

# Apprentissage Automatique I 60629

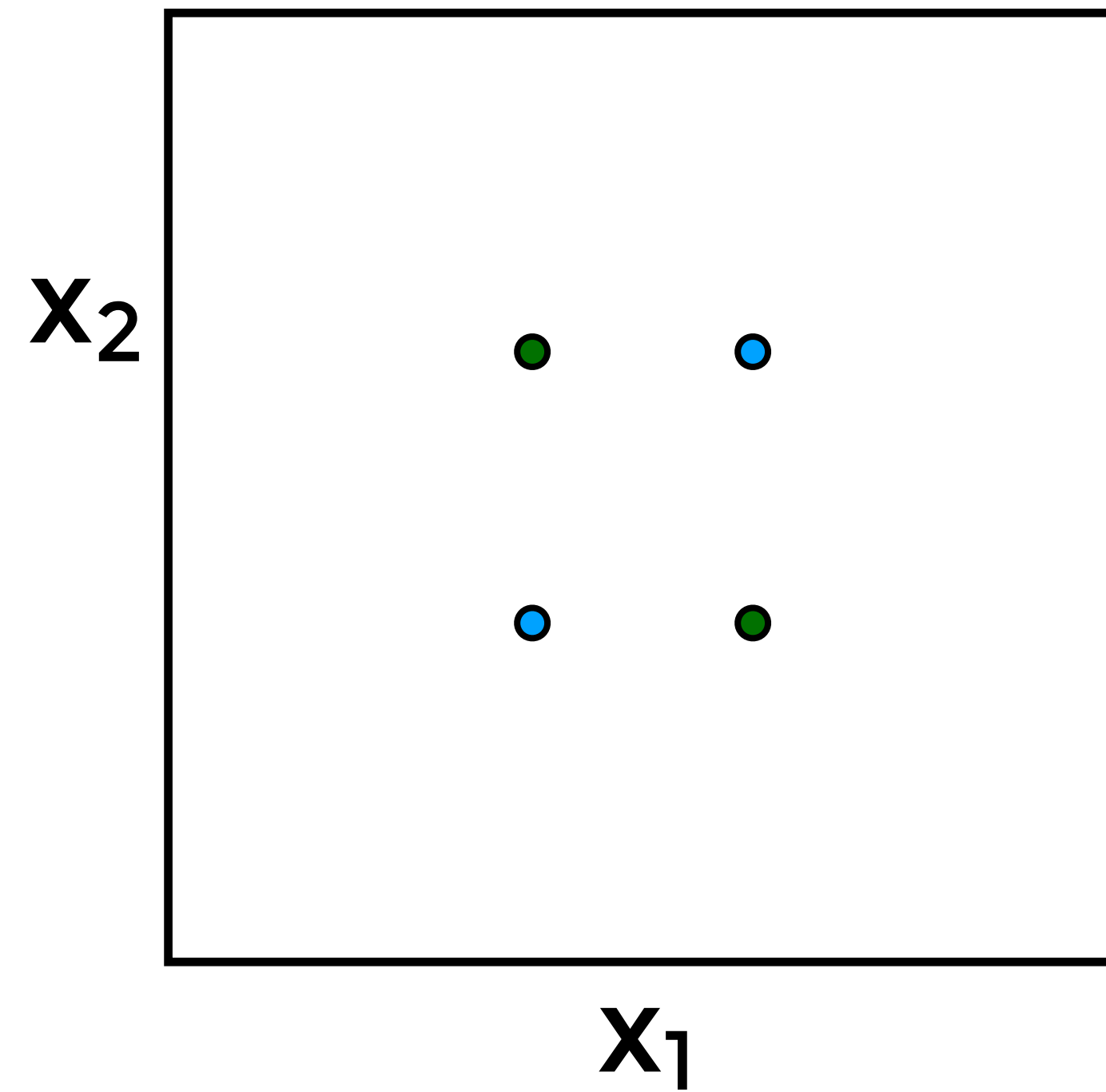
## **Sommaire**

Réseaux de neurones

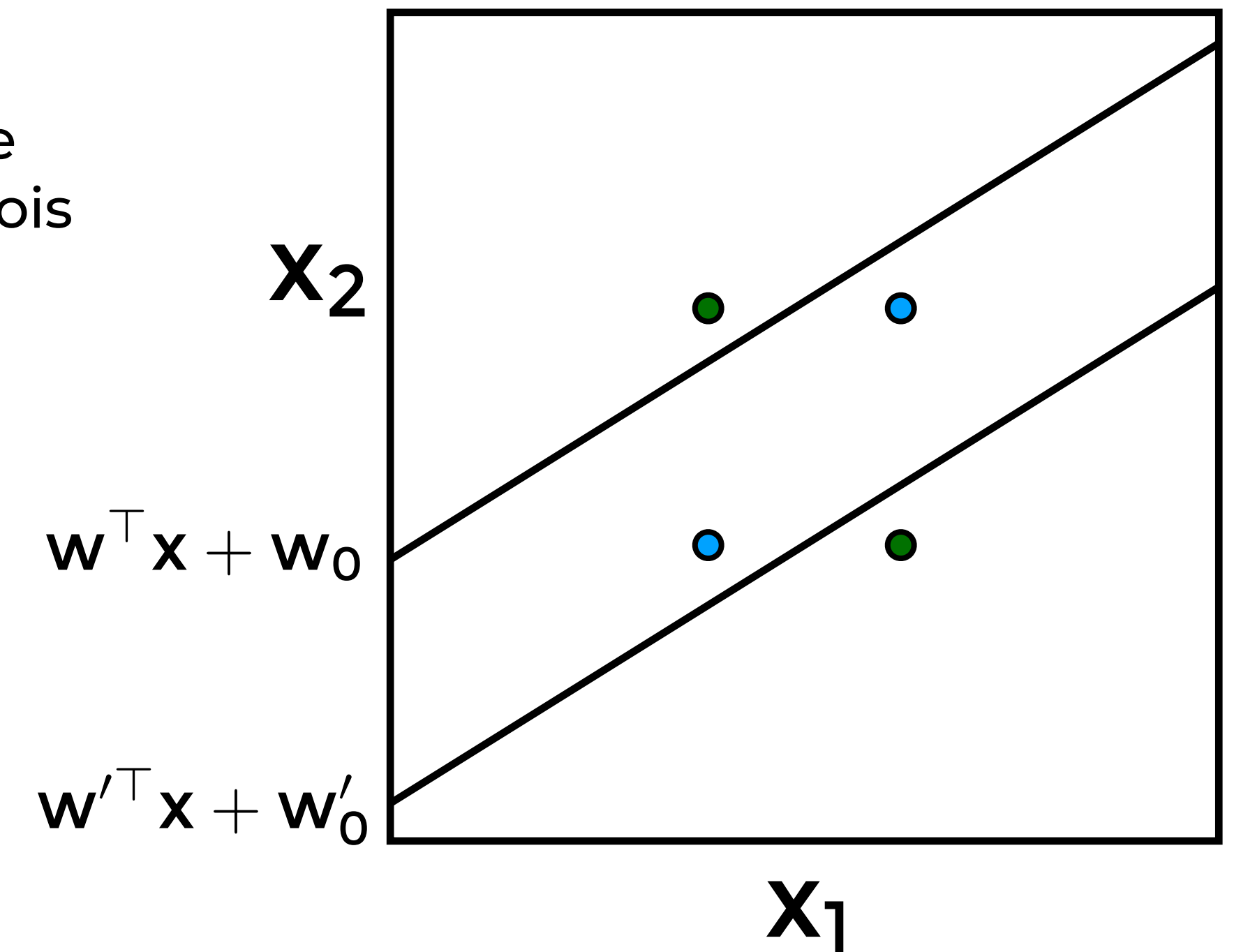
— Semaine #4

# Qu'en est-il si les données ne sont pas linéairement séparables?

