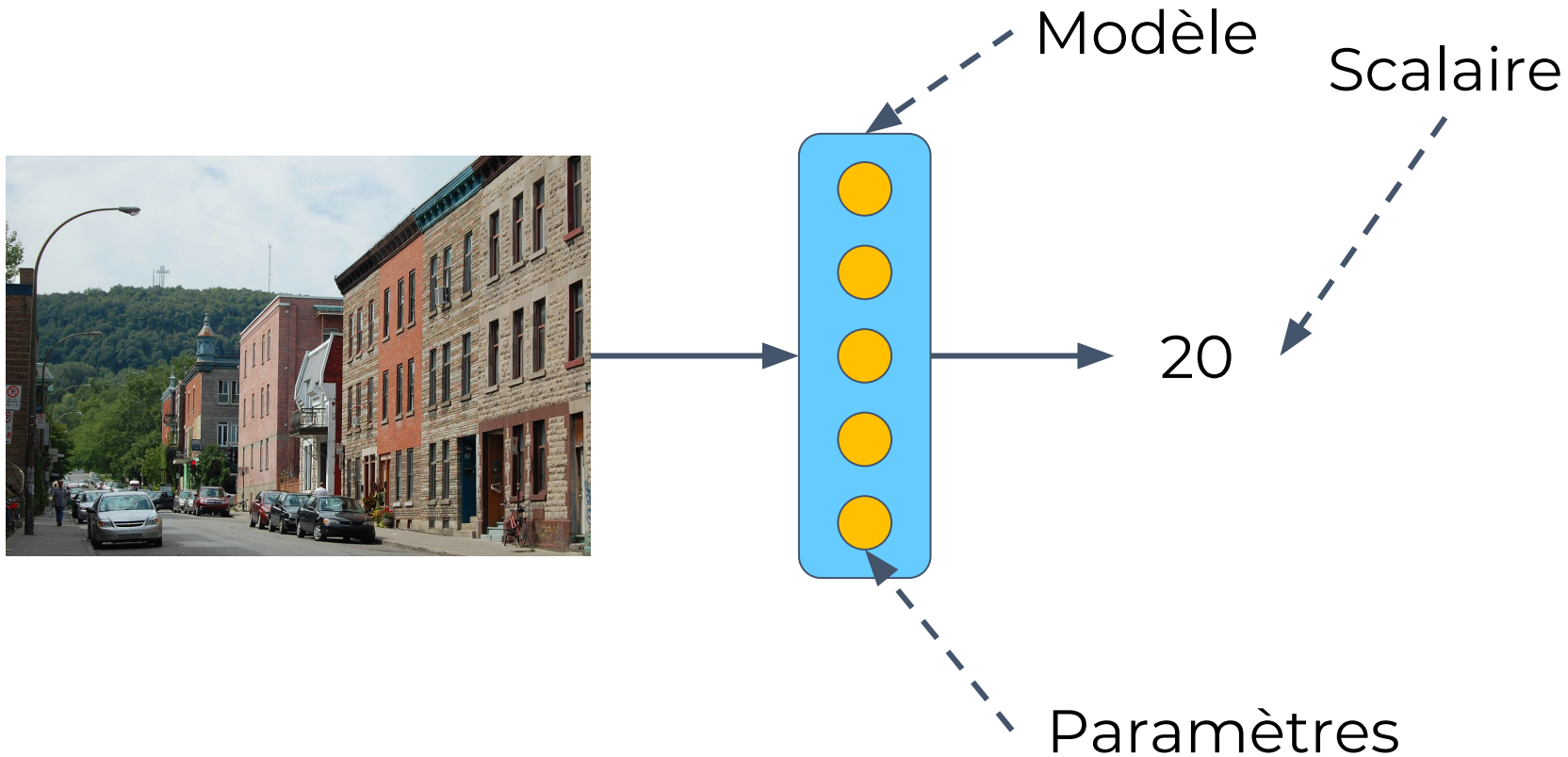
Exemple: déterminer l'animal



Exemple: déterminer la température



Exemple: déterminer la température



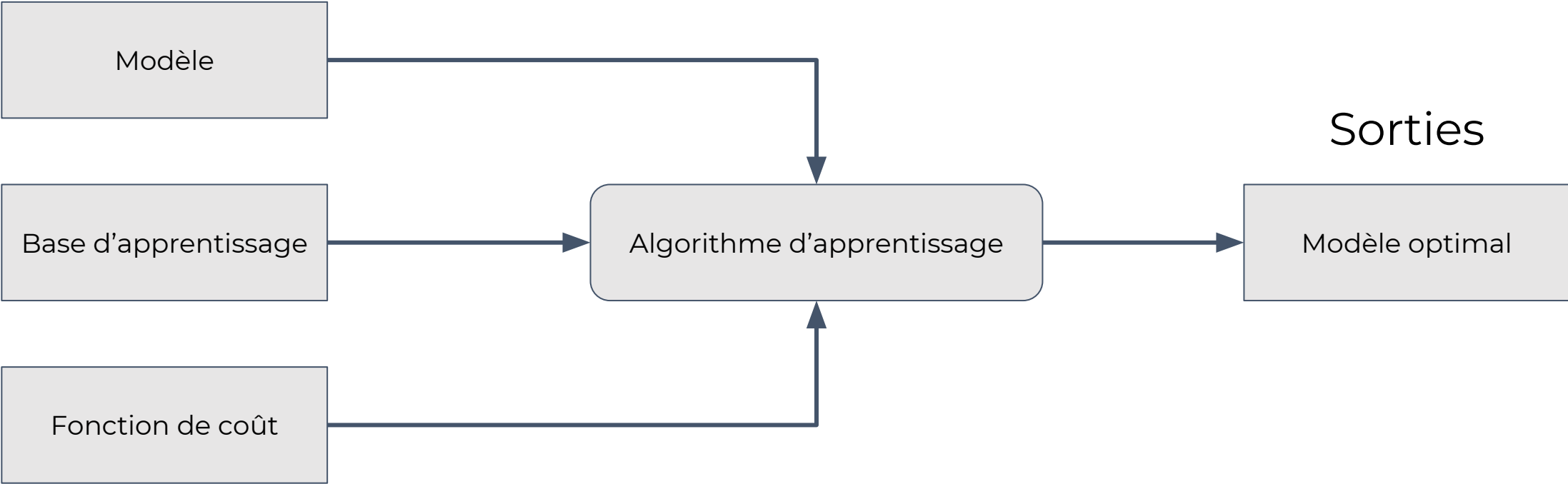
Apprentissage supervisé

Ingrédients:

- Base d'apprentissage
- Modèle
- **Algorithme d'apprentissage**

Algorithme d'apprentissage

Entrées



Une balade en million de dimensions

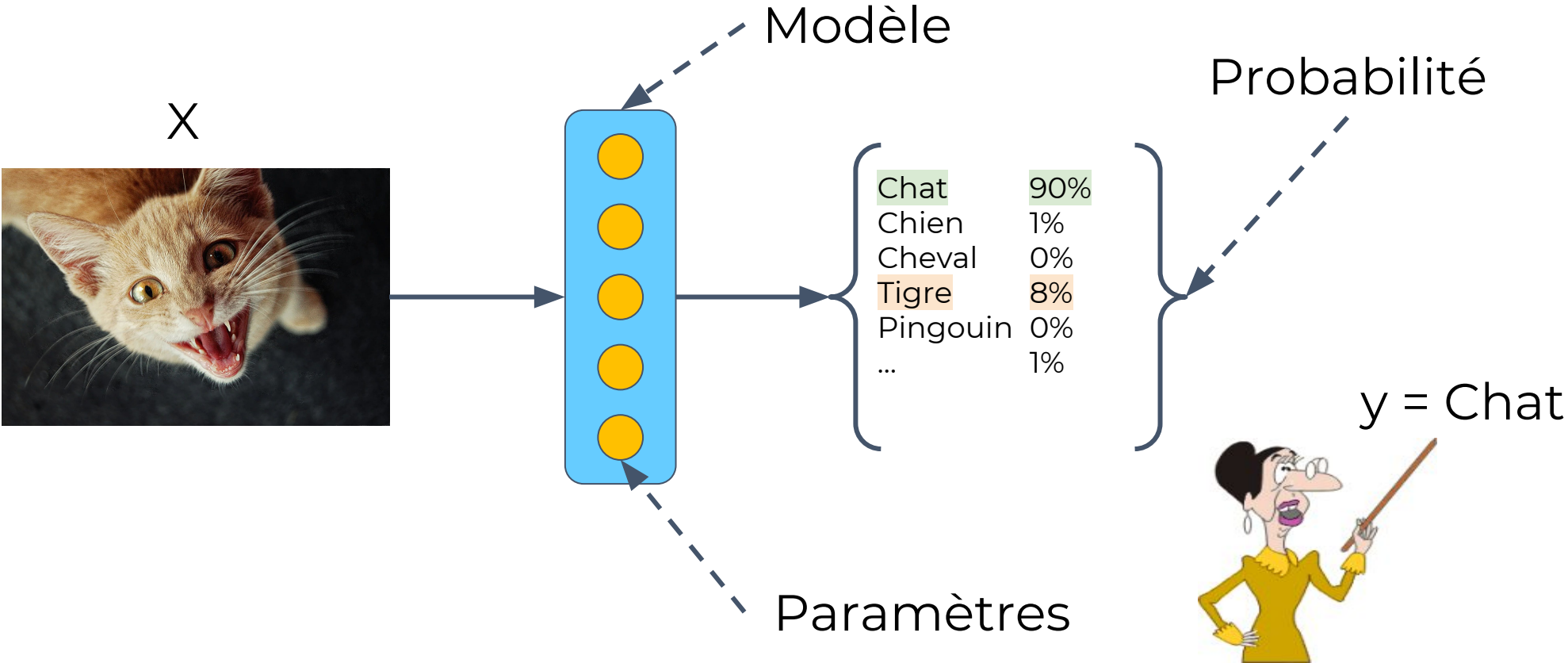
Optimisation	Alpinisme
Paramètre	Coordonnée
Coût	Altitude
Gradient	Pente
Mise à jour	Déplacement



Fonctions de coût

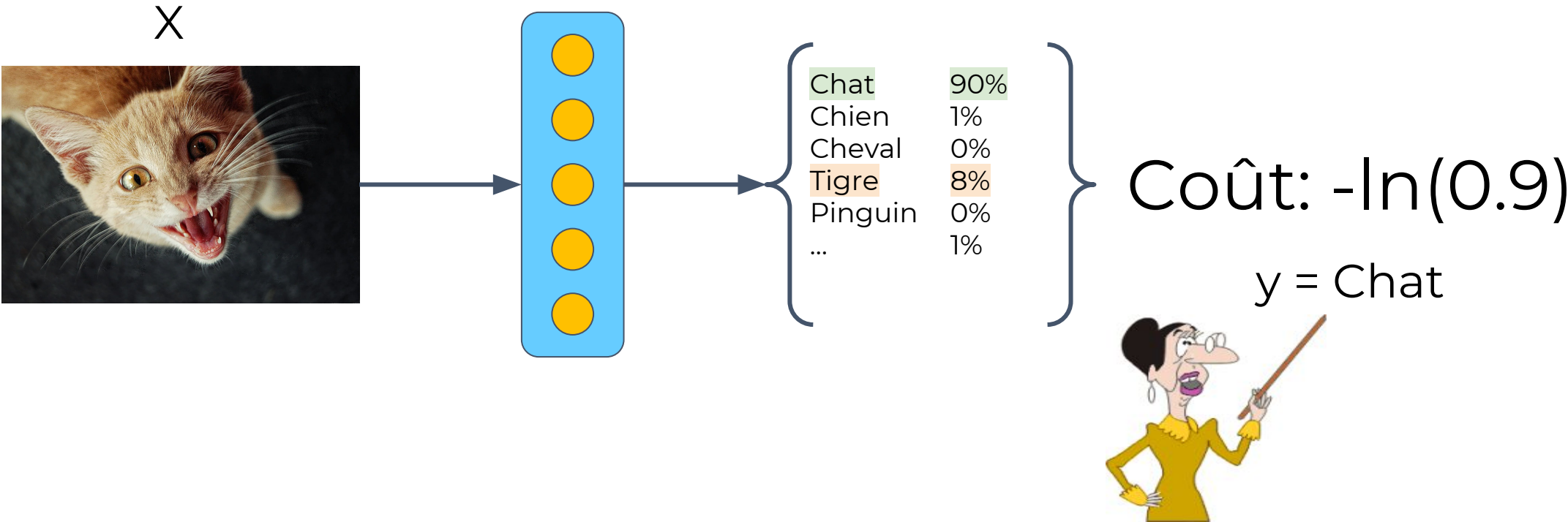
- Maximum de vraisemblance
- Méthode des moindres carrés
- Théorie de la décision
- ...

