UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE
ÉCOLE DE MANAGEMENT
DE LA SORBONNE

# Introduction aux Réseaux de Neurones

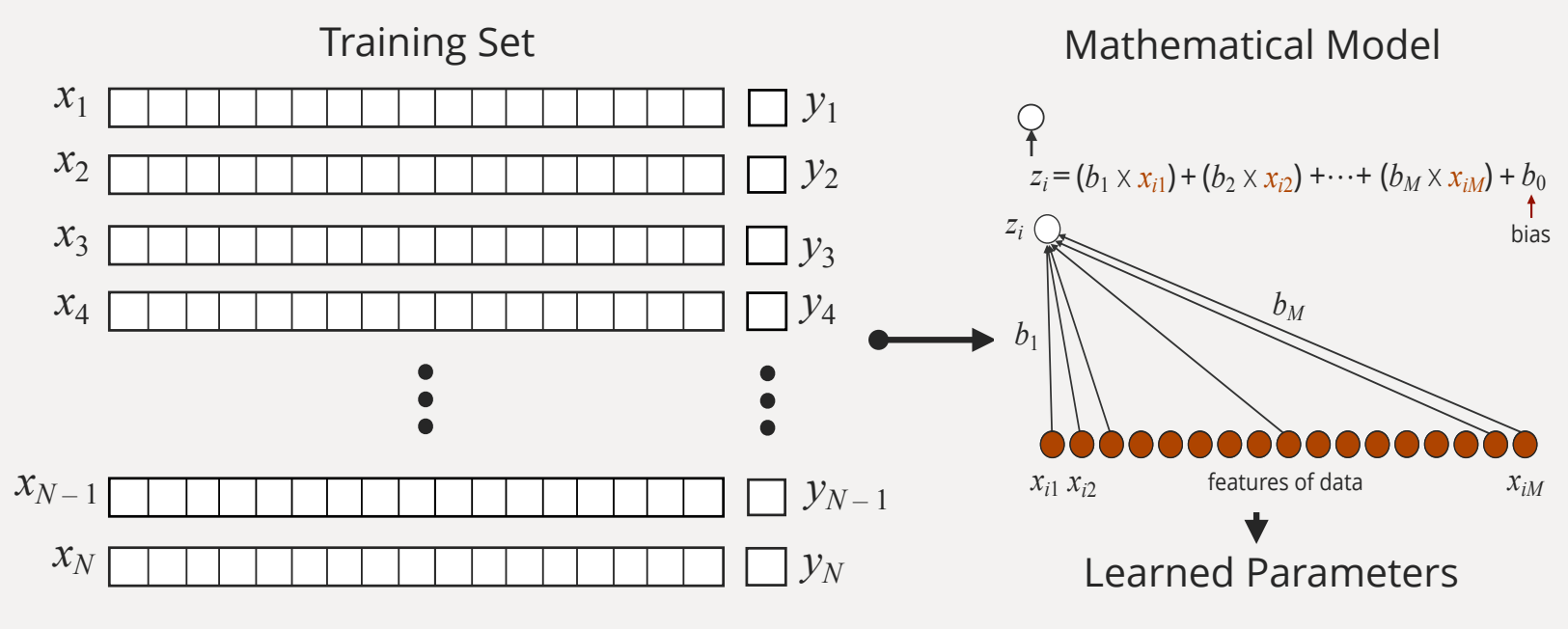
Fichiers sur

https://github.com/mkirschpin/CoursPython

http://kirschpm.fr/cours/PythonDataScience/

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE
ÉCOLE DE MANAGEMENT
DE LA SORBONNE

# De la régression linéaire aux neurones




Ici, la somme pondérée des features donne une estimation linéaire de la valeur attendue

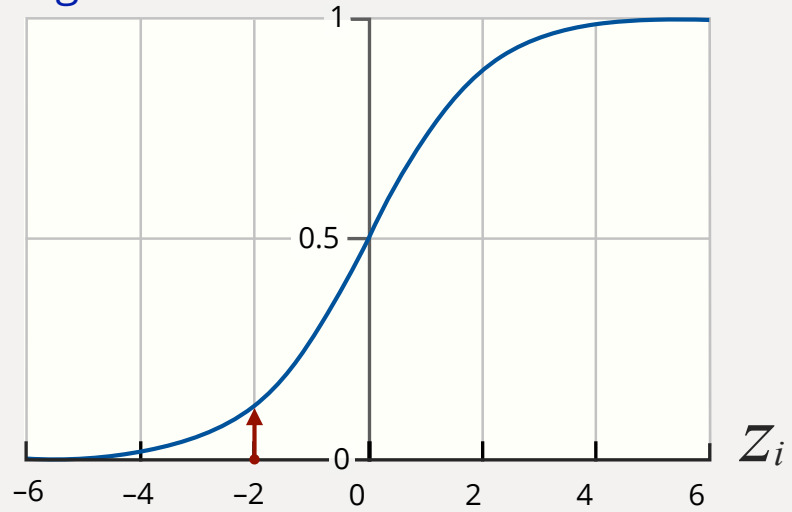
UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE
ÉCOLE DE MANAGEMENT
DE LA SORBONNE

# Et si on utilisait la régression pour la classification ?

- Au lieu d'un sortie linéaire, je cherche à établir une classification
- **Quelle est la probabilité d'appartenir à une classe cible ?**
- La fonction sigmoide - **σ(*zi*)**
- Représente les prédictions selon un point de vue probabiliste
  - Plus grand et positif est *zi*, plus probable sera la valeur *yi*

$$z_i = (b_1 \times x_{i1}) + (b_2 \times x_{i2}) + \cdots + (b_M \times x_{iM}) + b_0$$

Sigmoid Function $p(y_i = 1|x_i) = \sigma(z_i)$

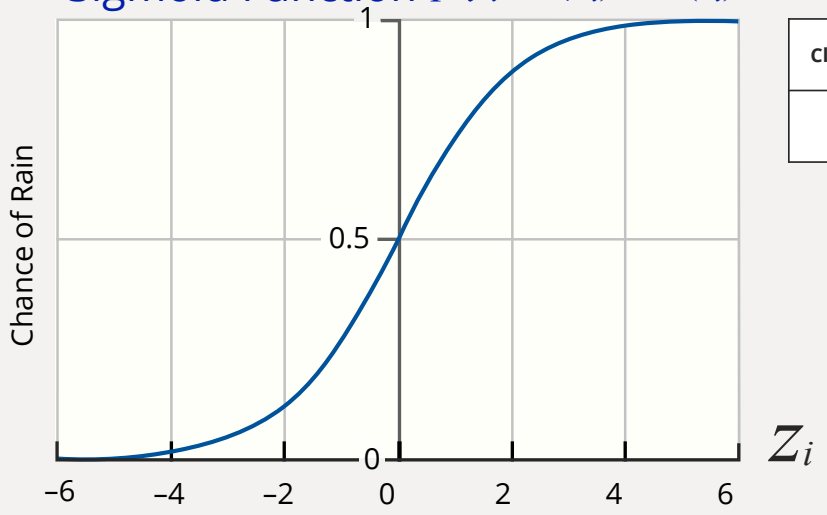


UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE
ÉCOLE DE MANAGEMENT
DE LA SORBONNE

# Exemple

$$z_i = (b_1 \times 0.5) + (b_2 \times 0.8) + (b_3 \times 75) + (b_4 \times 1.2) + b_0$$