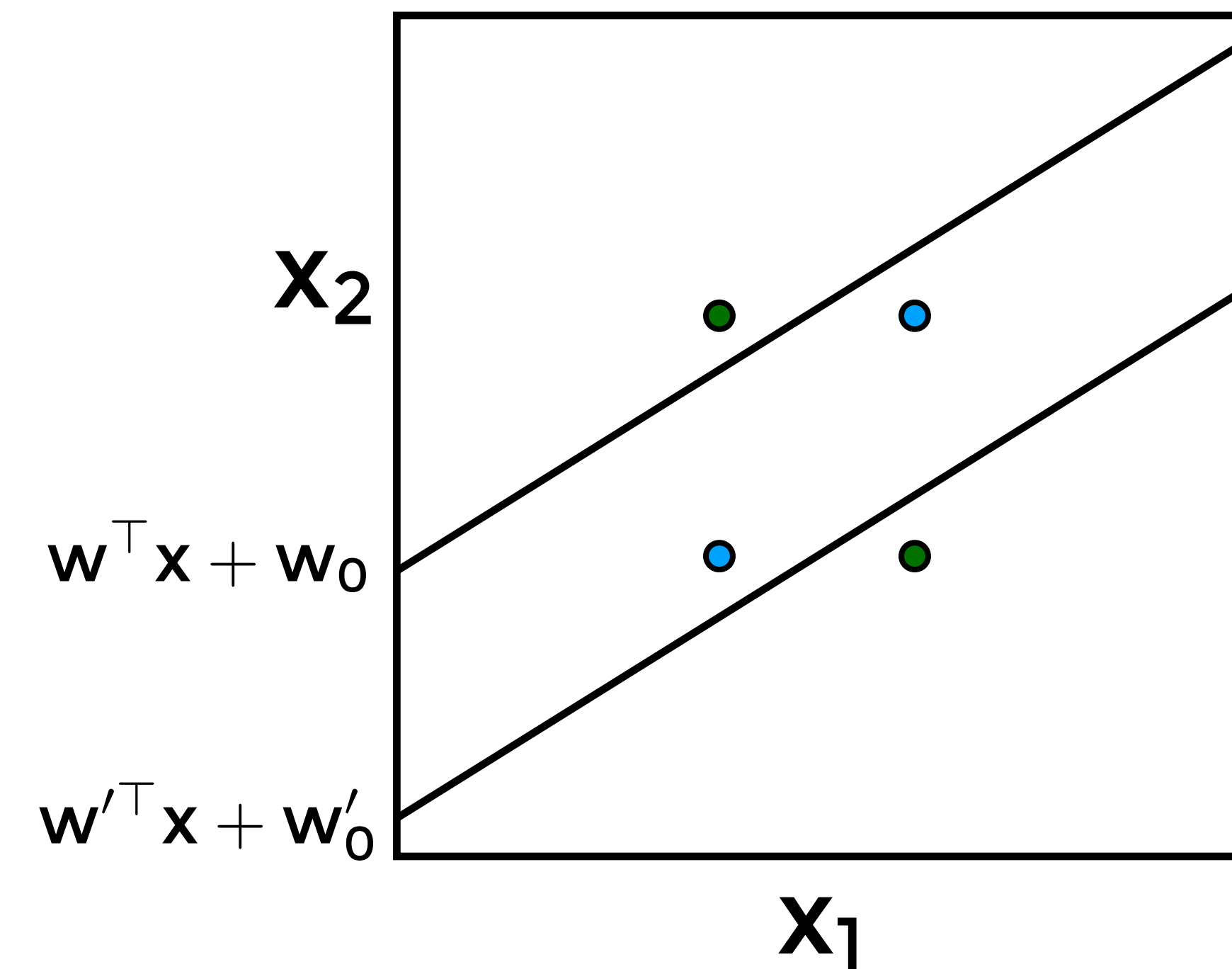
Exclusive OR (XOR)



On peut utiliser la décision de plusieurs classificateurs à la fois



# Combinaison de trois modèles



$$f(x): \begin{aligned} (w^T x + w_0) &> 0 \implies \bullet \\ (w^T x + w_0) &< 0 \implies \circ \end{aligned}$$

$$f'(x): \begin{aligned} (w'^T x + w'_0) &> 0 \implies \circ \\ (w'^T x + w'_0) &< 0 \implies \bullet \end{aligned}$$