Exemple: maximiser la probabilité



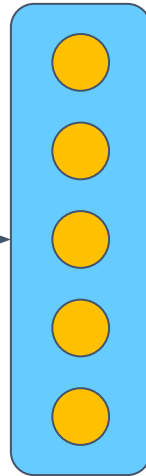
Exemple: maximiser la probabilité

Negative log-likelihood

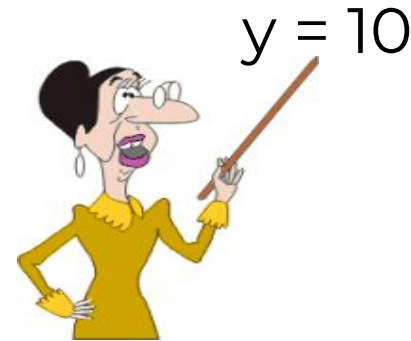


Exemple: réduire l'erreur au carrée

X



20

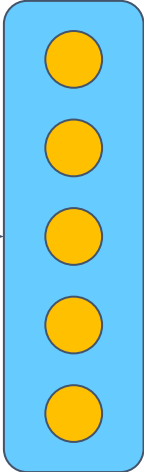


$y = 10$

Exemple: réduire l'erreur au carrée

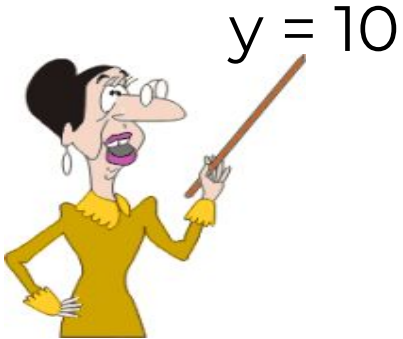
x

Mean Squared Error



20

Coût: $(20-10)^2$



y = 10

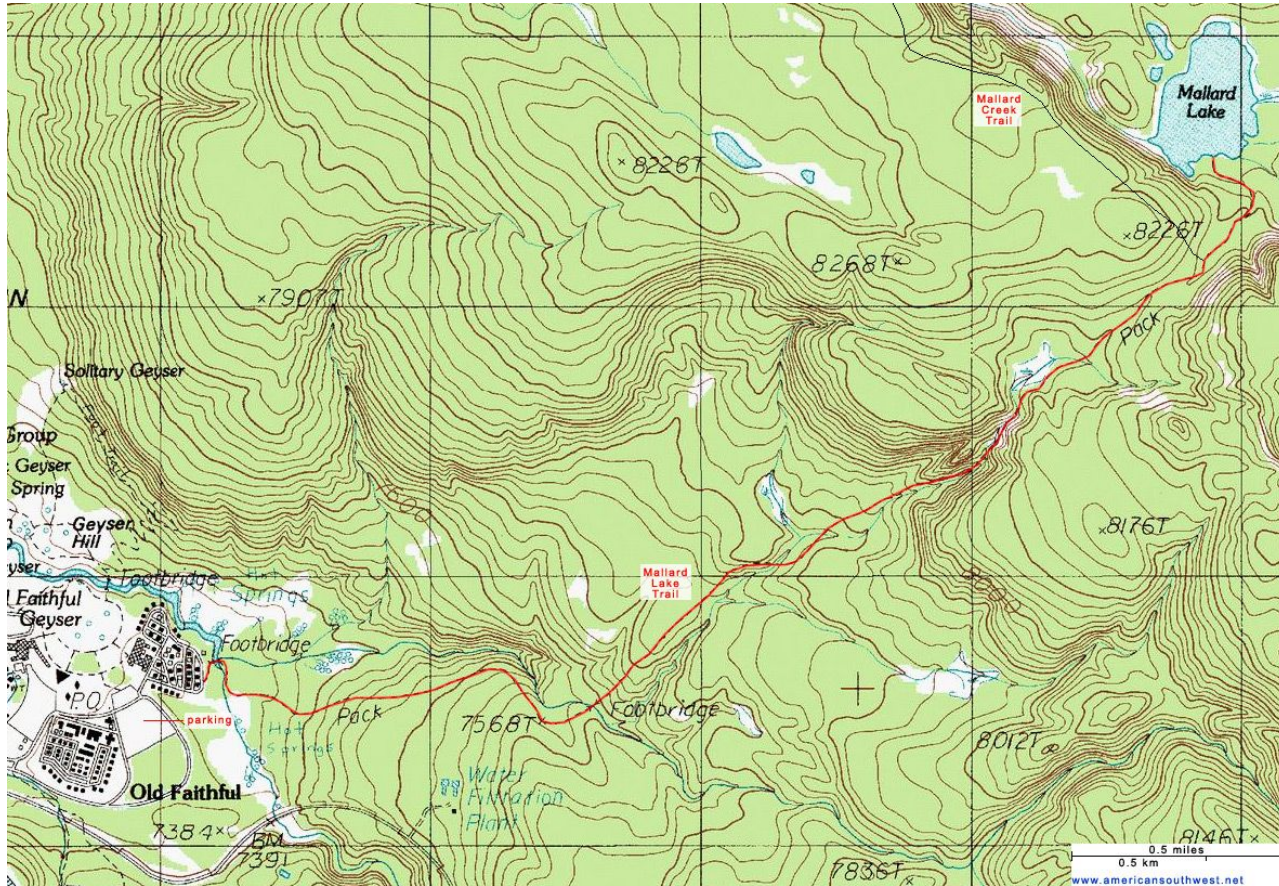
Fonction de coût

- Rétroaction sur la réponse du modèle
- Moyenne sur tous les exemples de l'ensemble d'entraînement
- Définie les courbes de niveaux (altitude)



Calcul de la pente

- Un petit pas dans chaque direction



Gradient: calcul de la pente

- Fonction différentiable: la fonction s'approxime bien avec une fonction linéaire (gradient)
- **Théorème**: le gradient donne la direction de la plus forte pente.
- Attention: information locale



Terminologie

- Calcul du gradient: **backpropagation**
- Se déplacer avec le gradient: **optimiseur**

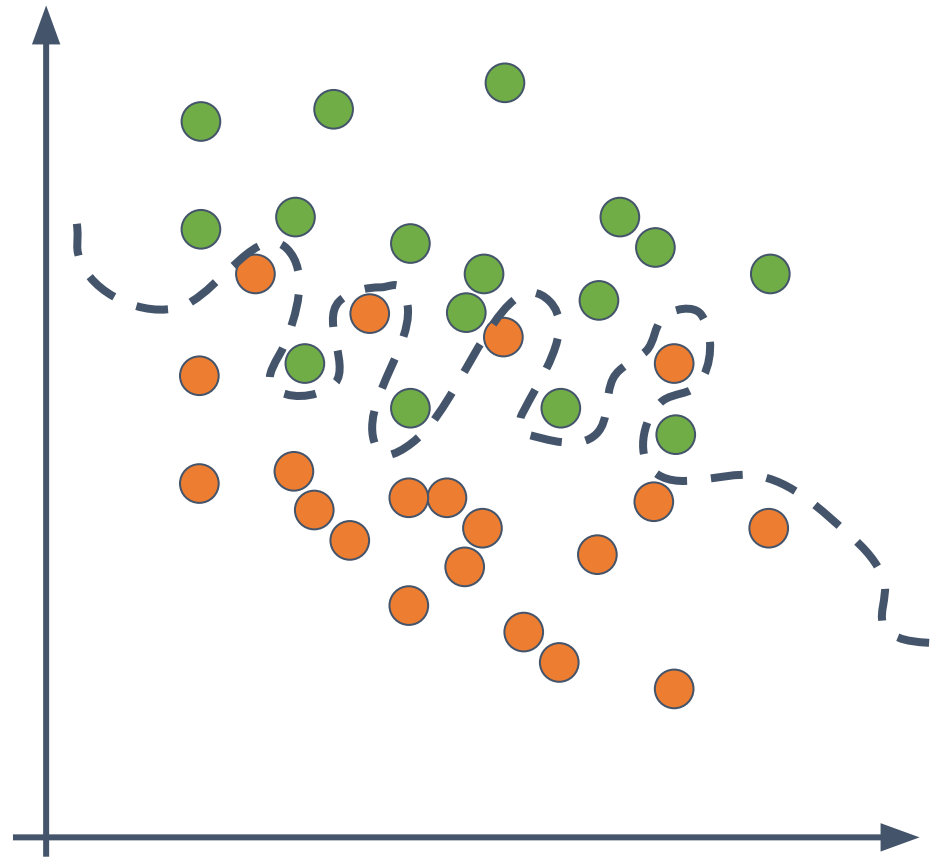
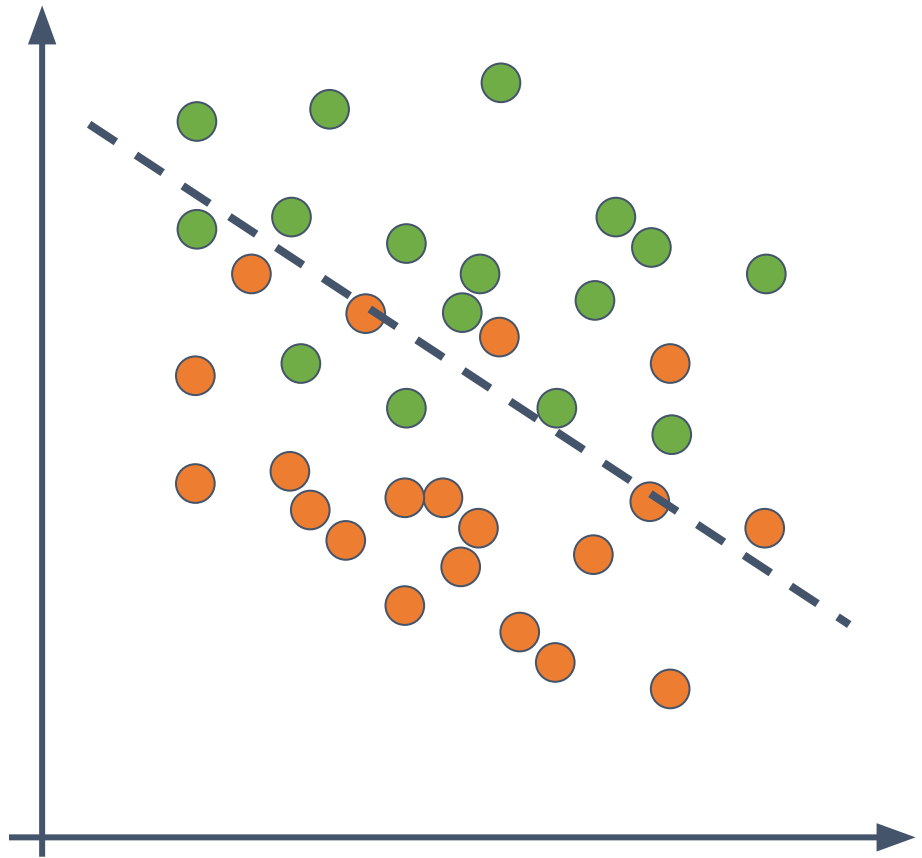
Base d'apprentissage

Décomposition en trois ensembles:

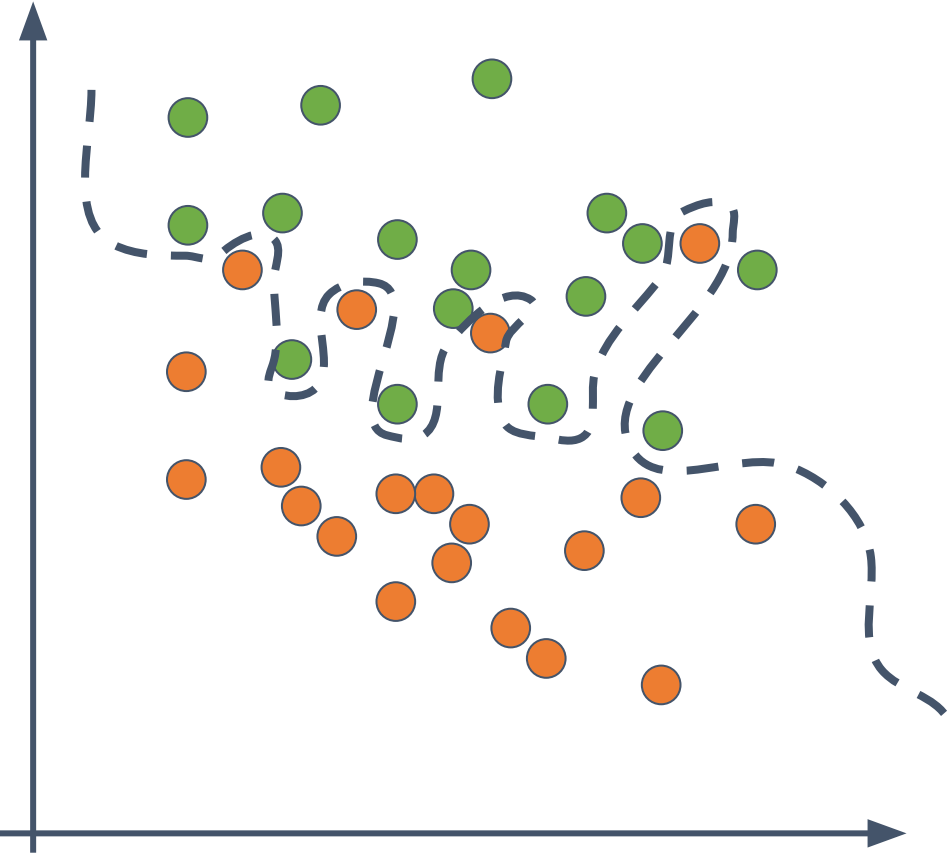
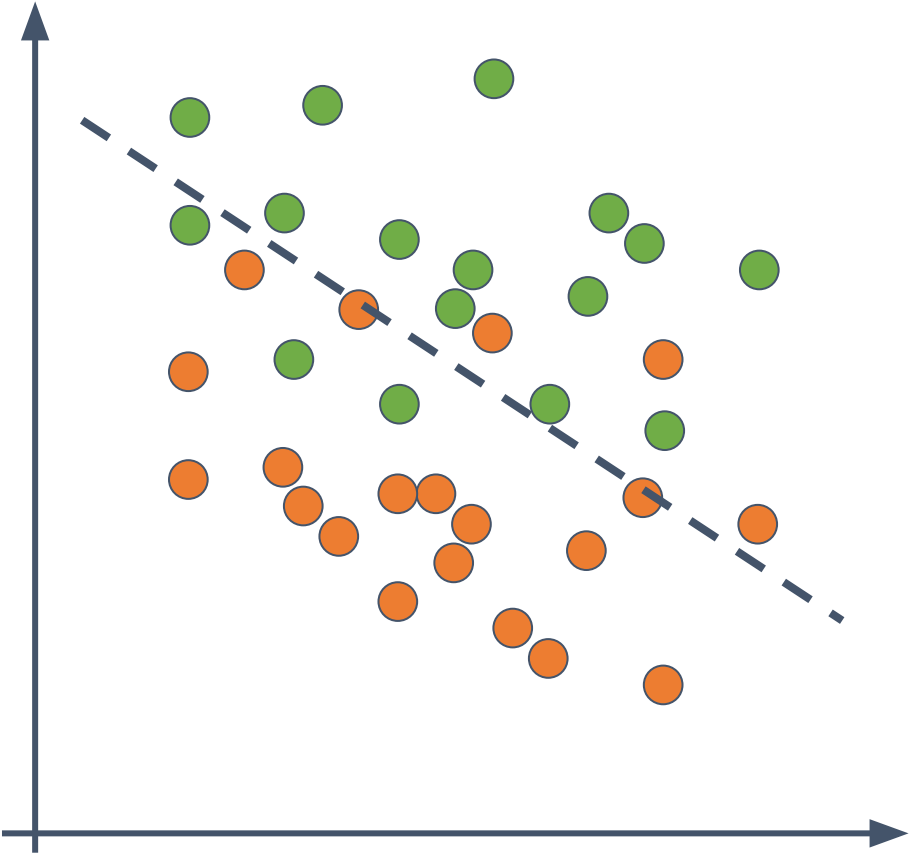
- ensemble **d'entraînement**
- ensemble de **validation**
- ensemble de **test**



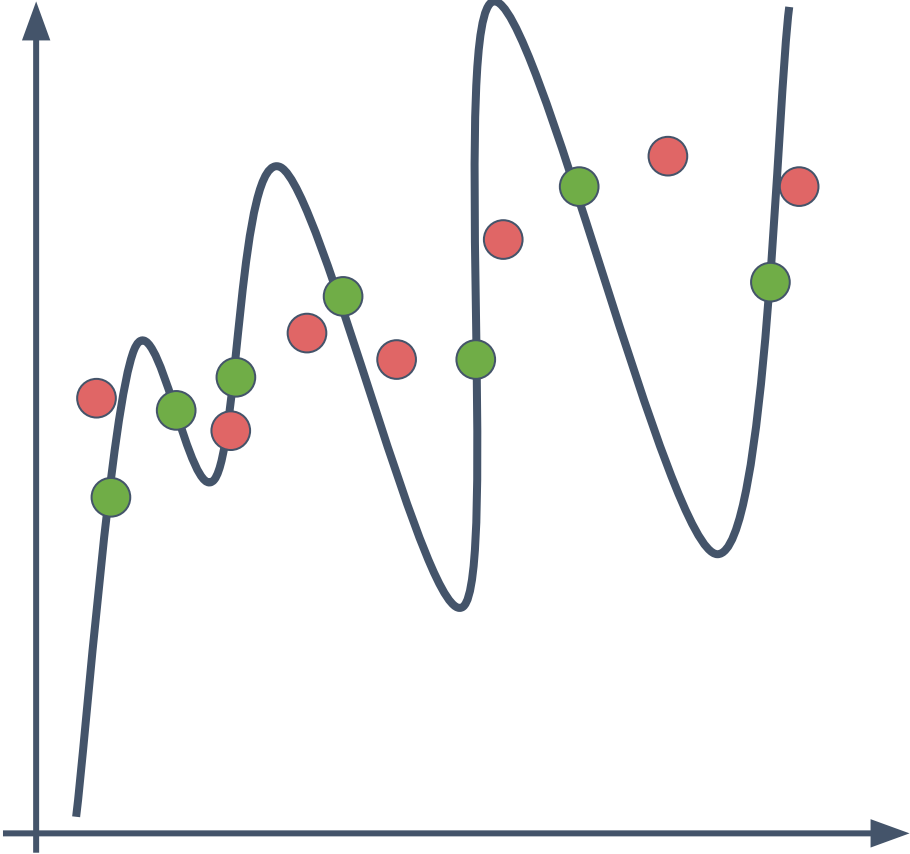
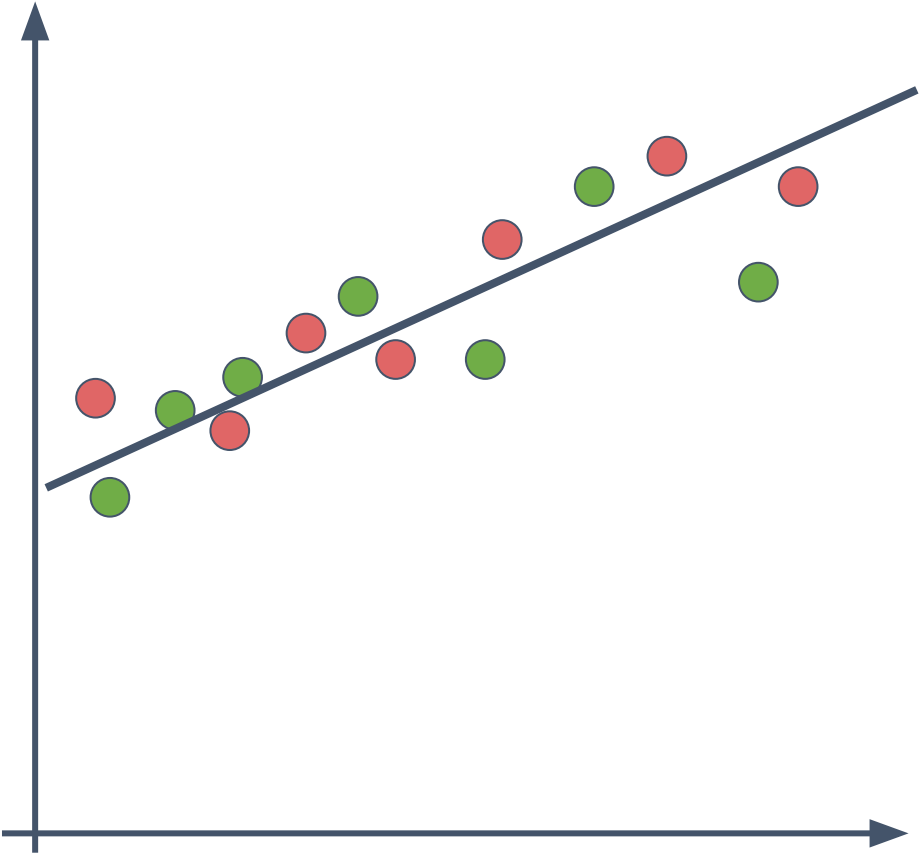
Sur-apprentissage: classification