Sigmoid Function $p(y_i = 1|x_i) = \sigma(z_i)$



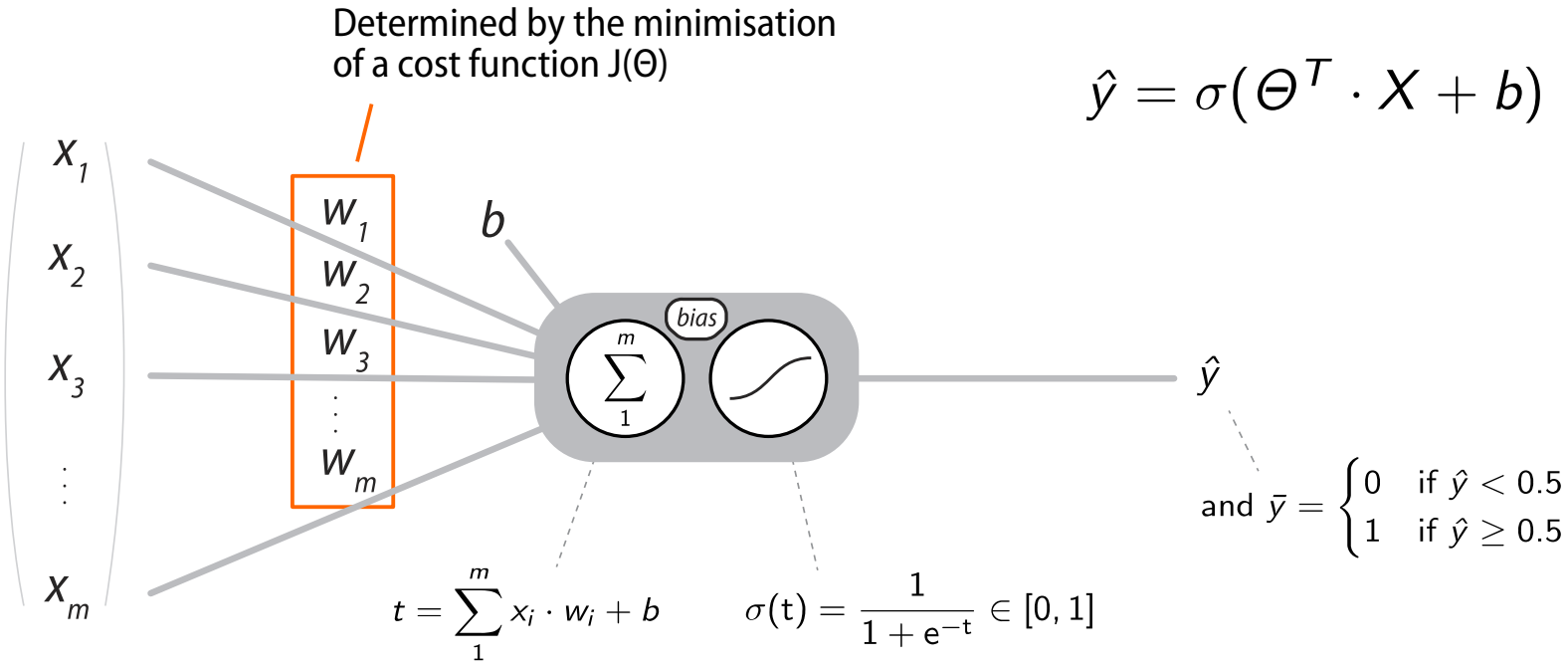

features

| Cloud Cover | Humidity | Temperature | Air Pressure |
|---|---|---|---|
| 0.5 | 80% | 75 | 1.2 |

- Les paramètres *b* indiquent combien les variables sont importantes pour la prédiction

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE
ÉCOLE DE MANAGEMENT
DE LA SORBONNE

# Schéma d'une Régression Logistique



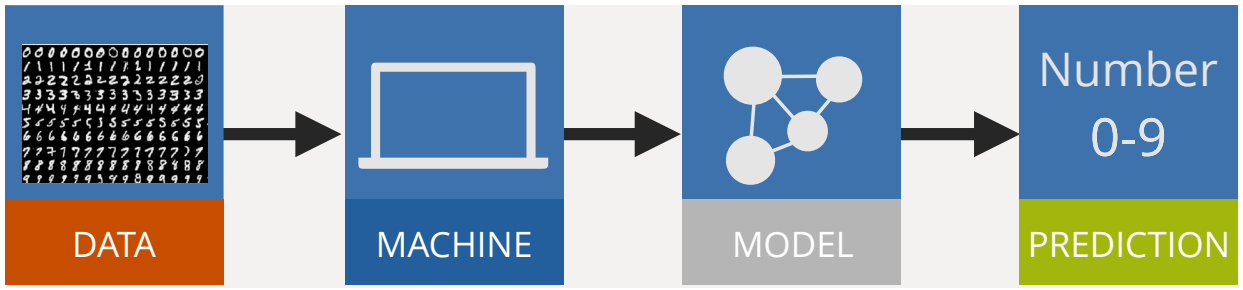

| Input | Bias / Weight | Activation function | Output |
|---|---|---|---|
| $X$ | $\Theta$ | $\sigma(t)$ | $\hat{y}$ |

Surprise !!! Ceci est le schéma d'un **neurone artificiel** (perceptron). On a **un réseau d'un neurone** !!

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE
ÉCOLE DE MANAGEMENT
DE LA SORBONNE

# Exemple d'usage pratique

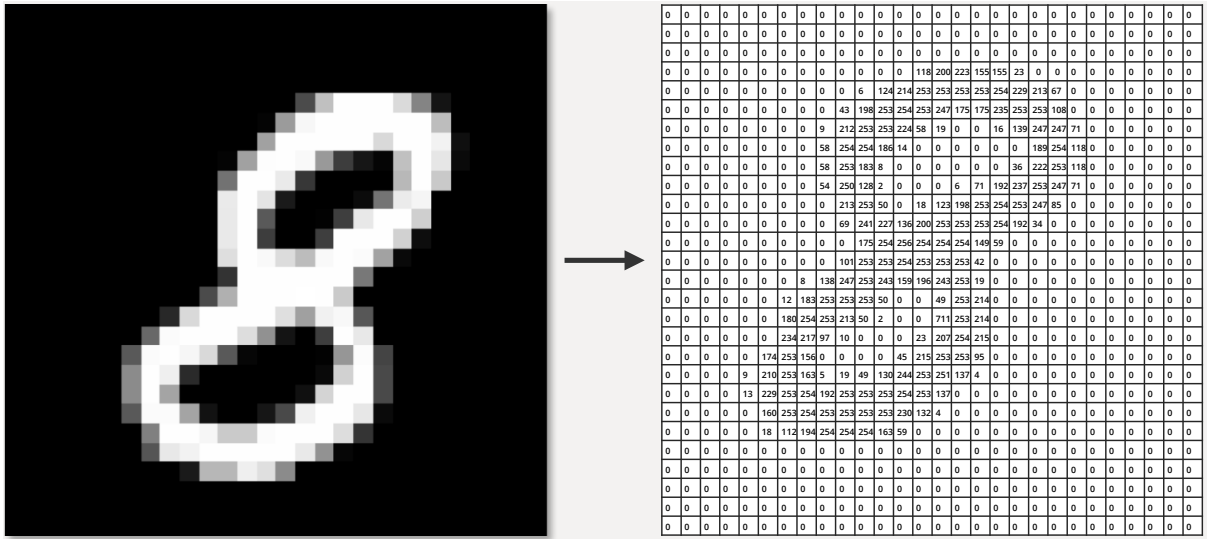
- Exemple : le dataset MNIST
  - Ensemble de chiffres écrits à la main
- Objectif : identifier le chiffre
  - 0 à 9



UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE
ÉCOLE DE MANAGEMENT
DE LA SORBONNE