$$f(x) = \bullet \text{ and } f'(x) = \circ \implies \bullet$$

$$f(x) = \circ \text{ and } f'(x) = \bullet \implies \bullet$$

$$f(x) = \circ \text{ and } f'(x) = \circ \implies \circ$$

1. On évalue chaque modèle

2. On combine les décisions des modèles

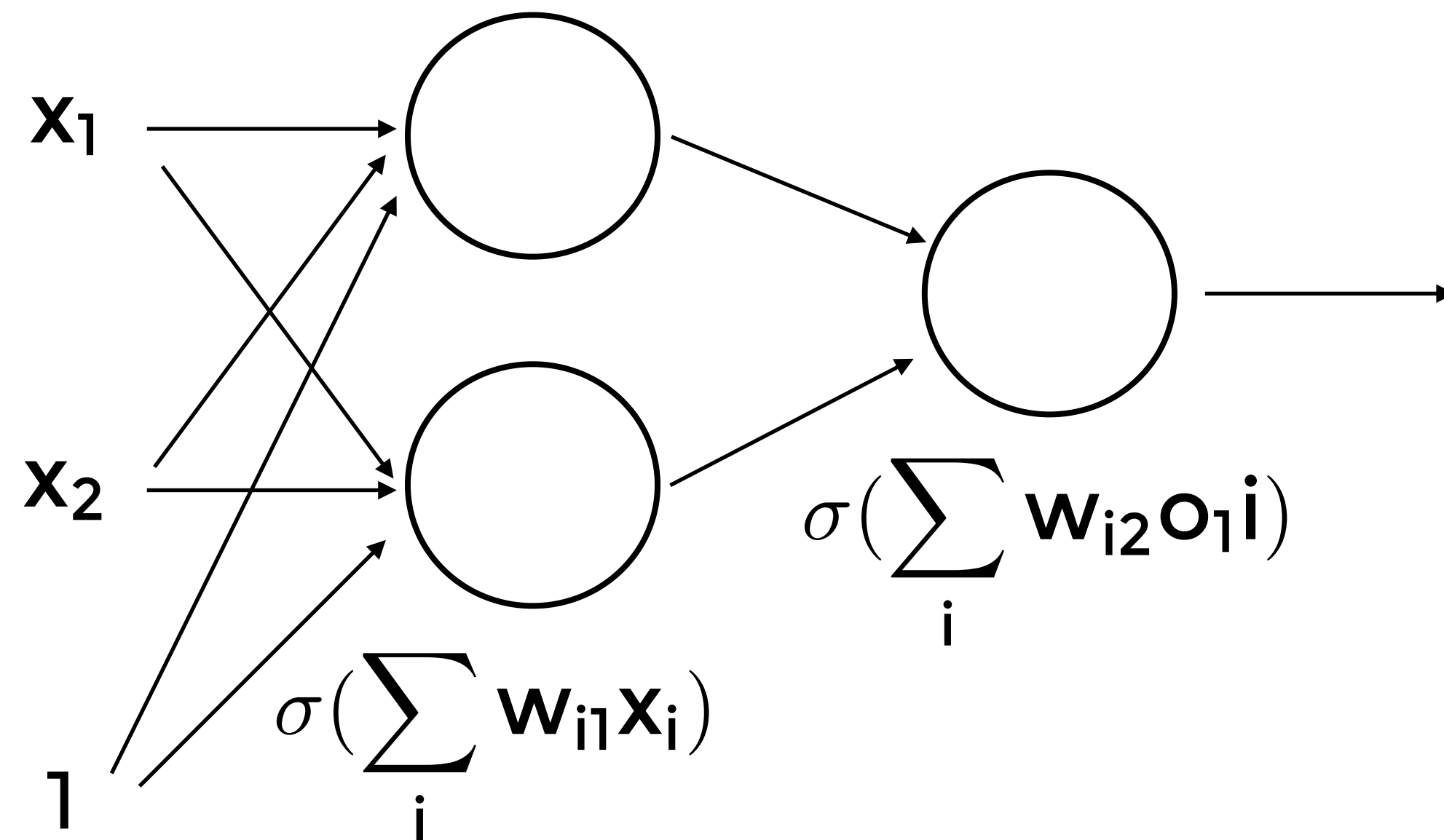
$$f''(x) = \text{threshold}\left(w''^T \begin{bmatrix} f(x) \\ f'(x) \end{bmatrix} + w''_0\right)$$

# Réseaux de neurones feed-forward

Couche d'entrées

Couche de sortie

Couche(s) cachée(s)



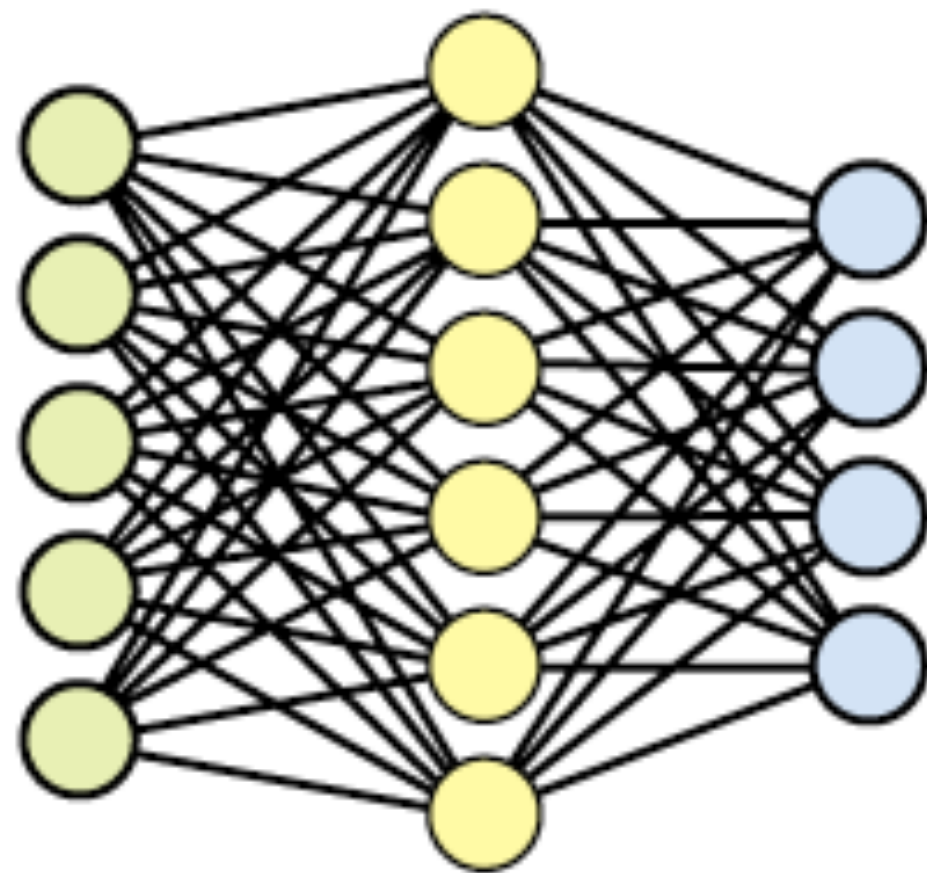
- Les flèches sont des connexions
- Un signal associé à un poids
- Chaque vertex calcule une somme pondérée de ses entrées suivie d'une activation non linéaire
- Les connexions vont de gauche à droite
- Il n'y a pas de connexion entre vertex d'une même couche
- Il n'y a pas de connexion vers « l'arrière » (récurrentes)