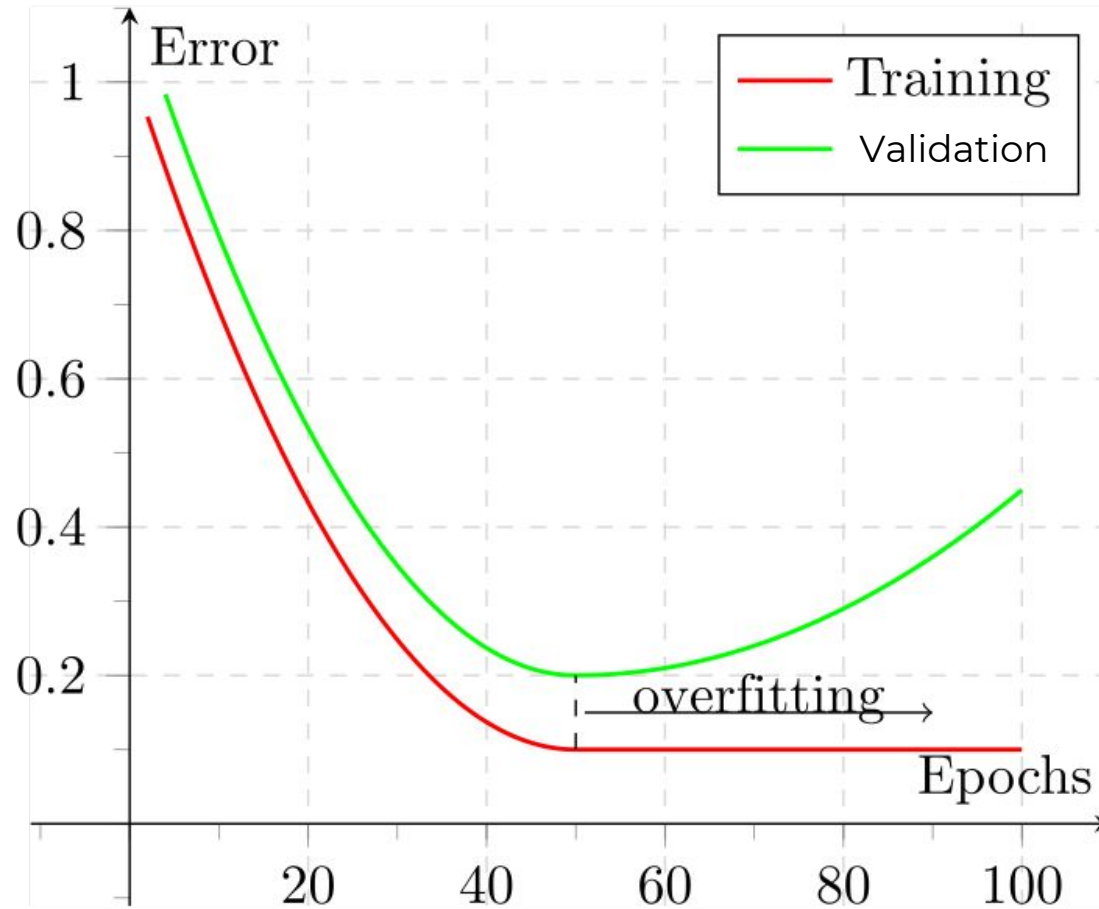
Sur-apprentissage: classification



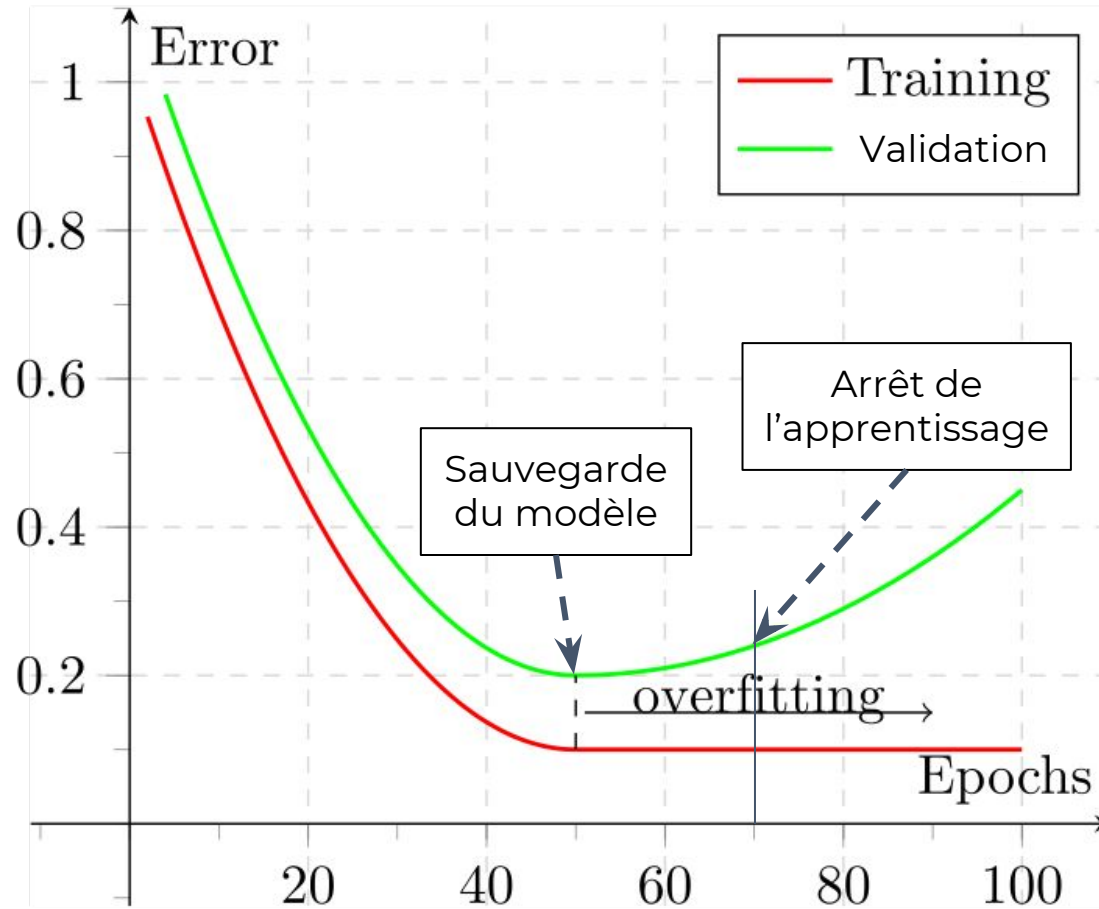
Sur-apprentissage: régression



Early Stopping



Early Stopping



Fonction de coût: régularisation

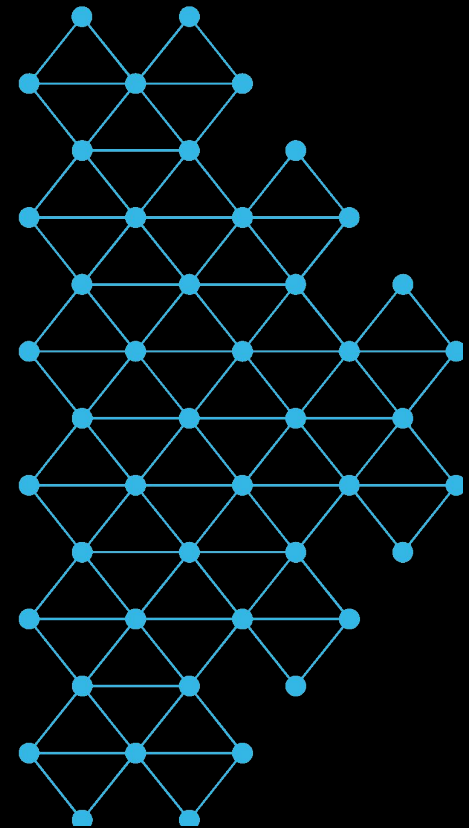
- Lutter contre le sur-apprentissage
- Pénalité sur les paramètres (coordonnées)
- Norme L1 ou L2
- D'autres types de régularisations plus efficaces

Hyperparamètres

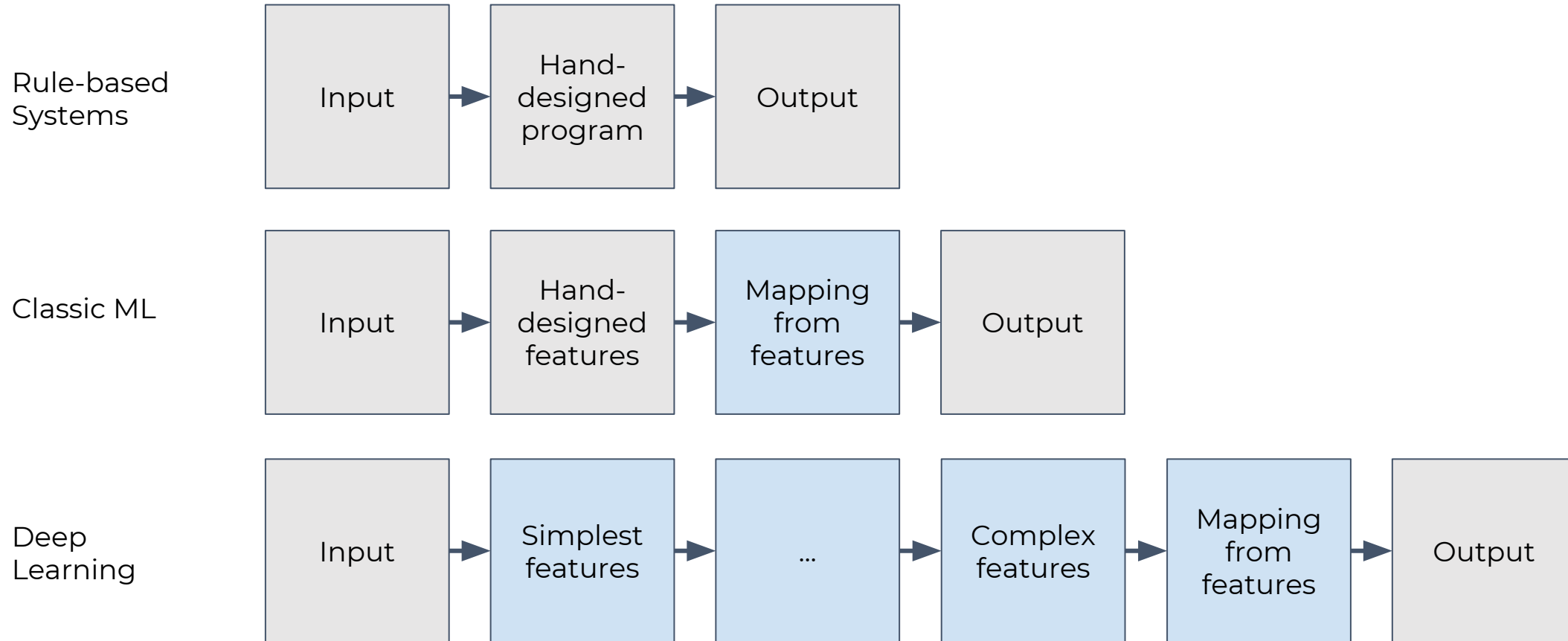
- Choix du modèle
- Structure du modèle
- Paramètres de l'optimiseur

Apprentissage profond

(ou apprentissage de représentations)

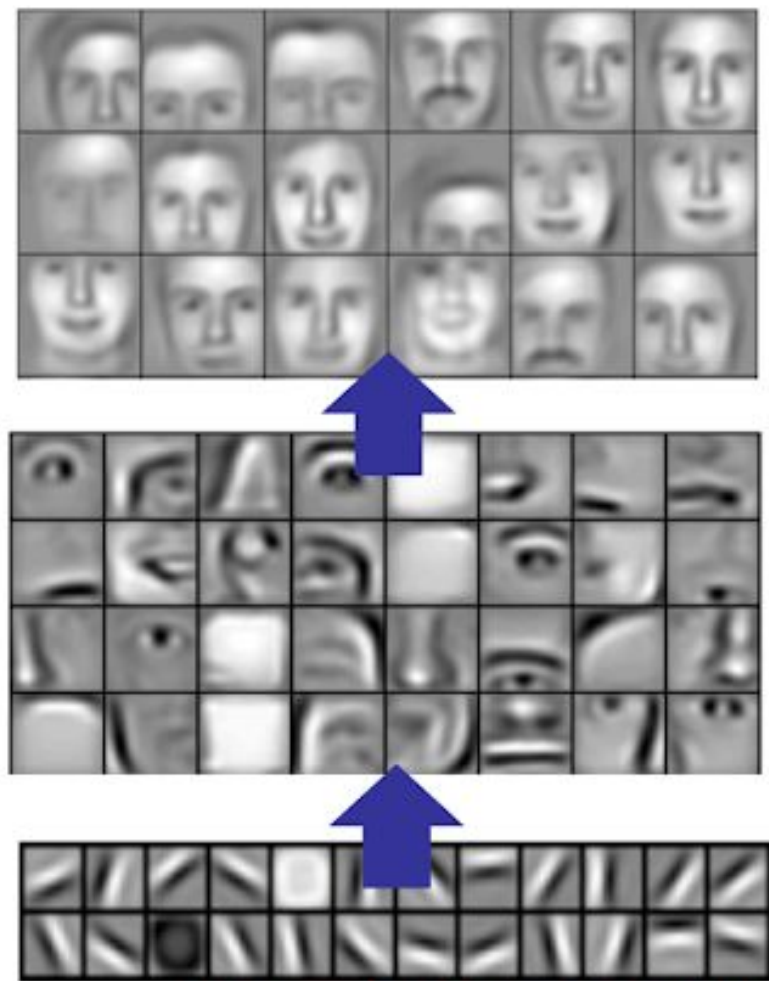


Modèle en apprentissage profond



Niveaux de représentations

(Lee, Largman, Pham & Ng, NIPS 2009)
(Lee, Grosse, Ranganath & Ng, ICML 2009)