# Mise en route

- Les chiffres sont des images, chaque pixel a une valeur numérique



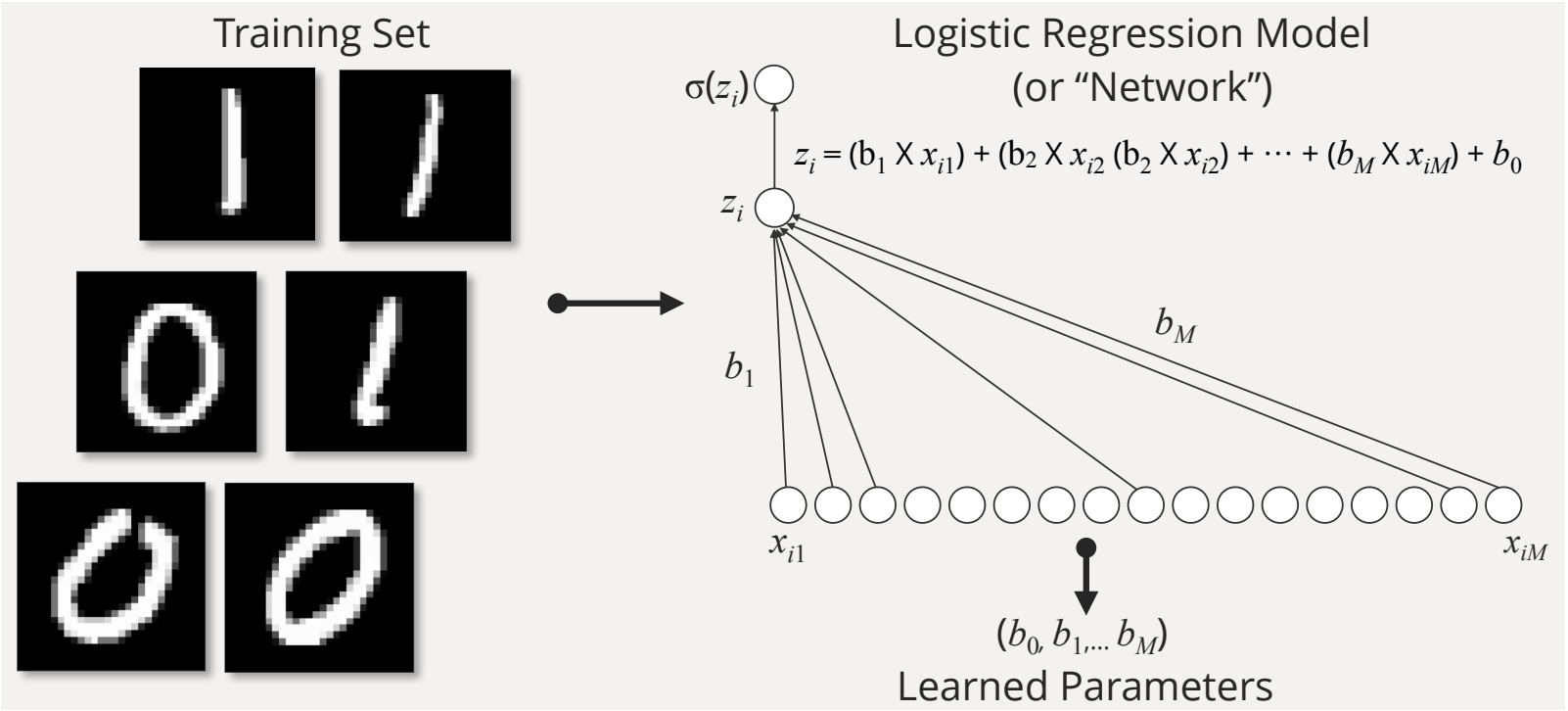
= X

**MNIST Dataset of Handwritten Digits (Images)**

Yann LeCun (Courant Institue, NYU) and Corinna Cortes (Google Labs, New York) CC-by-SA 3.0
http://yann.lecun.com/exdb/mnist/

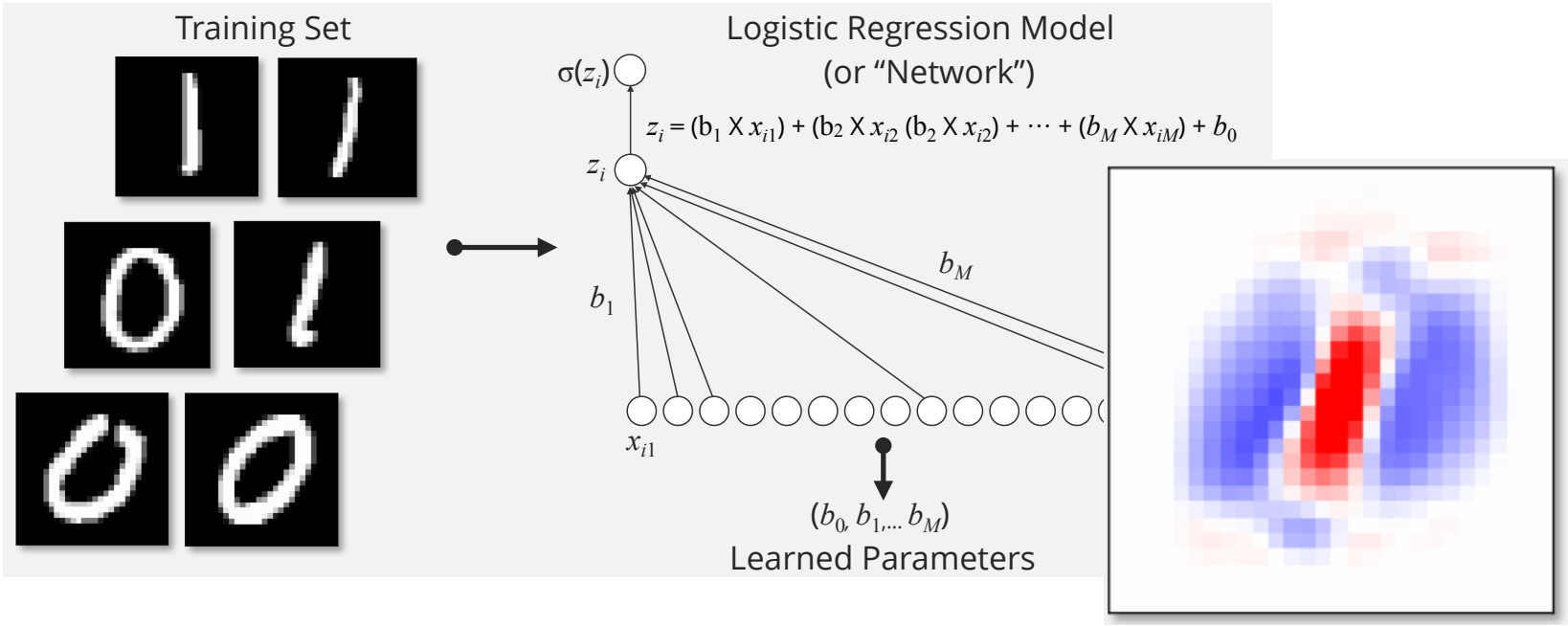
UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE
ÉCOLE DE MANAGEMENT
DE LA SORBONNE

# Apprentissage du MNIST




Chaque position de la matrice est étiquetée avec une valeur négative (0), positive (1) ou rien. Plus on a des 0 ou 1, plus la "case" est fortement marquée (poids de *b*)