# Descente de gradient

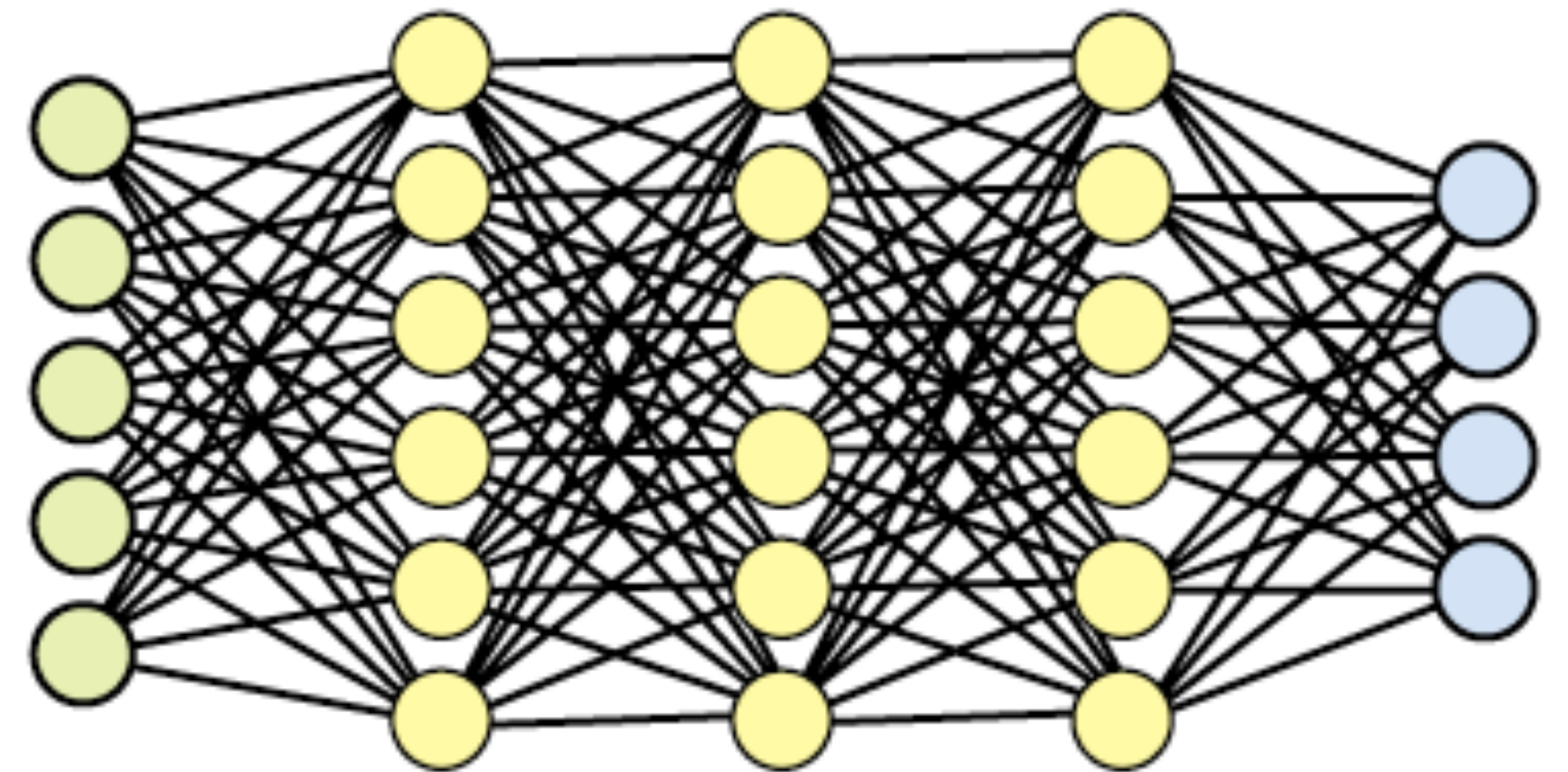
- Pas d'expression de forme fermée
- Répétez (pour  $t=0,1,2,\dots$  jusqu'à convergence):
  1. Calculer le gradient  $\nabla \mathbf{w}_{ij}^t$
  2. Mettre à jour les paramètres  $\mathbf{w}_{ij}^{t+1} = \mathbf{w}_{ij}^t - \alpha \nabla \mathbf{w}_{ij}^t$
- Descente de gradient stochastique
  - Un exemple à la fois
- Descente de gradient en lot (batch)
  - Tous les exemples à la fois