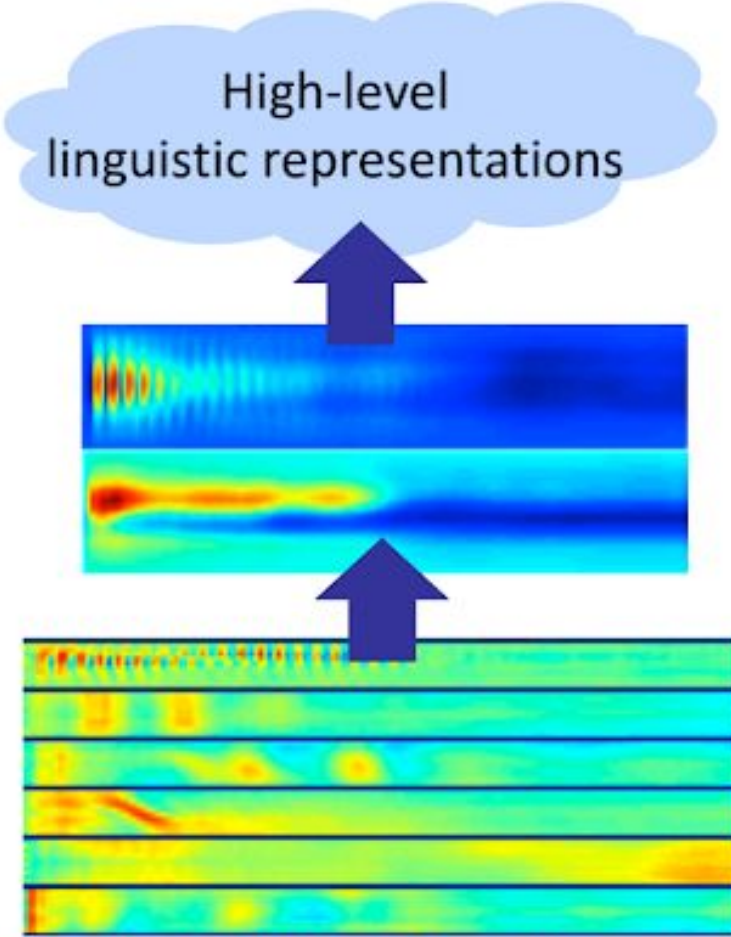


Layer 3

Parts combine to form objects

Layer 2

Layer 1



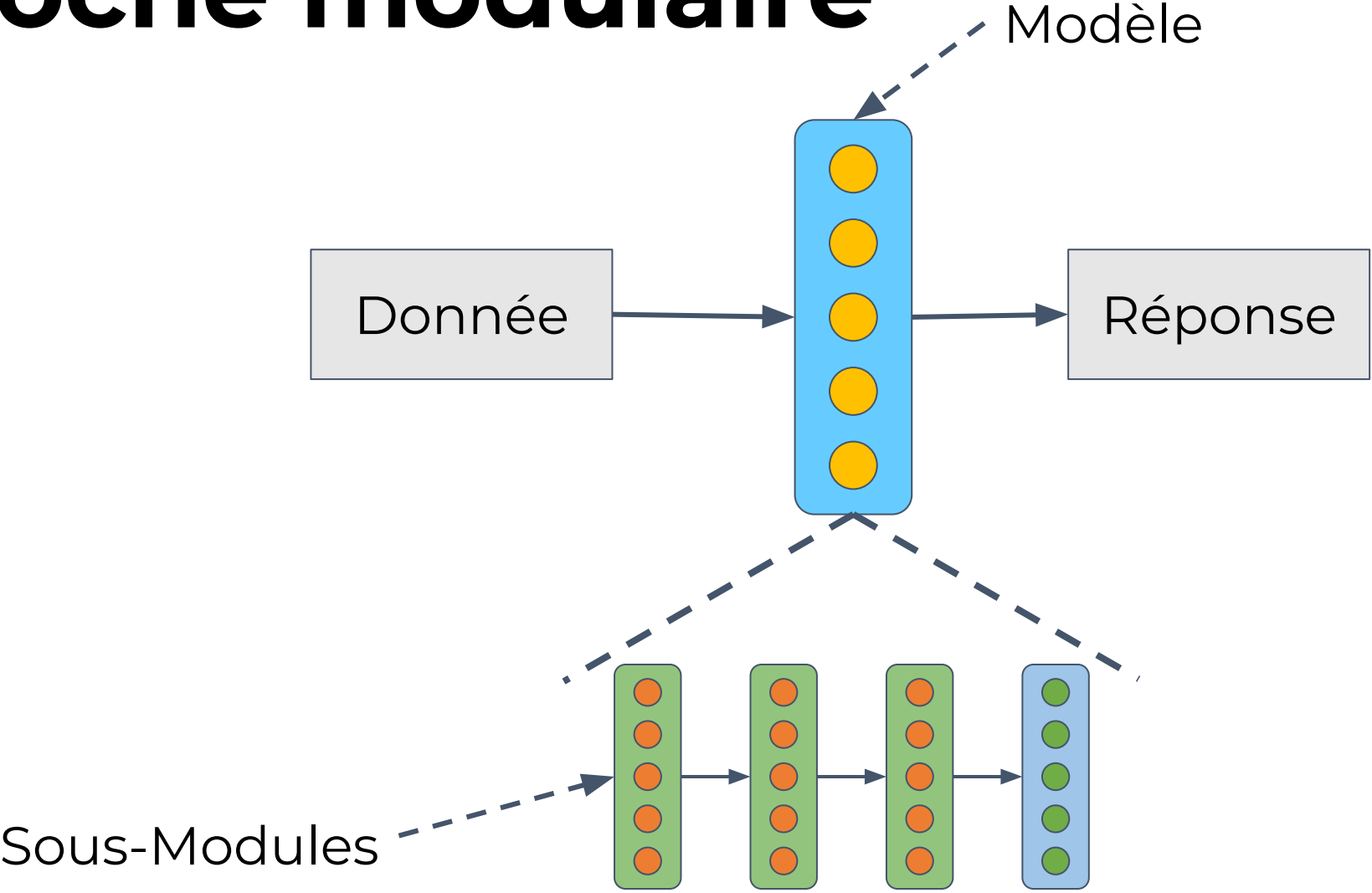
Principe de compositionnalité



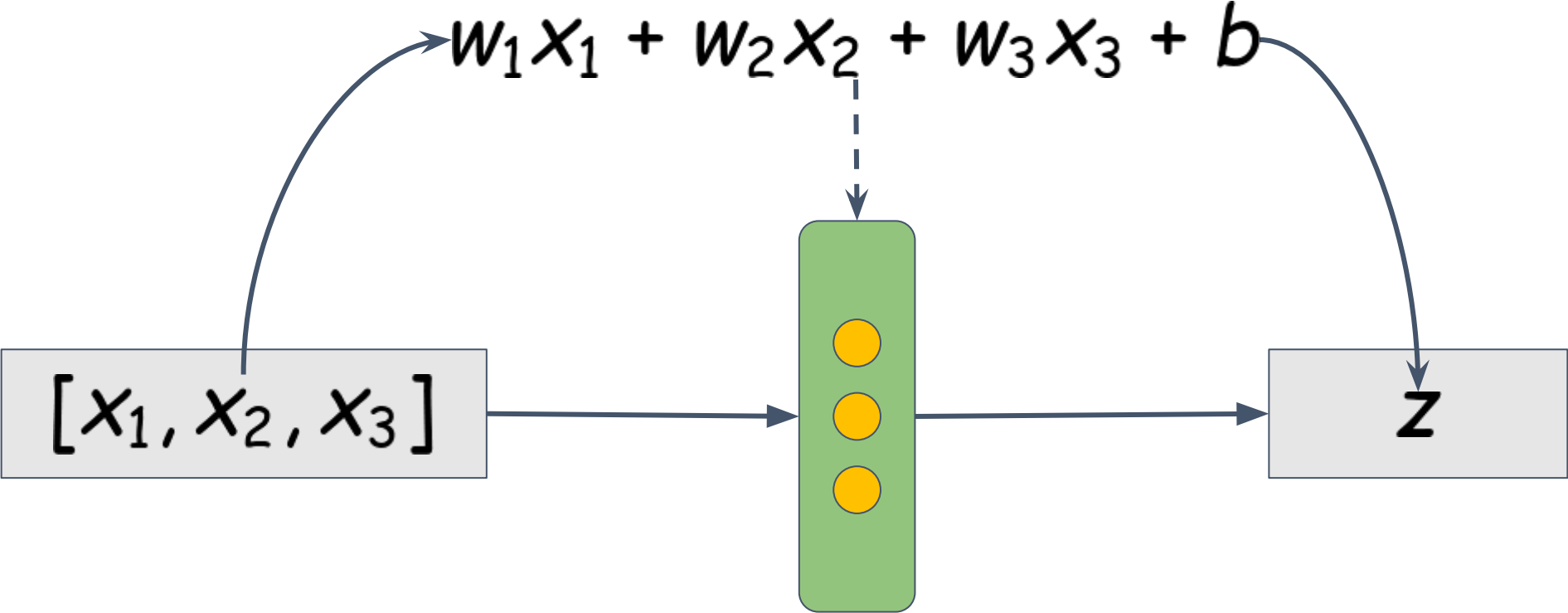
Ingrédients au DL

- Une **quantité importante** de données
- Graphe computationnel **différentiable** de bout en bout
- Un algorithme de calcul de gradient: **backpropagation**
- Des fonctions d'activation **non-linéaires**
- Un **optimiseur** itératif de paramètres

Approche modulaire

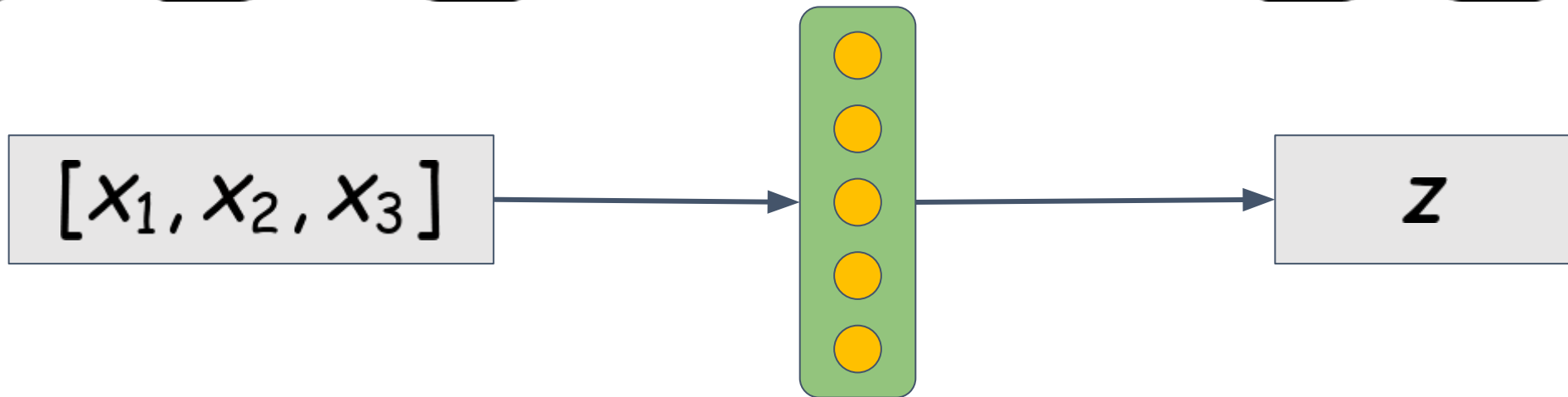


Transformation linéaire

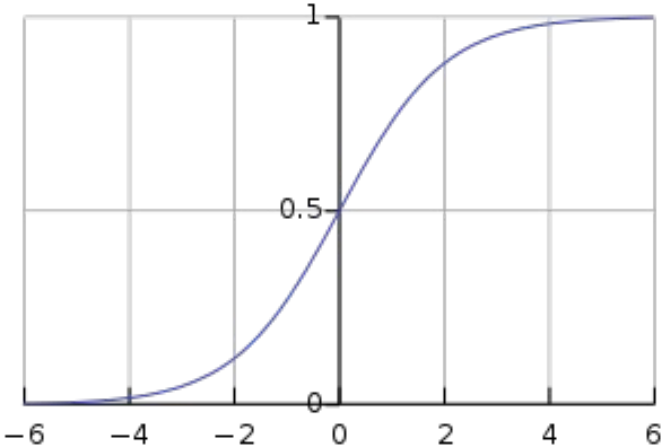
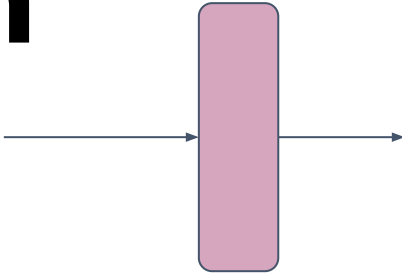


Transformation linéaire

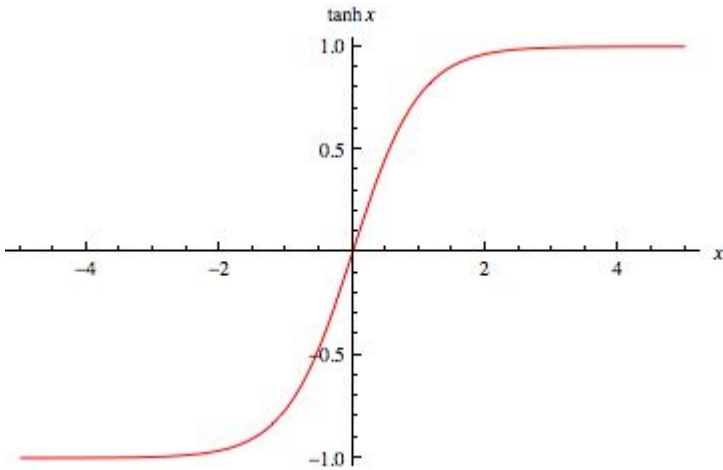
$$\begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & w_{13} & b_1 \\ w_{21} & w_{22} & w_{23} & b_2 \\ w_{31} & w_{32} & w_{33} & b_3 \\ w_{41} & w_{42} & w_{43} & b_4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ 1 \end{bmatrix}$$