# Apprentissage du MNIST

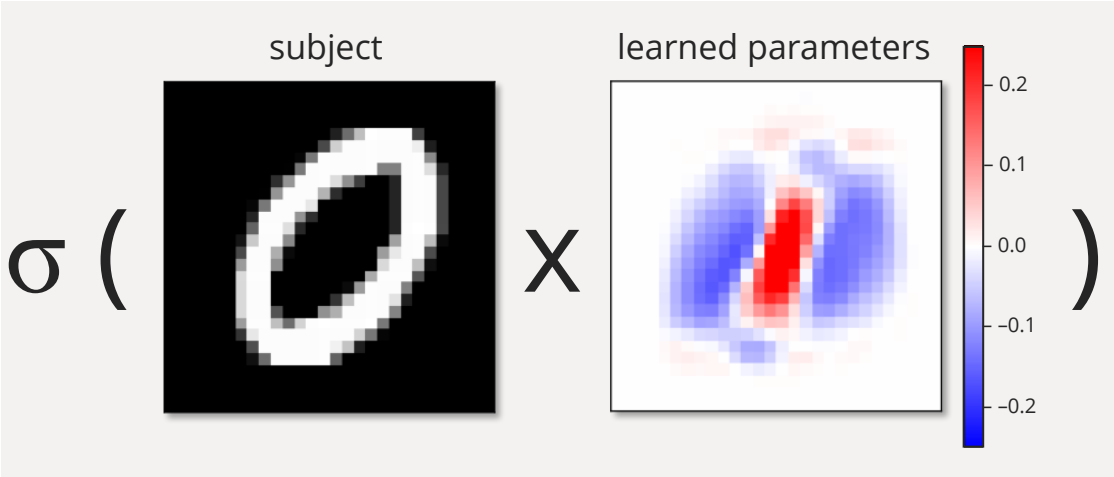



Chaque position de la matrice est étiquetée avec une valeur négative (0), positive (1) ou rien. Plus on a des 0 ou 1, plus la "case" est fortement marquée (poids de *b*)

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE
ÉCOLE DE MANAGEMENT
DE LA SORBONNE

# Quelle est la probabilité d'être un 1 ?

- Pour un chiffre donné, on compare sa matrice avec la matrice entraînée :

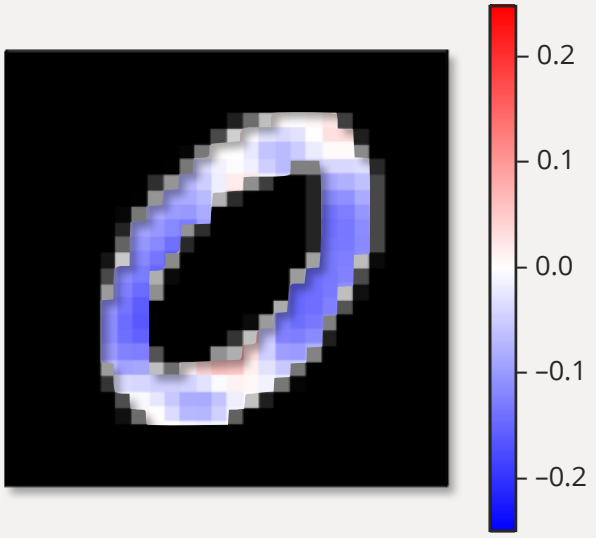


UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE
ÉCOLE DE MANAGEMENT
DE LA SORBONNE

# Probabilité d'être un 1

- La "superposition" donne une note pour les parties communes

$\sigma$ (




) = 0.006

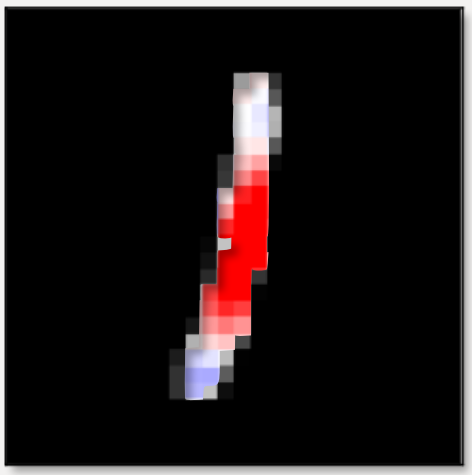
= 0.6% de chance d'être un 1

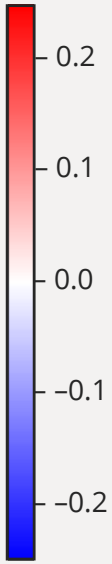
UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE
ÉCOLE DE MANAGEMENT
DE LA SORBONNE

# Probabilité d'un Un

- La "superposition" donne une note pour les parties communes

$$\sigma\left(\right) = 0.991$$



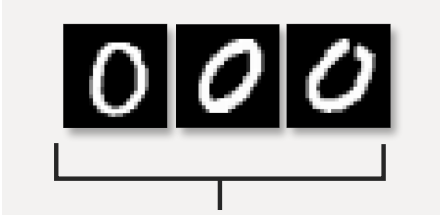



= 99,1% de chance d'être un 1

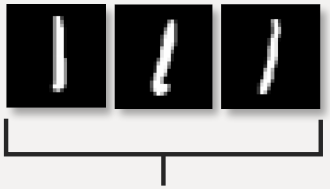
UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE
ÉCOLE DE MANAGEMENT
DE LA SORBONNE

# Interprétation du modèle

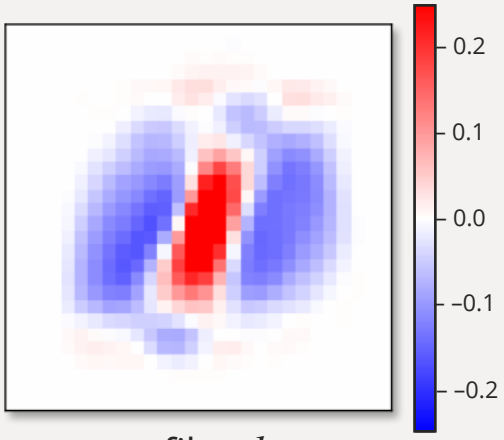
$$z_i = (b_1 \times x_{i1}) + (b_2 \times x_{i2}) + \cdots + (b_M \times x_{iM}) + b_0$$
$$= b_0 + x_i \odot b$$