# De réseaux de neurones aux réseaux profonds

Un réseau de neurones



Un réseau de neurones profond



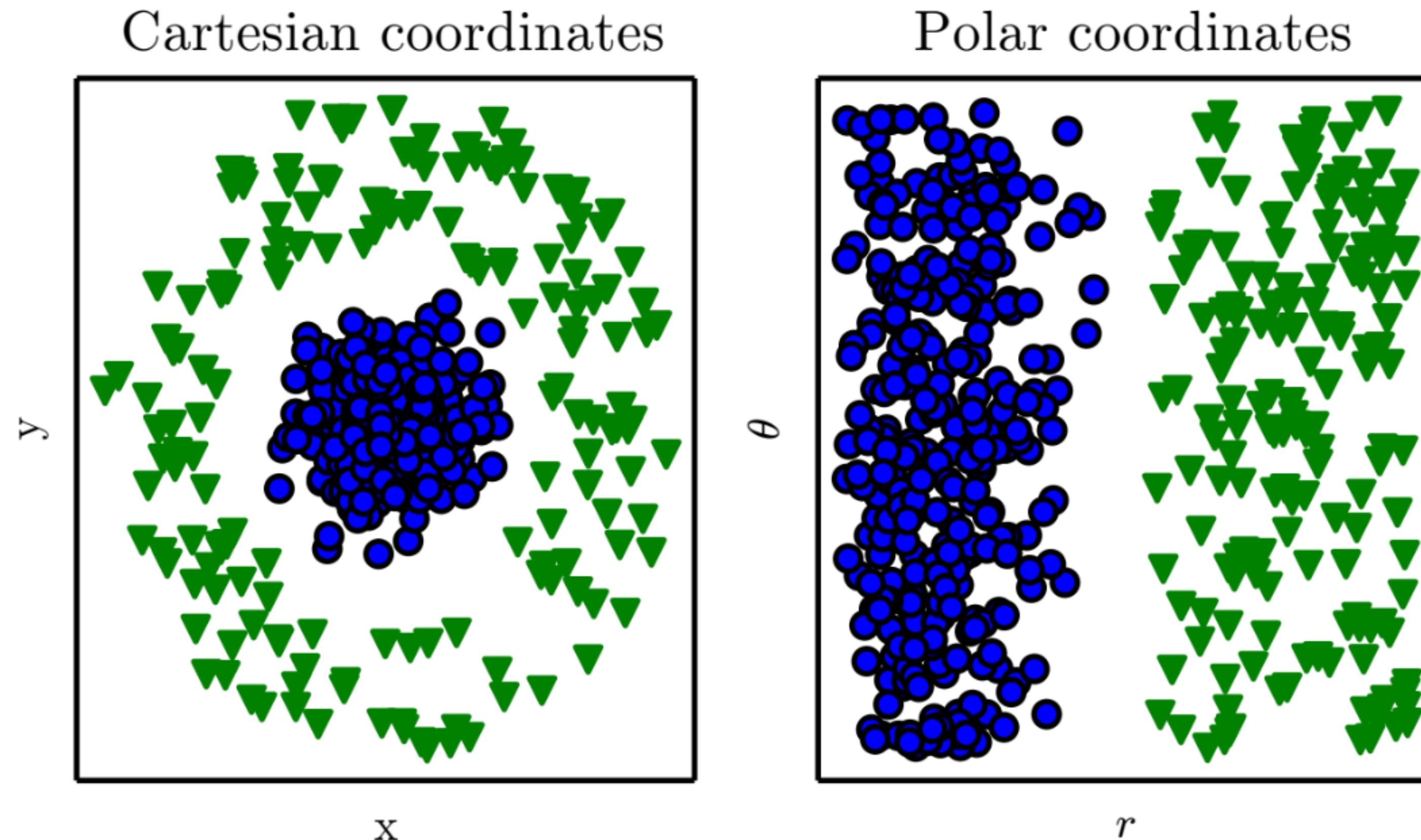
Modern deep learning provides a powerful framework for supervised learning. By adding more layers and more units within a layer, a deep network can represent functions of increasing complexity.