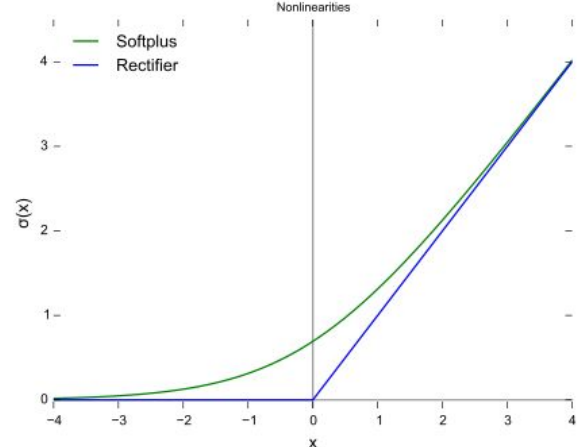
Fonctions d'activation



Sigmoid



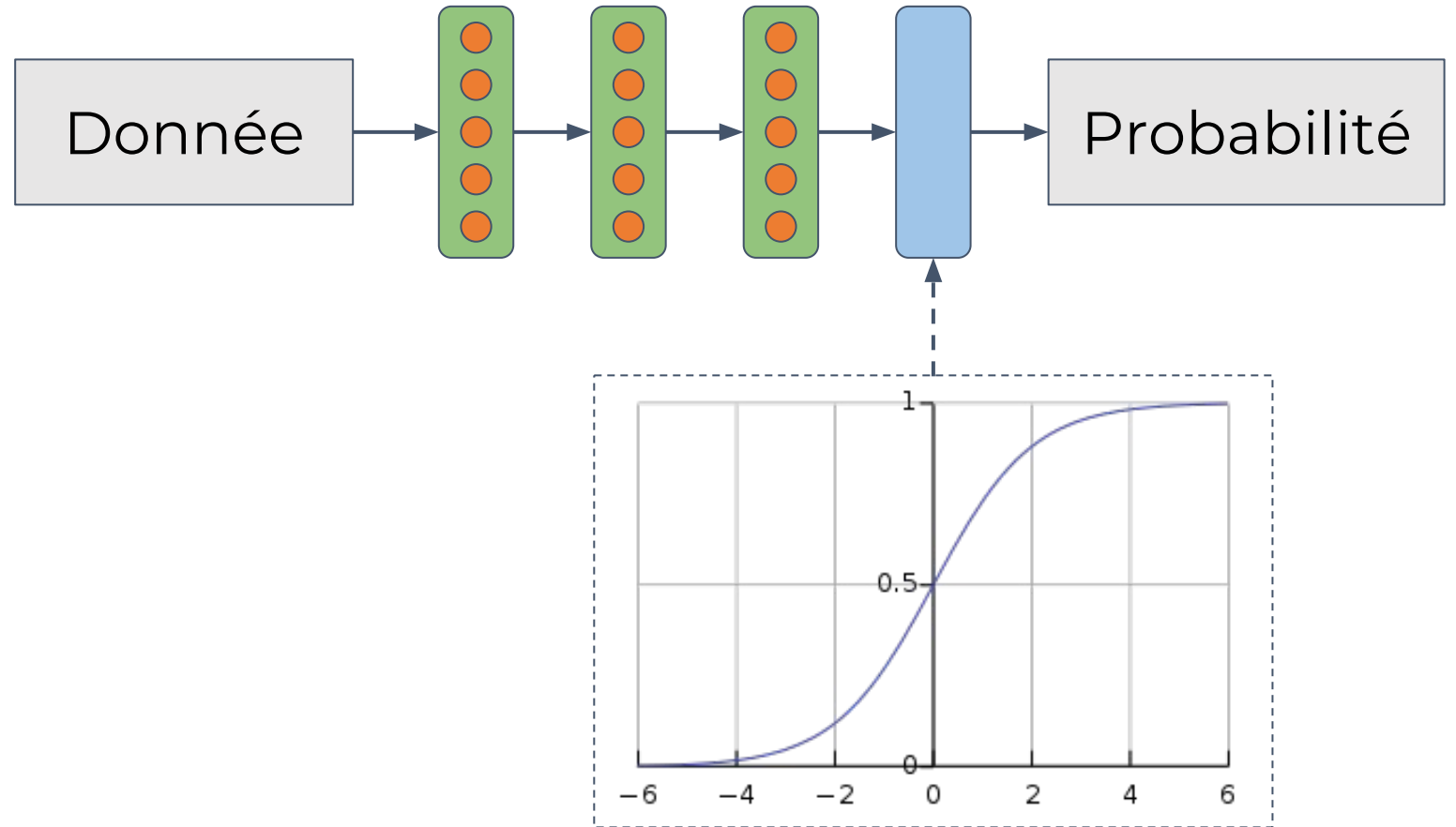
Tanh



ReLU

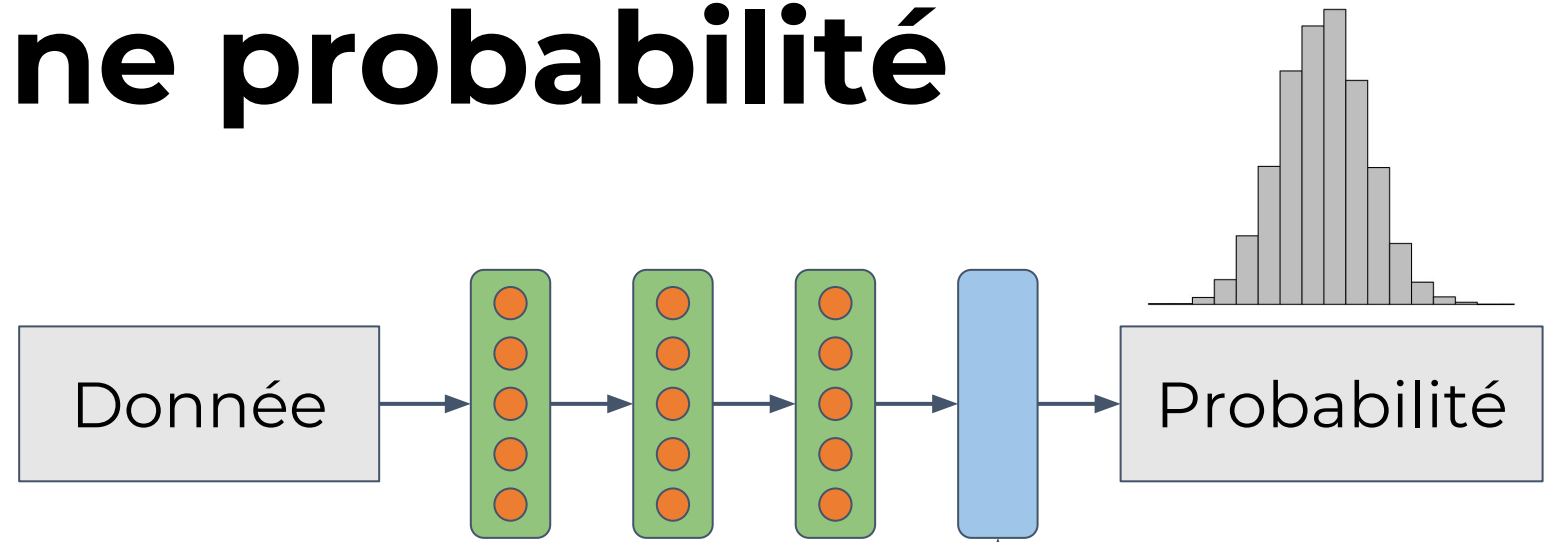
Modéliser une probabilité

- **Sigmoid**
- Softmax
- Modèle gaussien



Modéliser une probabilité

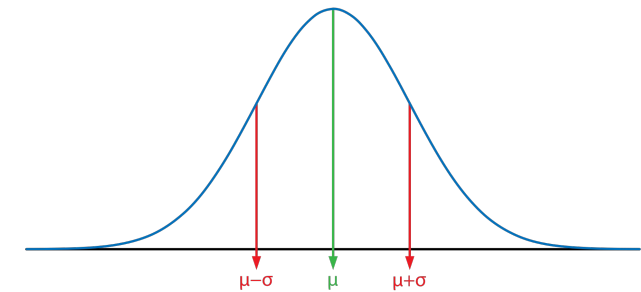
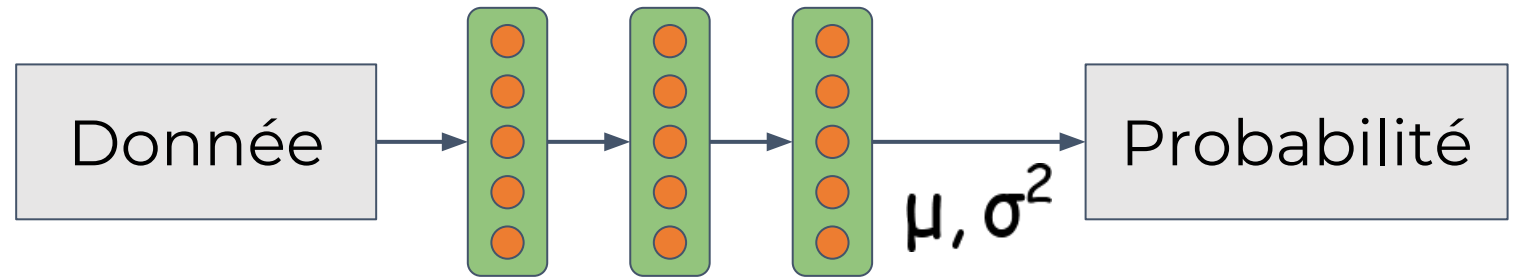
- Sigmoid
- **Softmax**
- Modèle gaussien



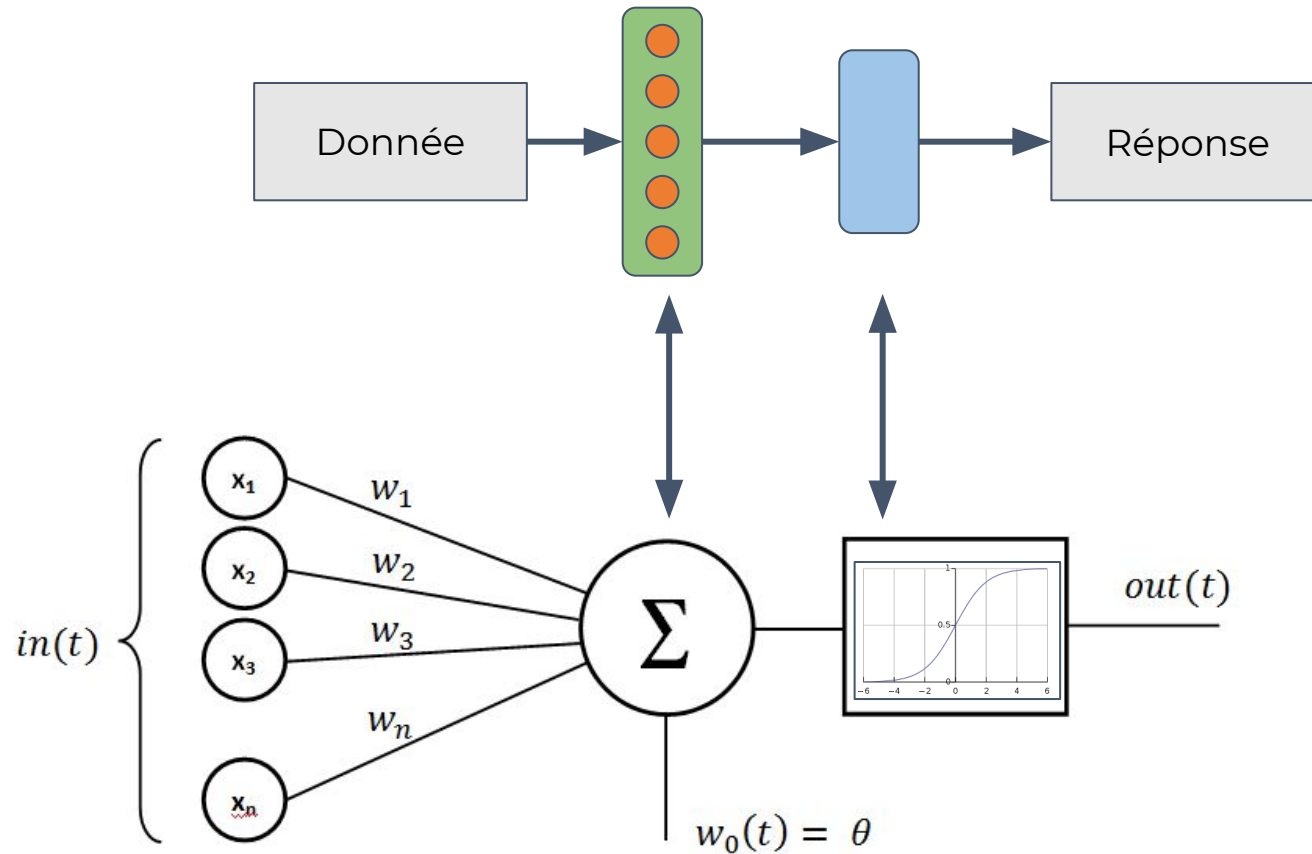
$$p_i = \frac{\exp(z_i)}{\sum_j \exp(z_j)}$$

Modéliser une probabilité

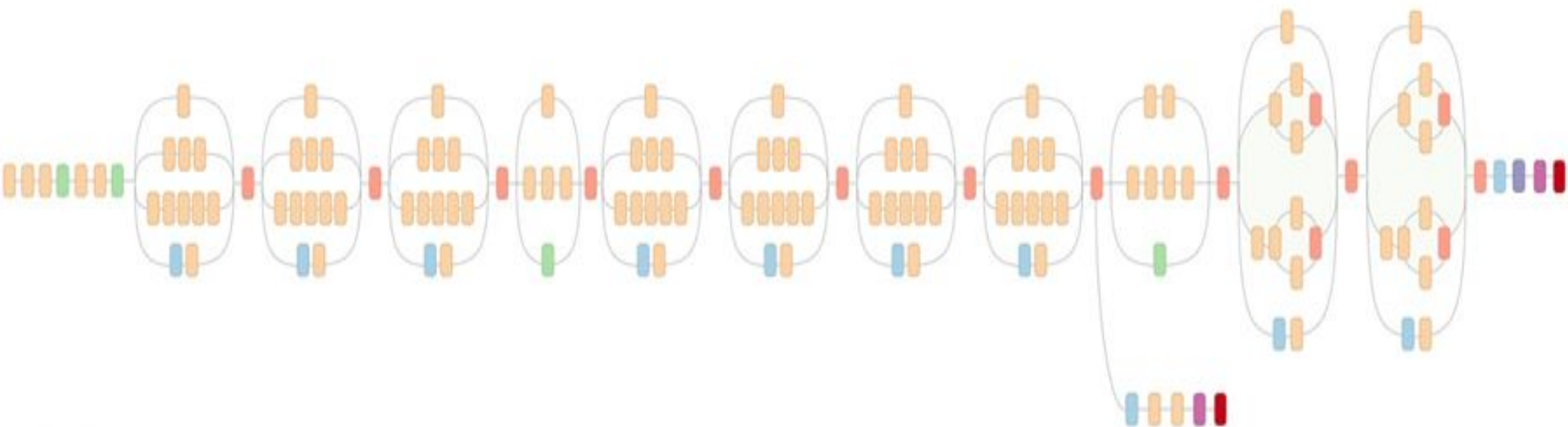
- Sigmoid
- Softmax
- **Modèle gaussien**



Exemple: régression logistique



Example: Inception V1

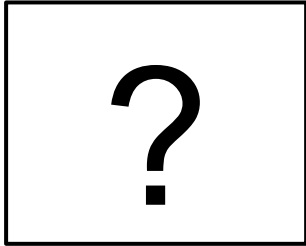
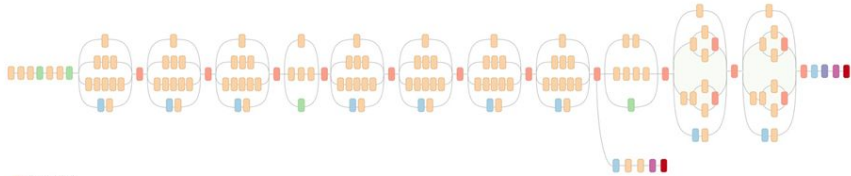