examples of class $y = 0$



examples of class $y = 1$




filter $b$

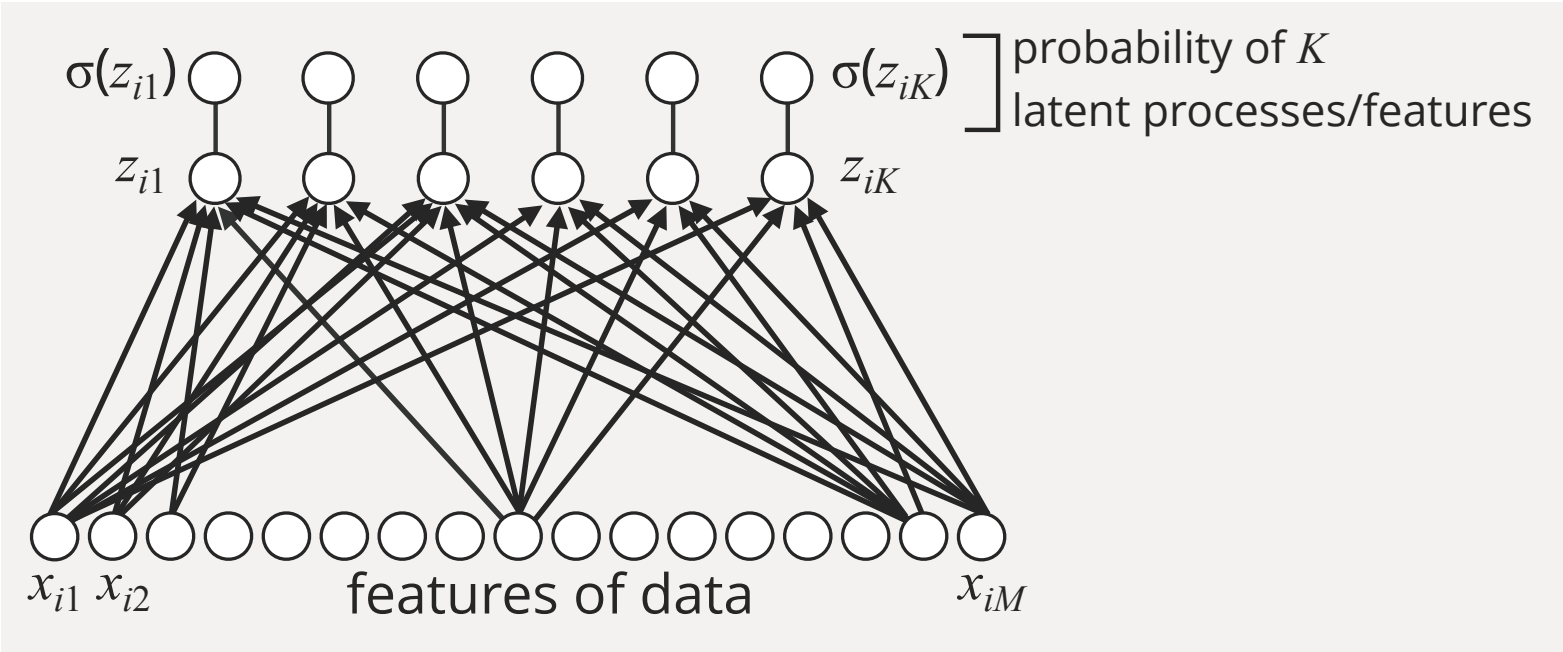
Une grande correspondance entre le filtre et les données signifie que $p(yi = 1|xi)$ est élevé

Une petite correspondance entre le filtre et les données signifie que $p(yi = 1|xi)$ est bas

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE
ÉCOLE DE MANAGEMENT
DE LA SORBONNE

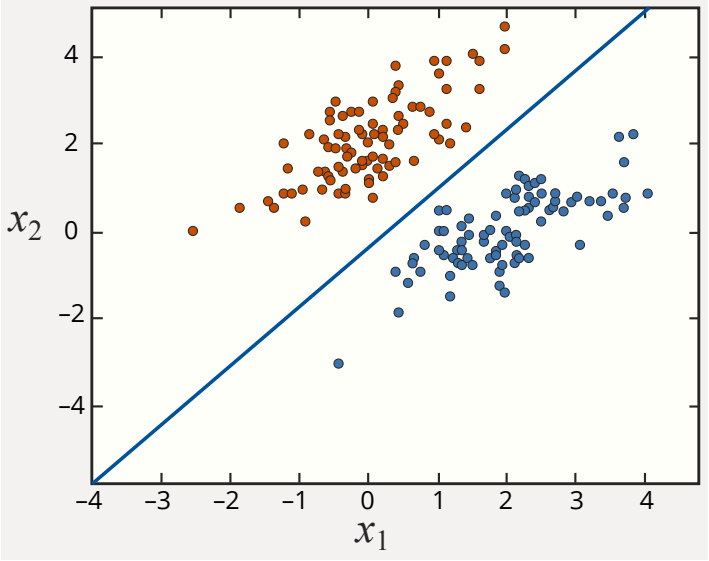
# Généralisation de la Classification - Plusieurs Classes

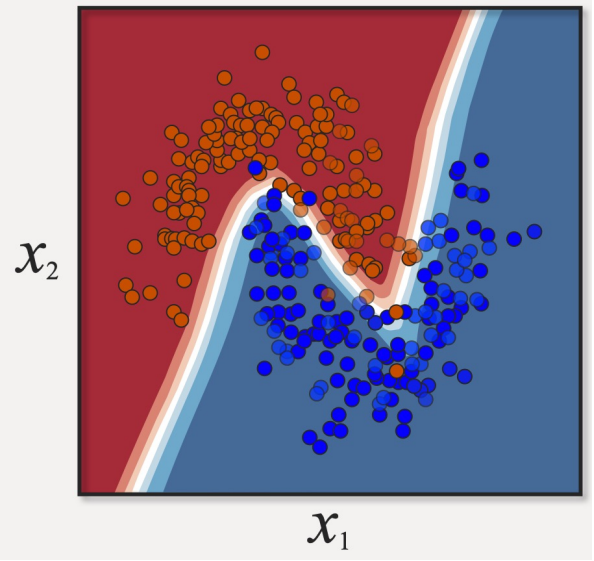
- Pour plusieurs chiffres, il faudra créer plusieurs modèles pour dire "**oui**" ou "**non**" sur chaque choix
  - On parle de "dimension de la sortie"
  - On se retrouve avec plusieurs sorties, chacune avec un % exprimé
    - La fonction *softmax()* permet de normaliser les sorties dans un intervalle [0,1]



# Limitations des réseaux à 1 couche

- Les classifieurs linéaires sont limités dans leurs possibilités
- Souvent on veut classer des données non-linéaires

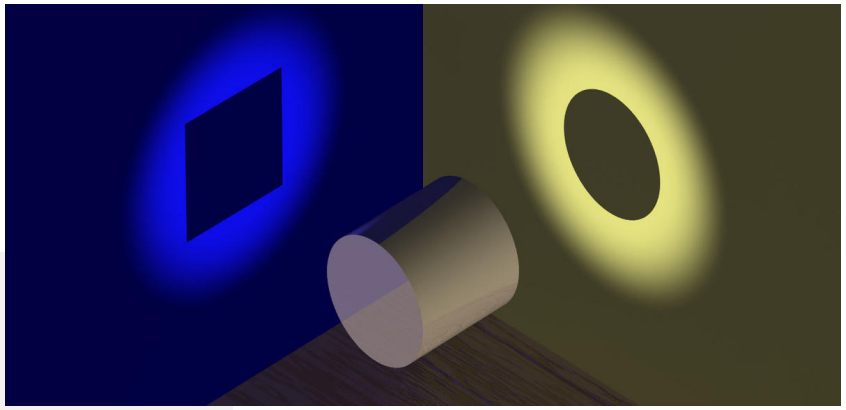


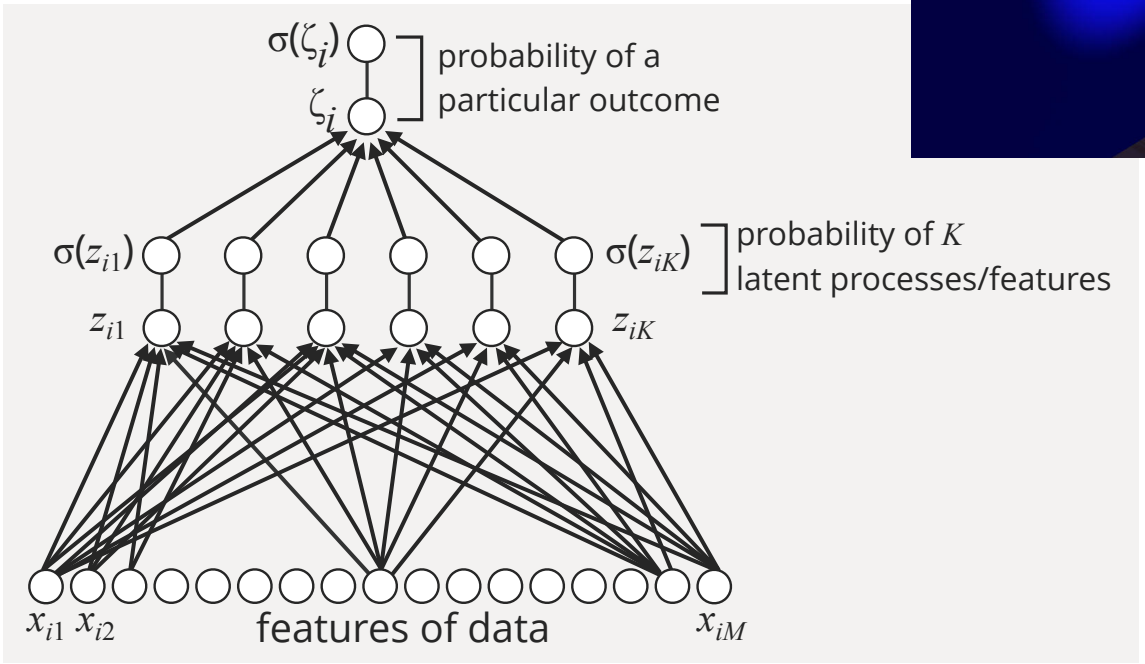



UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE
ÉCOLE DE MANAGEMENT
DE LA SORBONNE