Deep Learning — Part II, p.163

[http://www.deeplearningbook.org/contents/part\\_practical.html](http://www.deeplearningbook.org/contents/part_practical.html)

# Une autre façon de comprendre les réseaux profonds

- L'importance des représentations



# Hyperparamètres

## 1. Modèle

- Fonctions d'activation (sortie & cachées), taille du réseau

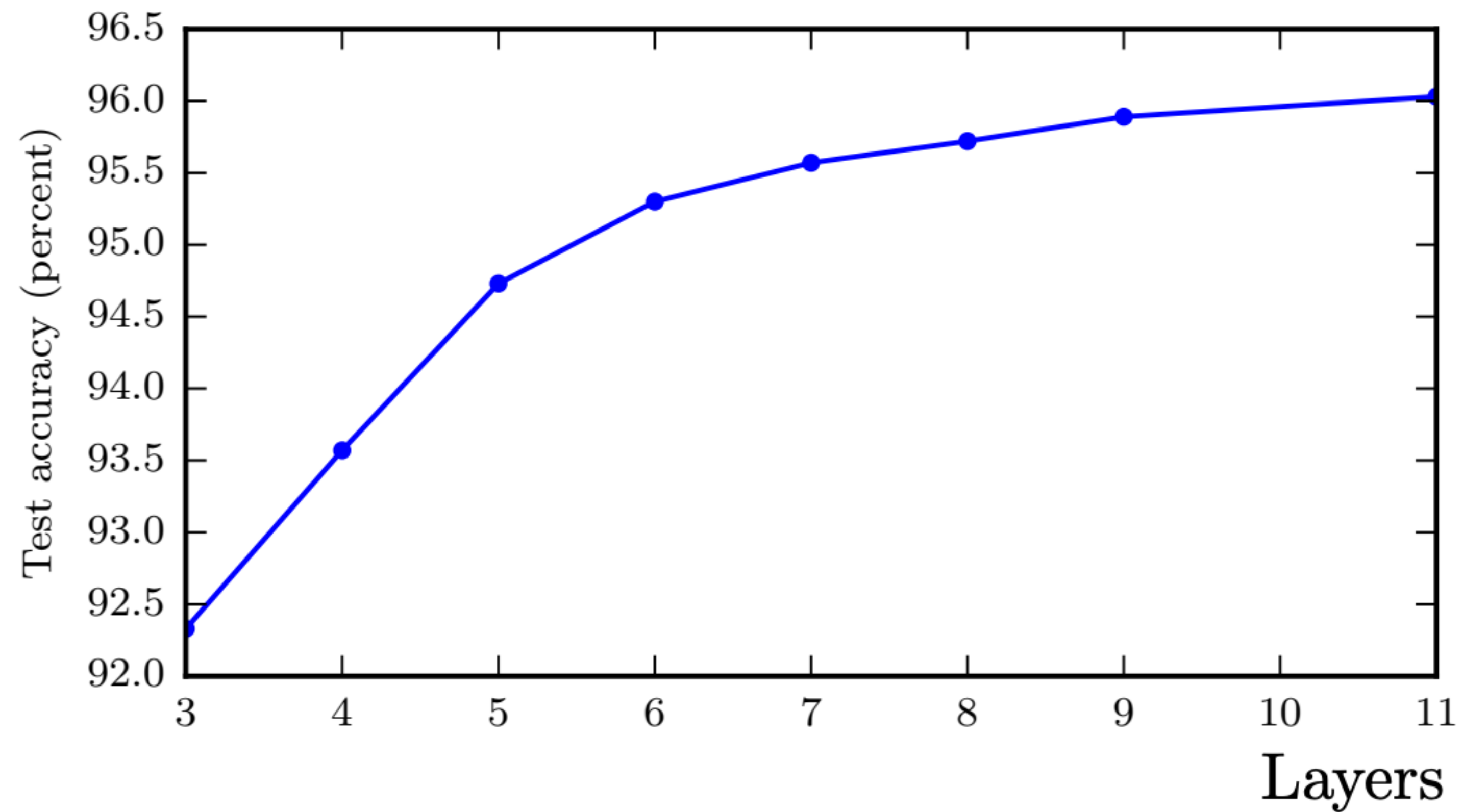
## 2. Objectif à optimiser

- Régularisation, Early-stopping, Dropout

## 3. Procédure d'optimisation

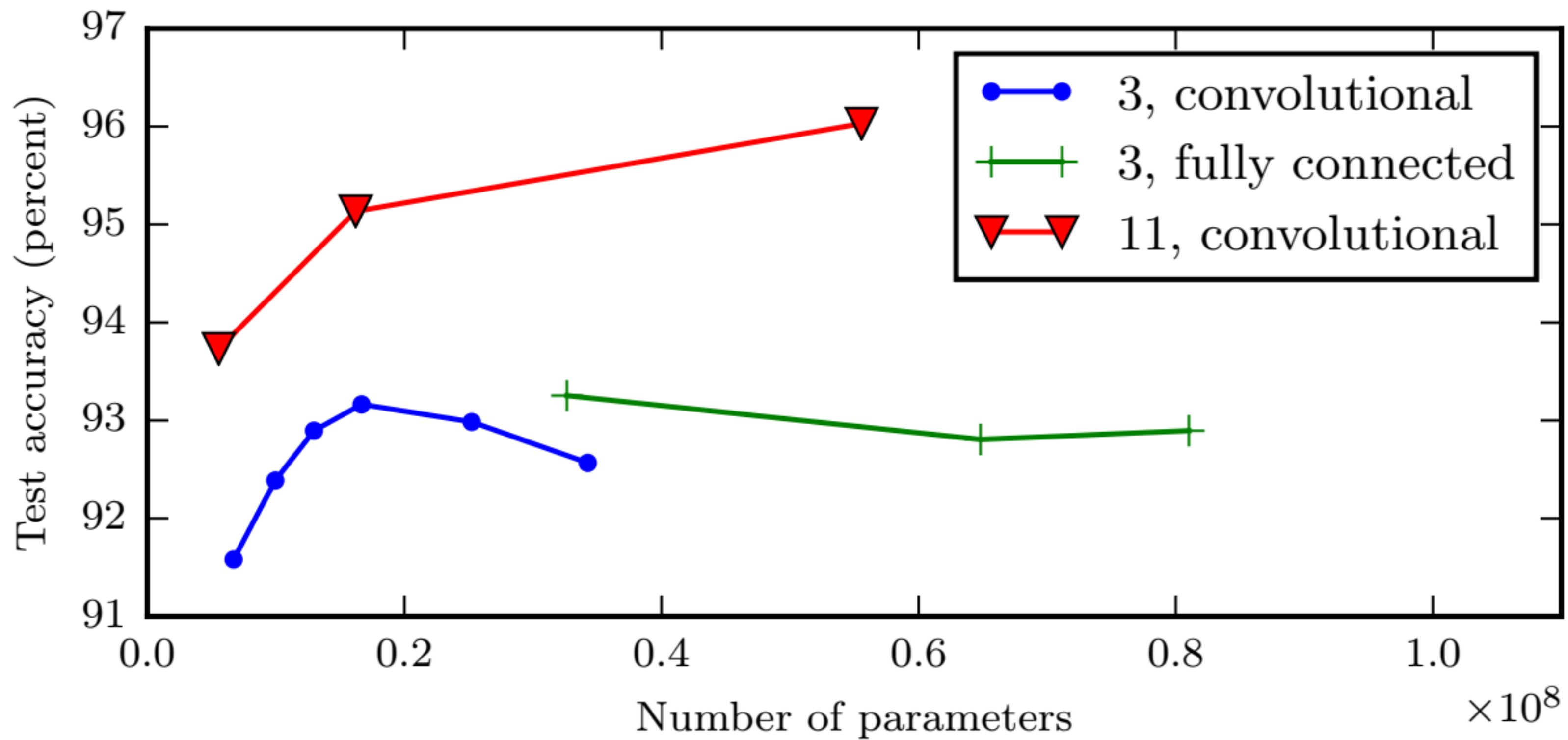
- Momentum, Adaptive learning rates

# Réseaux profonds?



[Figure 6.6, Deep Learning, book]

# Large ou Profond?



[Figure 6.7, *Deep Learning*, book]