State of the science

Empirisme

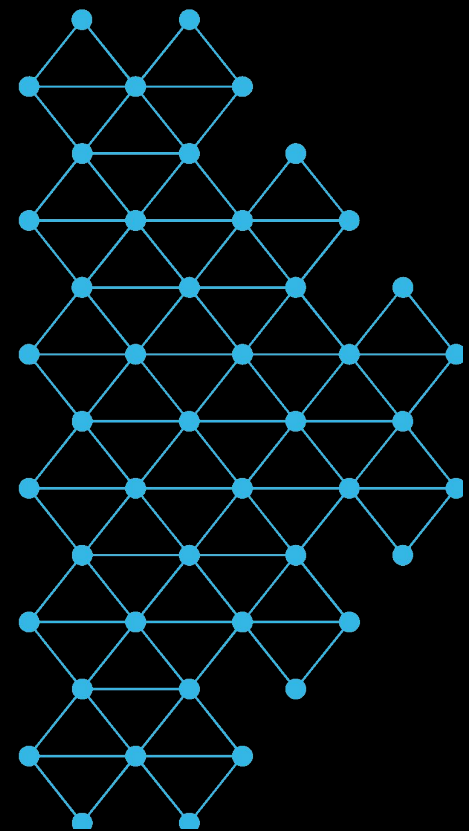
Théorie

DL



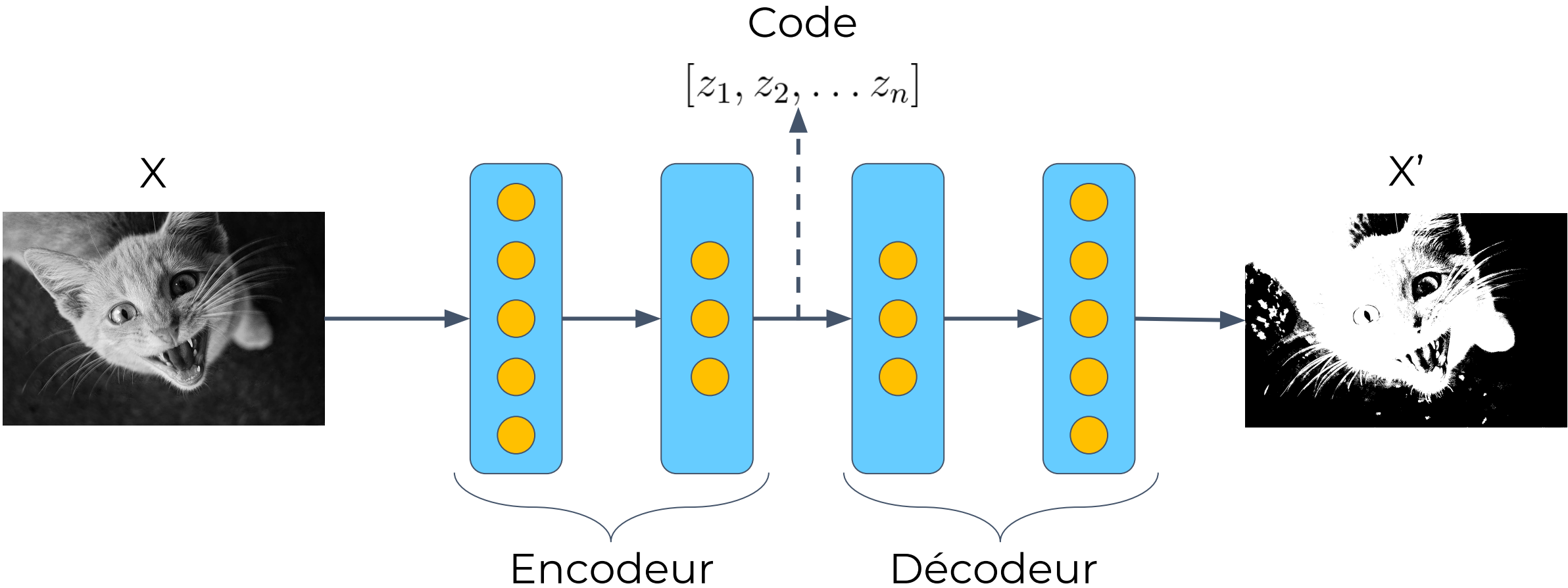
Aviation



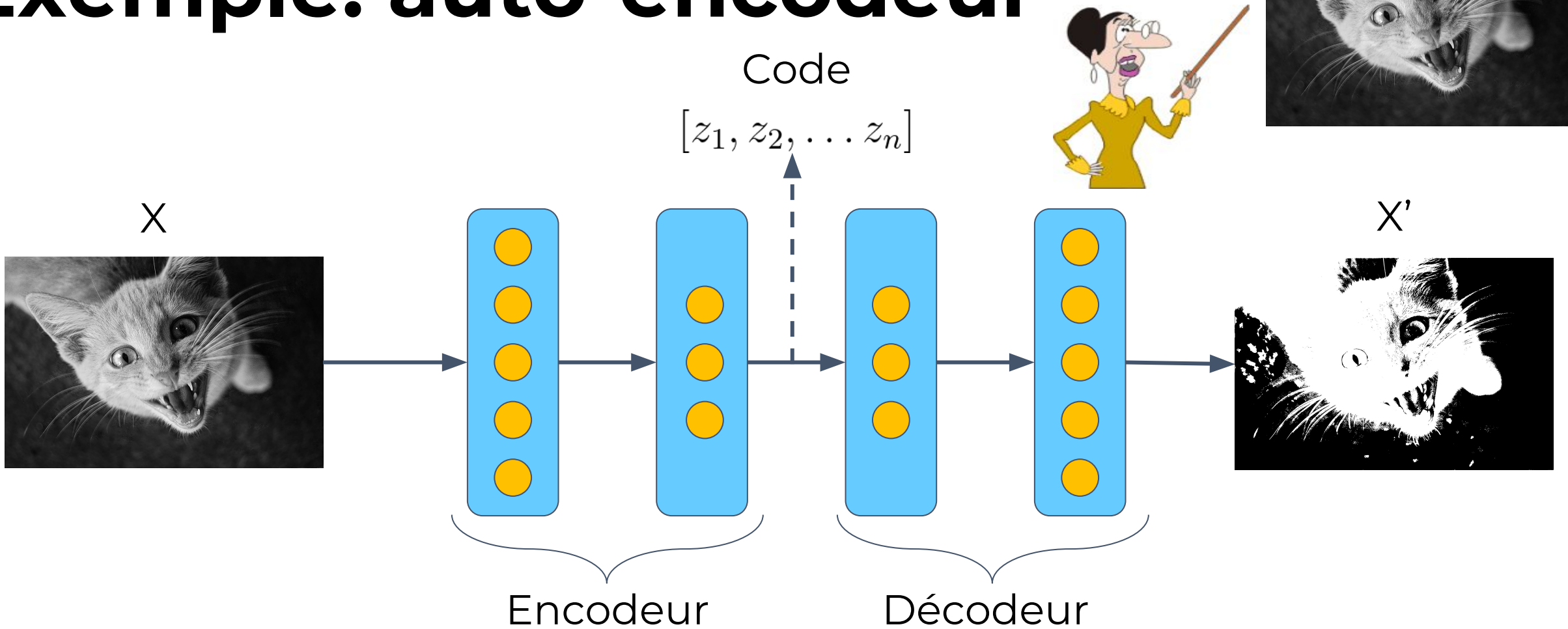


Exemples

Exemple: auto-encodeur



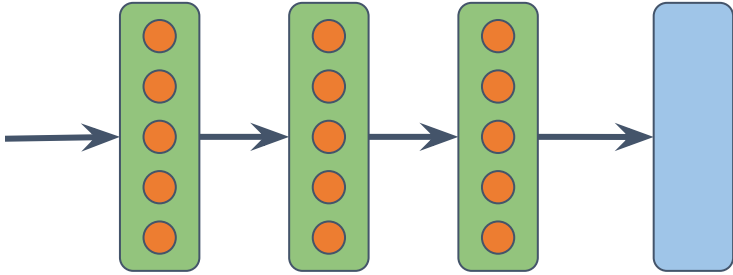
Exemple: auto-encodeur



Transfer learning



(a) MNIST



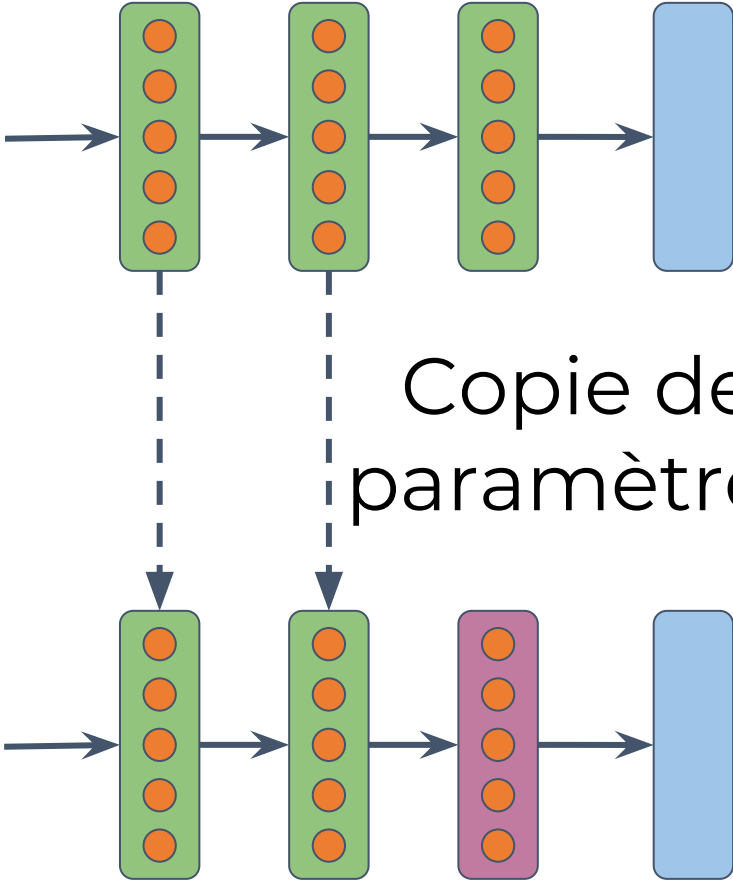
Transfer learning



(a) MNIST



(b) SVHN

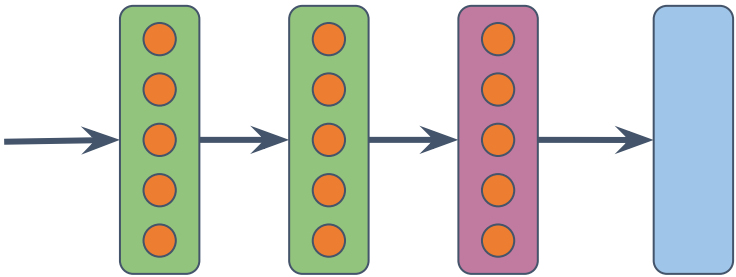


Transfer learning

On entraîne tout le réseau



(b) SVHN



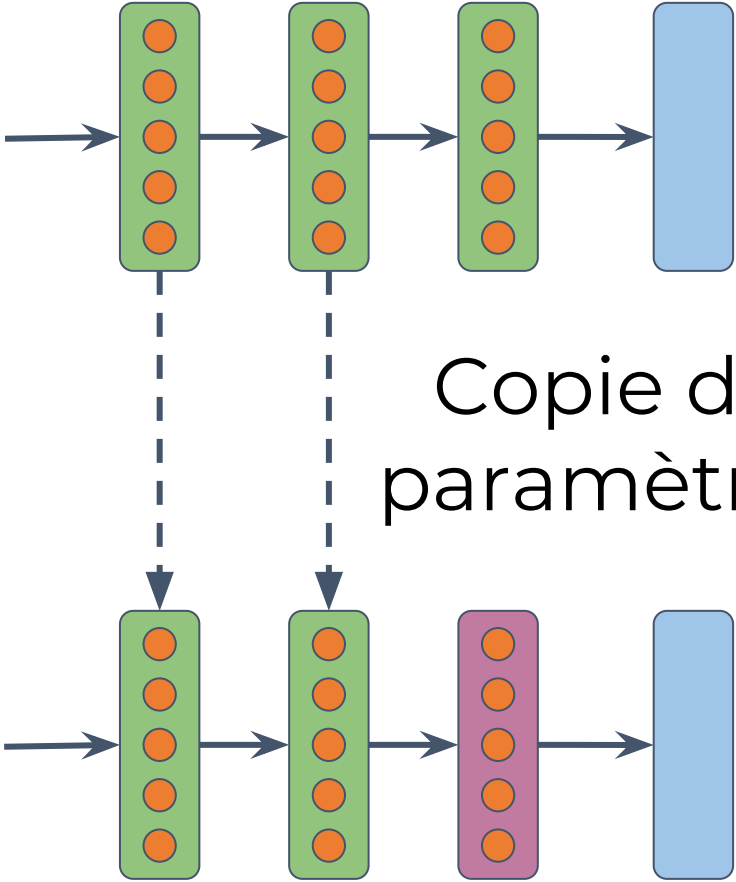
Transfer learning



(a) MNIST



(b) SVHN

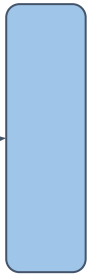
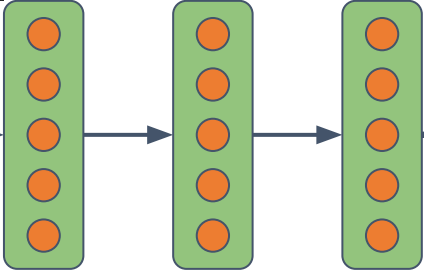