# Réseaux Multi-couche

- Et si on faisait un 2ème niveau de neurones ?



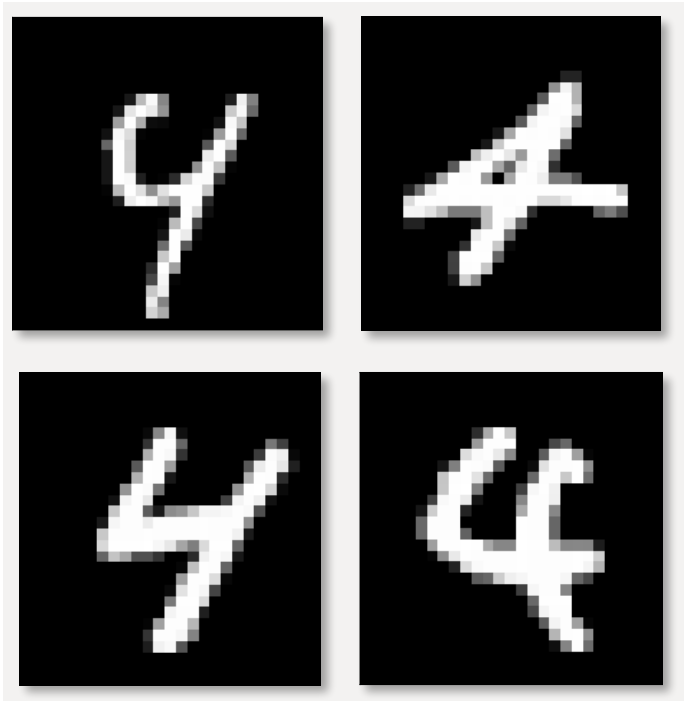



On a une classification sur K features latentes, plutôt que sur M données brutes

Chaque neurone a une vue légèrement différente des autres, et c'est l'union de leurs points de vue que mène au résultat

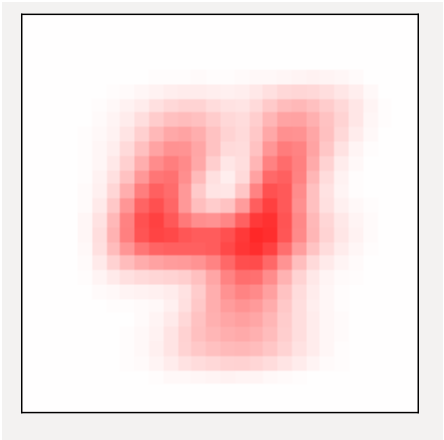
UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE
ÉCOLE DE MANAGEMENT
DE LA SORBONNE

# Exemple : mieux indentifier un 4



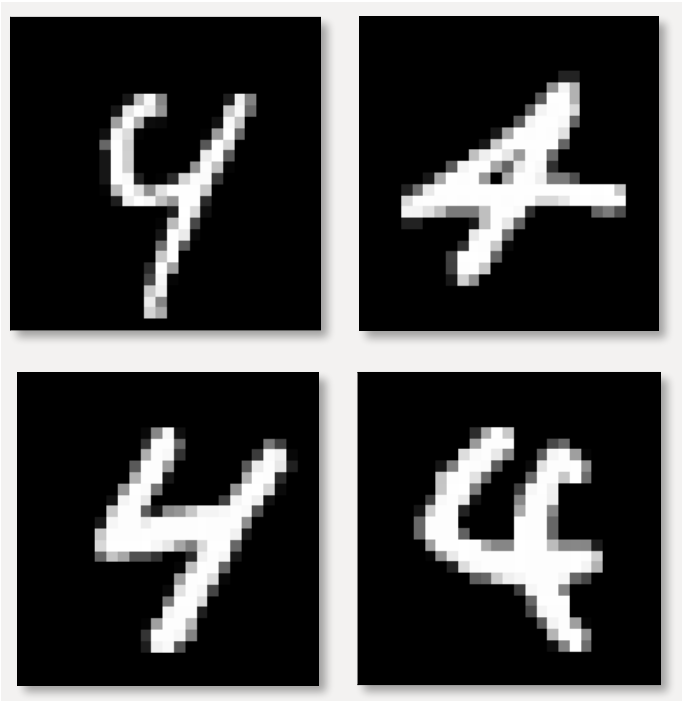
Filtre unique (shallow)

- pas très précis, adapté juste au cas moyen



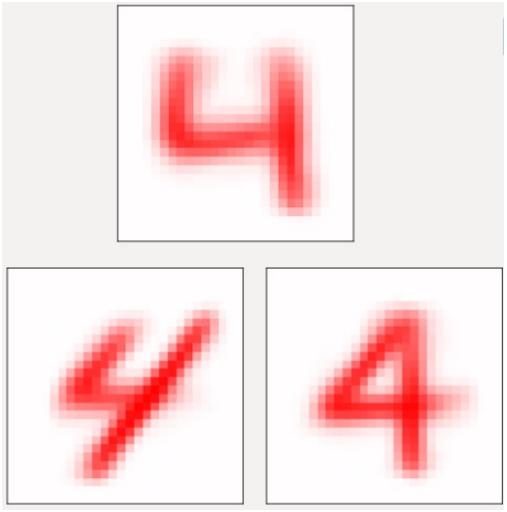
UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE
ÉCOLE DE MANAGEMENT
DE LA SORBONNE

# Plusieurs façons d'écrire 4



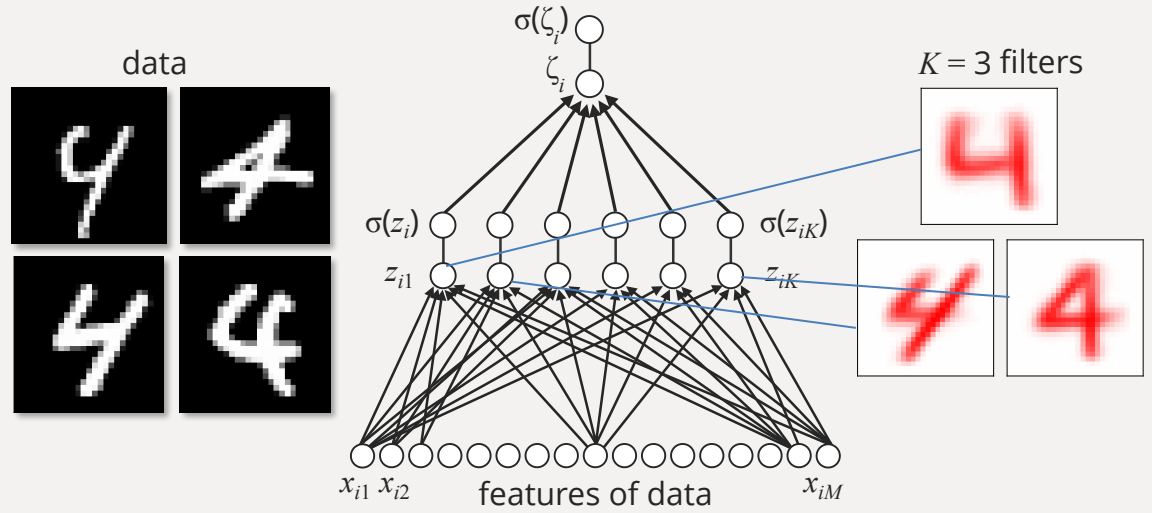
Plusieurs filtres

- Chacun est spécialisé sur une forme d'écrire 4



UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE
ÉCOLE DE MANAGEMENT
DE LA SORBONNE

# Réseau Dense de Neurones

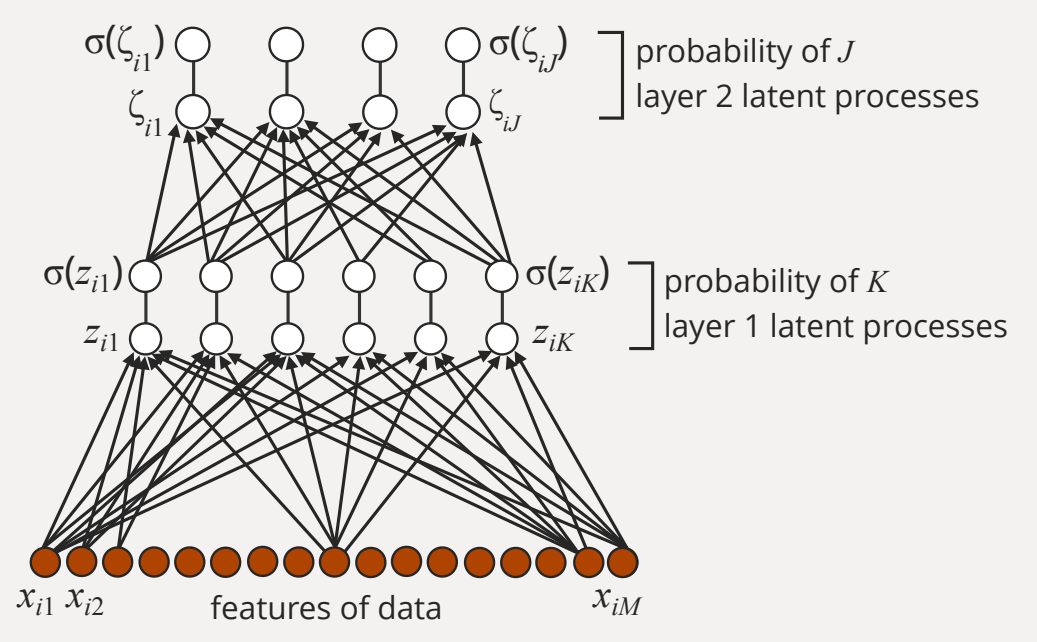


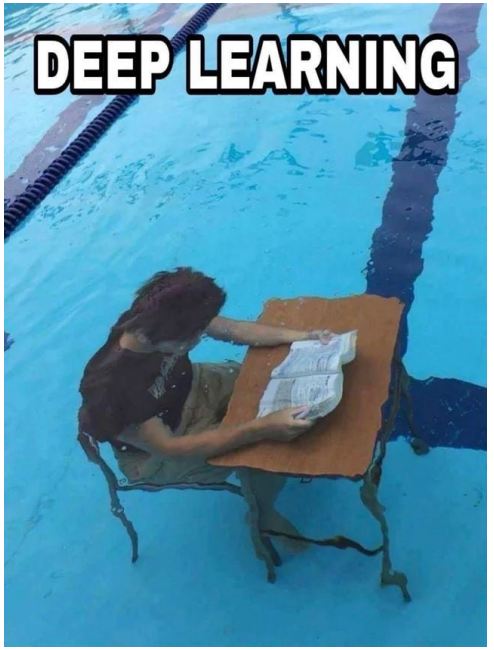

- Dans la pratique, les neurones sont initialisés avec des poids aléatoires
- L'entraînement permet à chacun de regarder les données avec des points de vue différents
- C'est la somme de ces points de vue qui donne une sortie

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE
ÉCOLE DE MANAGEMENT
DE LA SORBONNE

# Pourquoi s'arrêter là ?

- Différentes couches peuvent mettre en valeur des features distinctes
  - Plus de couches, plus de "poids" à optimiser
    - Il faut alléger les modèles
  - Les réseaux modernes sont parfois très profonds

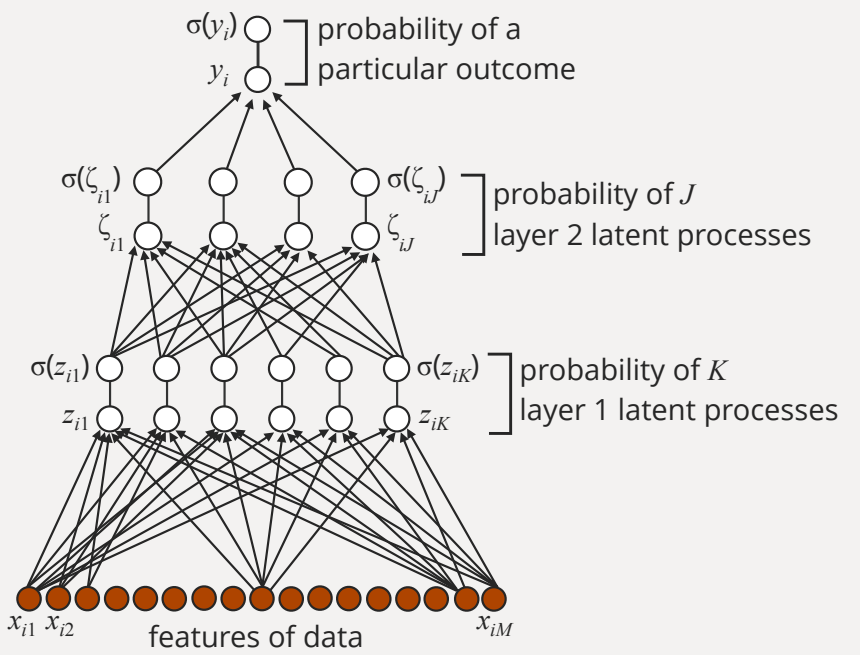


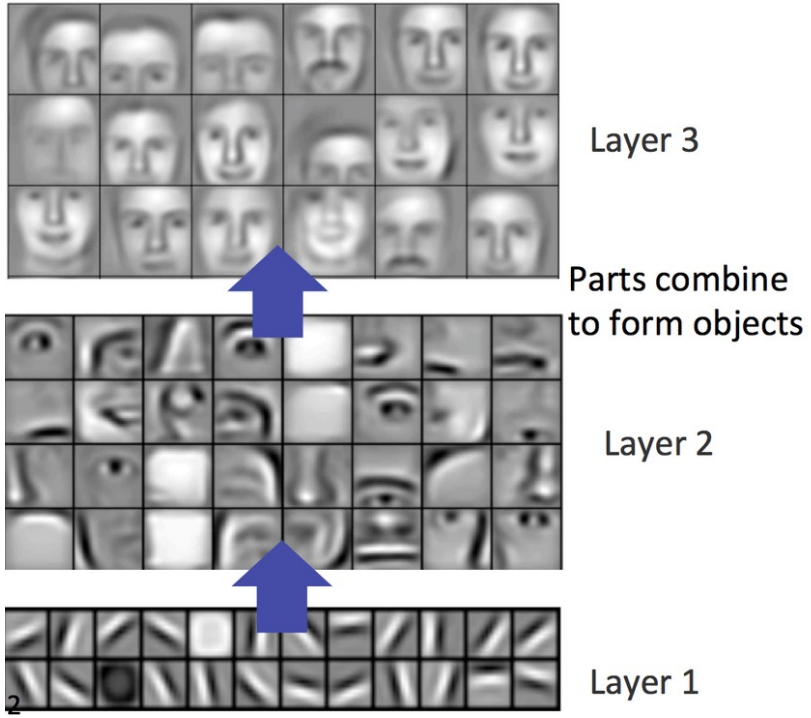



UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE
ÉCOLE DE MANAGEMENT
DE LA SORBONNE