Copie de paramètres

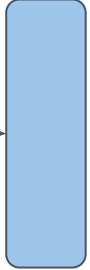
Multi-task learning



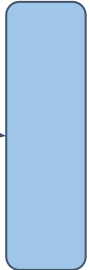
Donnée



Réponse1



Réponse2



Réponse3



$y_1 = \text{Chat}$



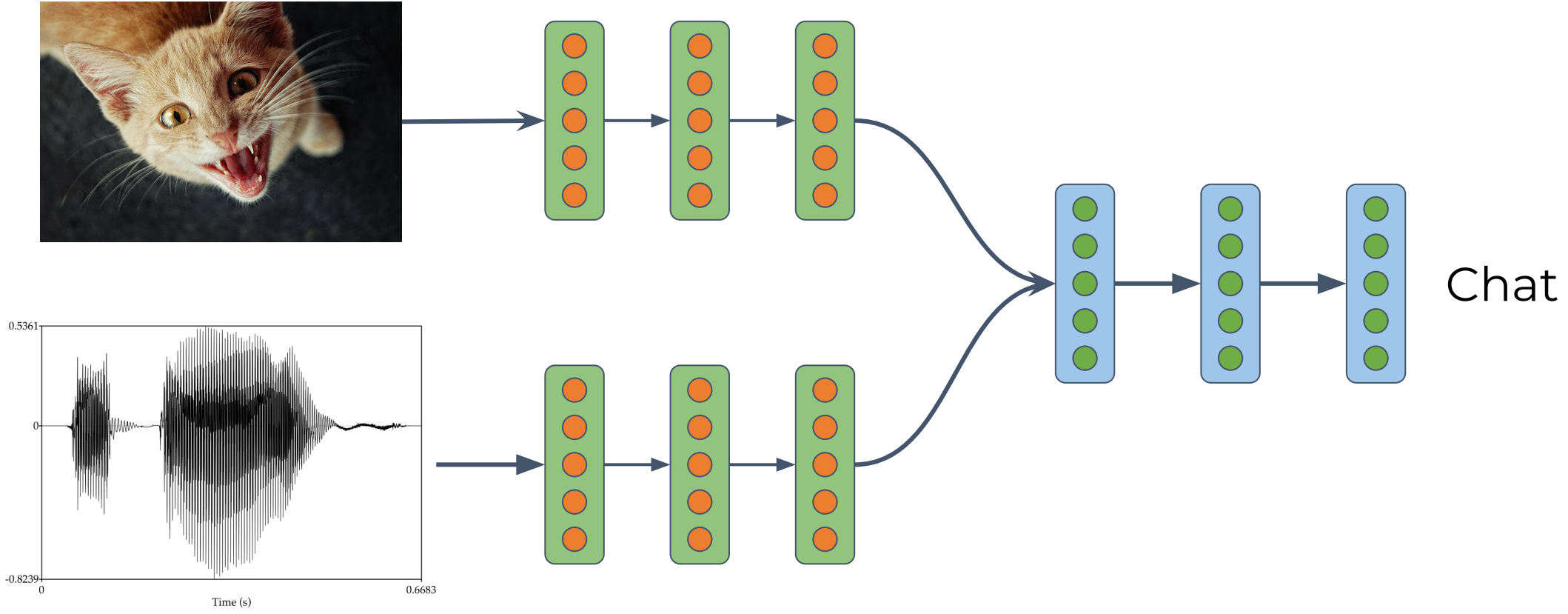
$y_2 = \text{Jaune}$



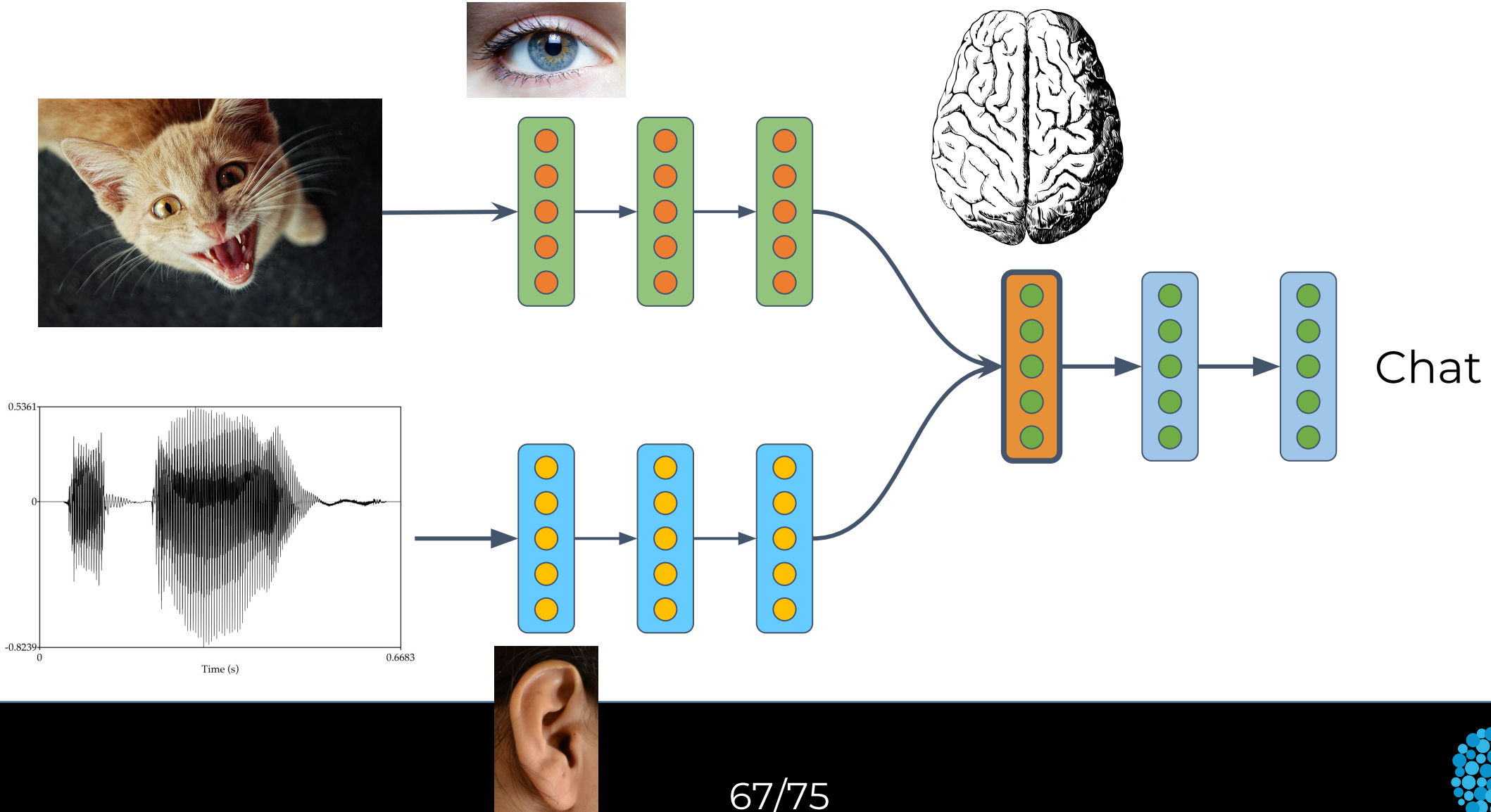
$y_3 = \text{Danger}$



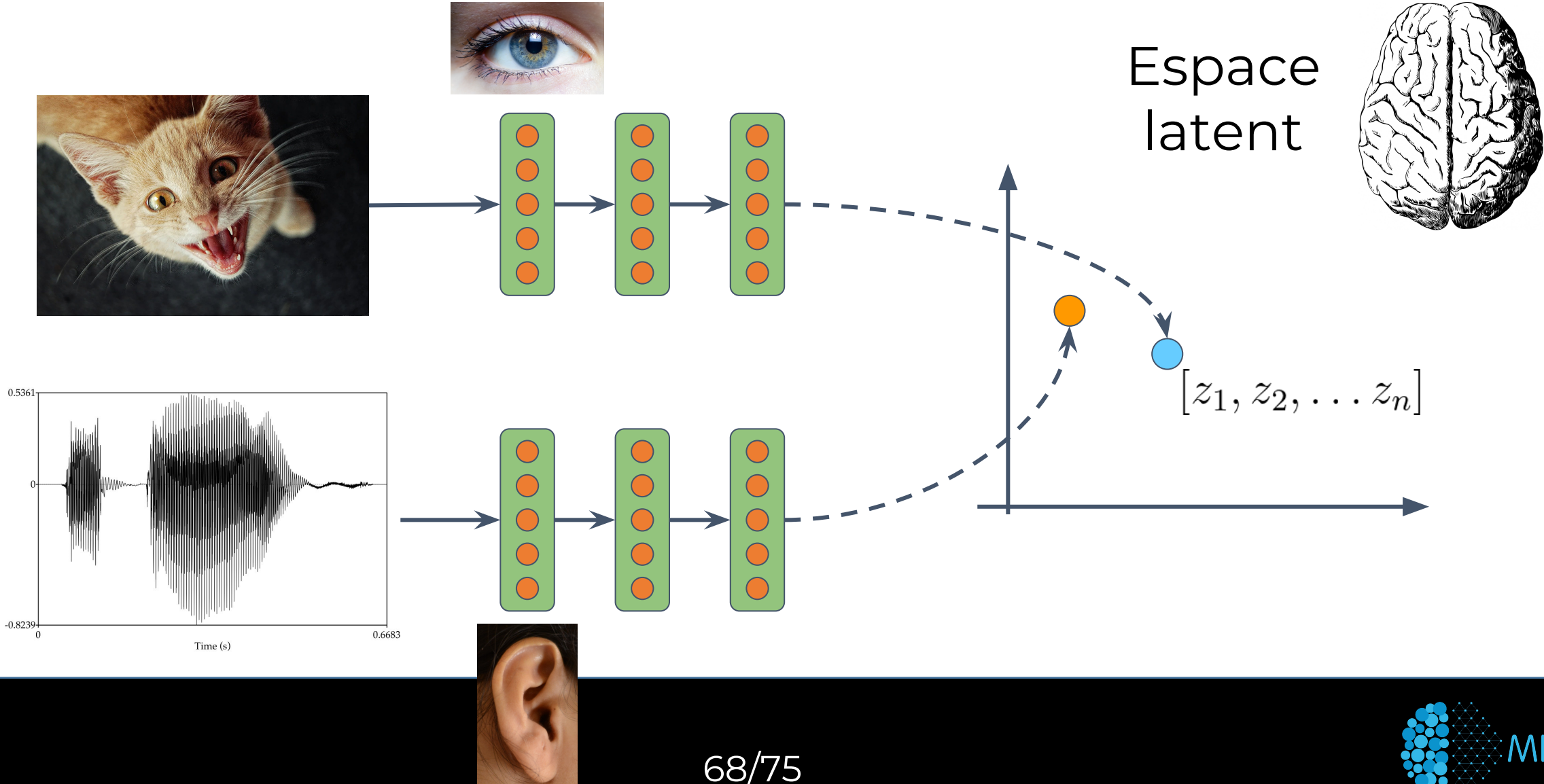
Représentation latente commune



Représentation latente commune



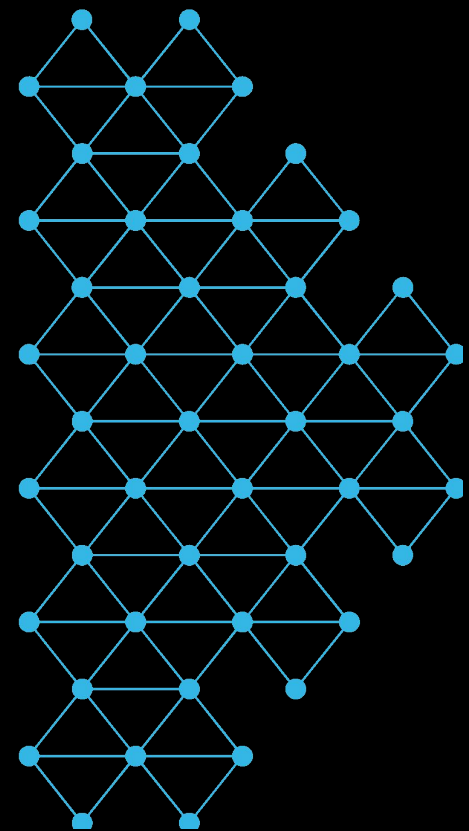
Représentation latente commune



Références

- Leçon inaugurale de Stéphane Mallat au Collège de France: <https://goo.gl/7Lq3q6>
- 3Blue1Brown: But what *is* a Neural Network? | Chapter 1, deep learning: <https://goo.gl/BMRoPo>

Merci pour votre
attention!



Contacts

<https://mila.quebec/>

Gaétan Marceau Caron, Membre MILA R&D,
gaetan.marceau.caron@rd.mila.quebec

Myriam Côté, Directrice équipe MILA R&D,
myriam.cote@rd.mila.quebec



Merci pour votre
attention!