# Réseaux Multi-couches avec Convolutions (CNN)

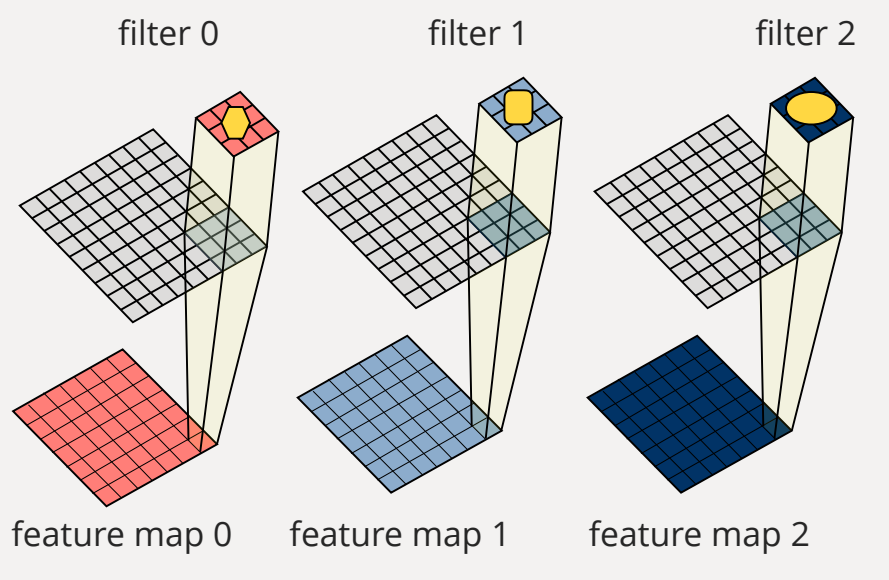
- Plutôt que des filtres complexes, on peut démarrer avec filtres simples (formes de base)
- Leur composition mène à des solutions complexes

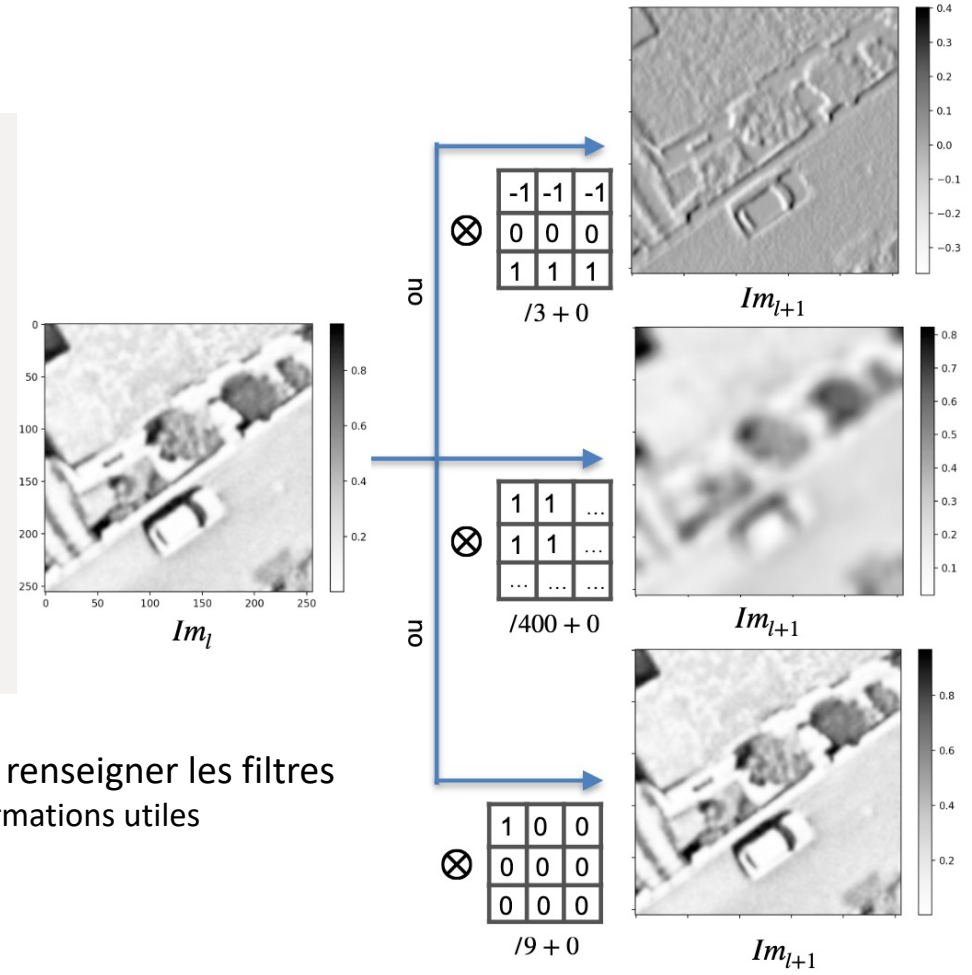


UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE
ÉCOLE DE MANAGEMENT
DE LA SORBONNE

# Multiples filtres de convolution




- Bonus : la plupart du temps, on n'a pas besoin de renseigner les filtres
  - Génération aléatoire mais qui souvent extrait d'informations utiles

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE
ÉCOLE DE MANAGEMENT
DE LA SORBONNE

# MLP dans Scikit-learn

- Sklearn inclut un module MLPClassifier qui permet de faire des réseaux de neurones

```
# Create a MLP with 2 hidden layers
clf = MLPClassifier(solver='lbfgs', alpha=1e-5,
                    hidden_layer_sizes=(5, 2), random_state=1)

# Train the MLP
clf.fit(X_train_std, y_train)
```

Le choix du nombre e de la forme des couches cachées a un très grand impact sur le résultat

```
: # Apply the trained perceptron on the X data to make predicts for the y test data
  y_pred = clf.predict(X_test_std)
  # View the accuracy of the model, which is: 1 - (observations predicted wrong / total observations)
  print('Accuracy: %.2f' % accuracy_score(y_test, y_pred))
```

```
Accuracy: 0.93
```

# Aller plus loin

- Les réseaux de neurones sont une solution avec beaucoup de potentiel
  - Pas toujours meilleure que les autres méthodes
  - Mas très efficace pour de images (à cause des convolutions)
- Scikit Learn n'est pas optimisé pour les réseaux de neurones
  - Ne supporte pas l'accélération sur GPU
- Certaines bibliothèques telles que **Keras** vous permettent de programmer en Tensorflow ou Pytorch avec une syntaxe proche de celle de Sklearn

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE
ÉCOLE DE MANAGEMENT
DE LA SORBONNE

# Hands-on

- Récupérer l'exercice "IntroNN_avec_Sklearn" sur le bureau virtuel