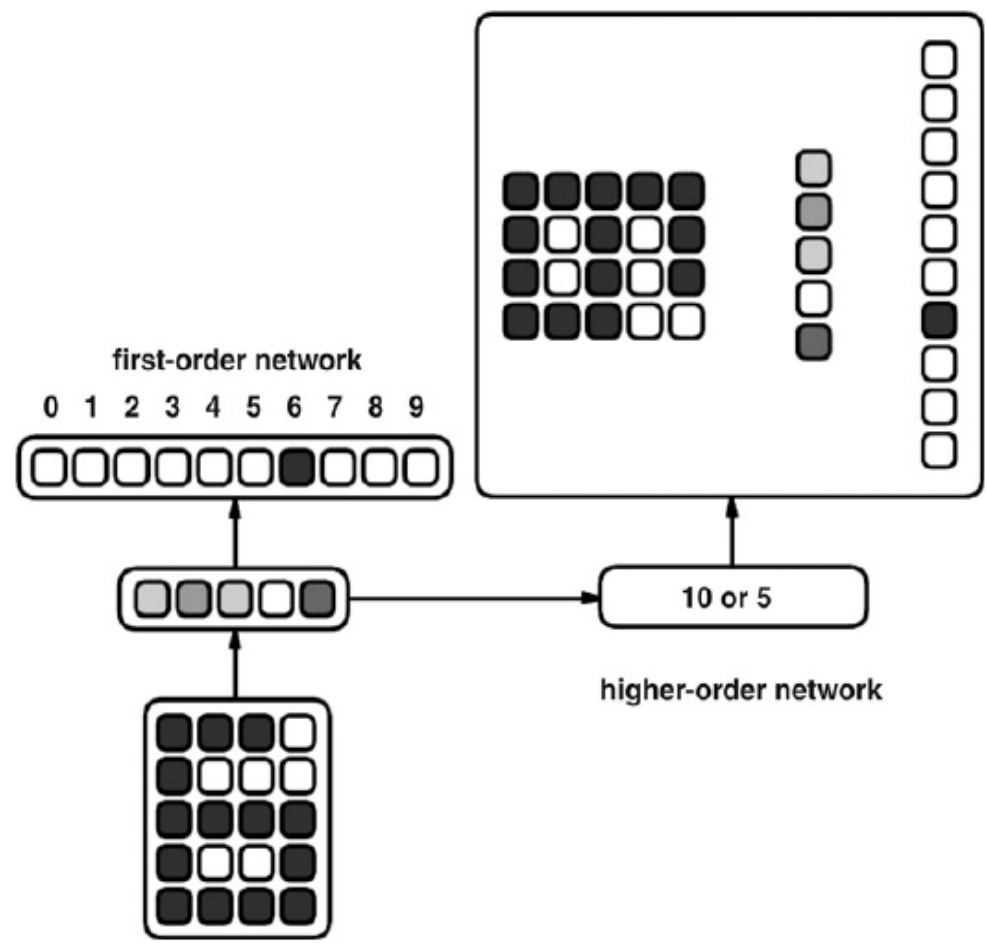


Simulation 1 : Reconnaissance de chiffre



Principe

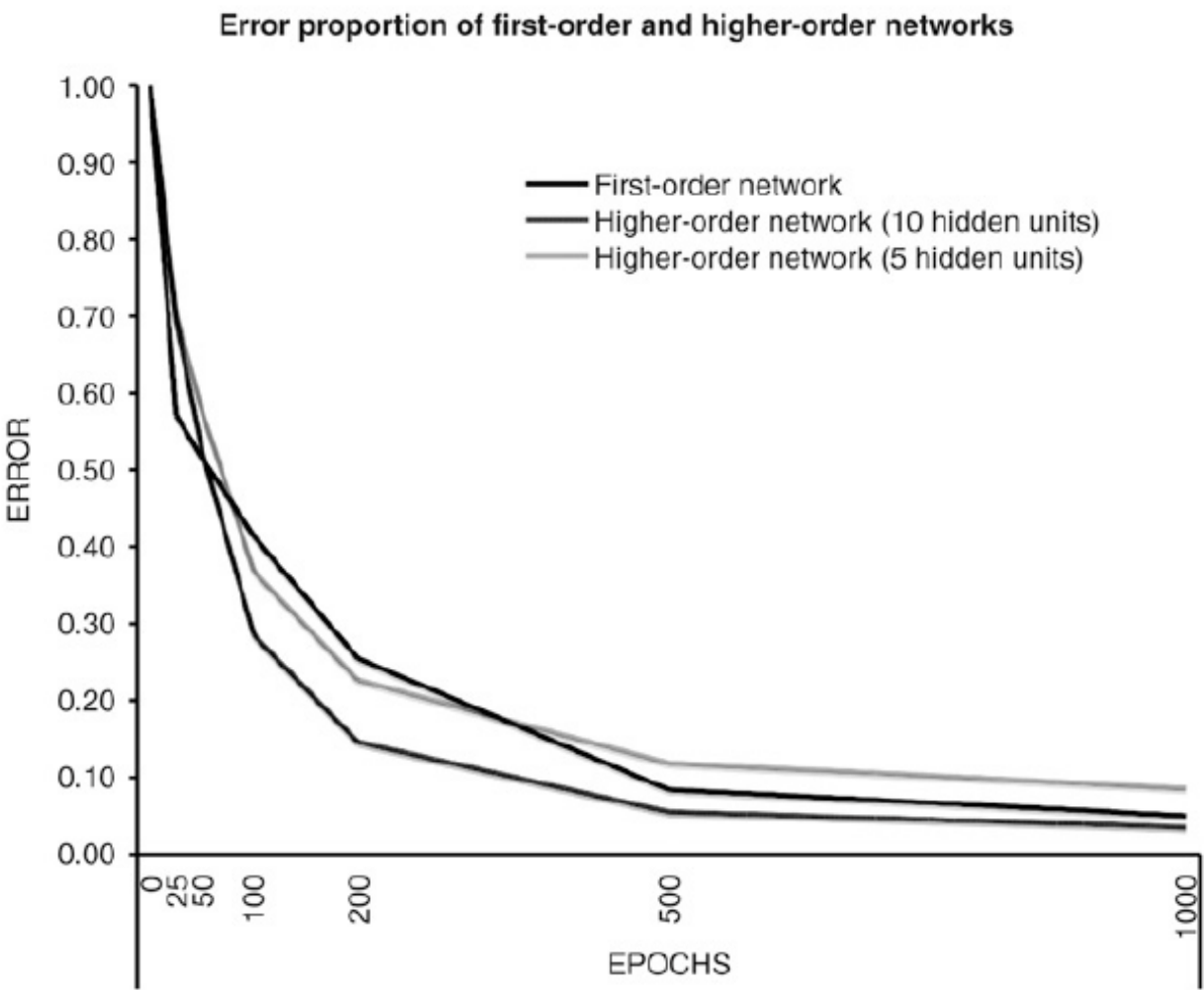
Un premier réseau apprend à associer une forme à un chiffre. En même temps, un second réseau apprend à retrouver l'ensemble du premier réseau, à partir, seulement, de la couche caché du premier.

Noter que lors de l'apprentissage du second réseau, il a aussi accès aux entrées/sorties du premier. ( mais pas à la sortie désirée du premier, si le premier reconnaît mal un chiffre, cela sera répercuté sur le second )

Paramètres

Momentum : 0.9	Taux d'apprentissage : 0.1 ( constant )	5 unités caché pour le premier
		10/5 unités caché pour le second
Apprentissage : 1000 epochs x 10 formes (shuffle)		Moyenne sur 5 réseaux

Résultat de l'article



Conclusion

Le réseau de second ordre apprend plus rapidement à retrouver l'état entier du premier, que celui-ci n'apprend à retrouver le bon chiffre.

Attention : le réseau de second ordre n'apprend pas plus rapidement à retrouver le bon chiffre.

Note : l'erreur RMS est définie comme :

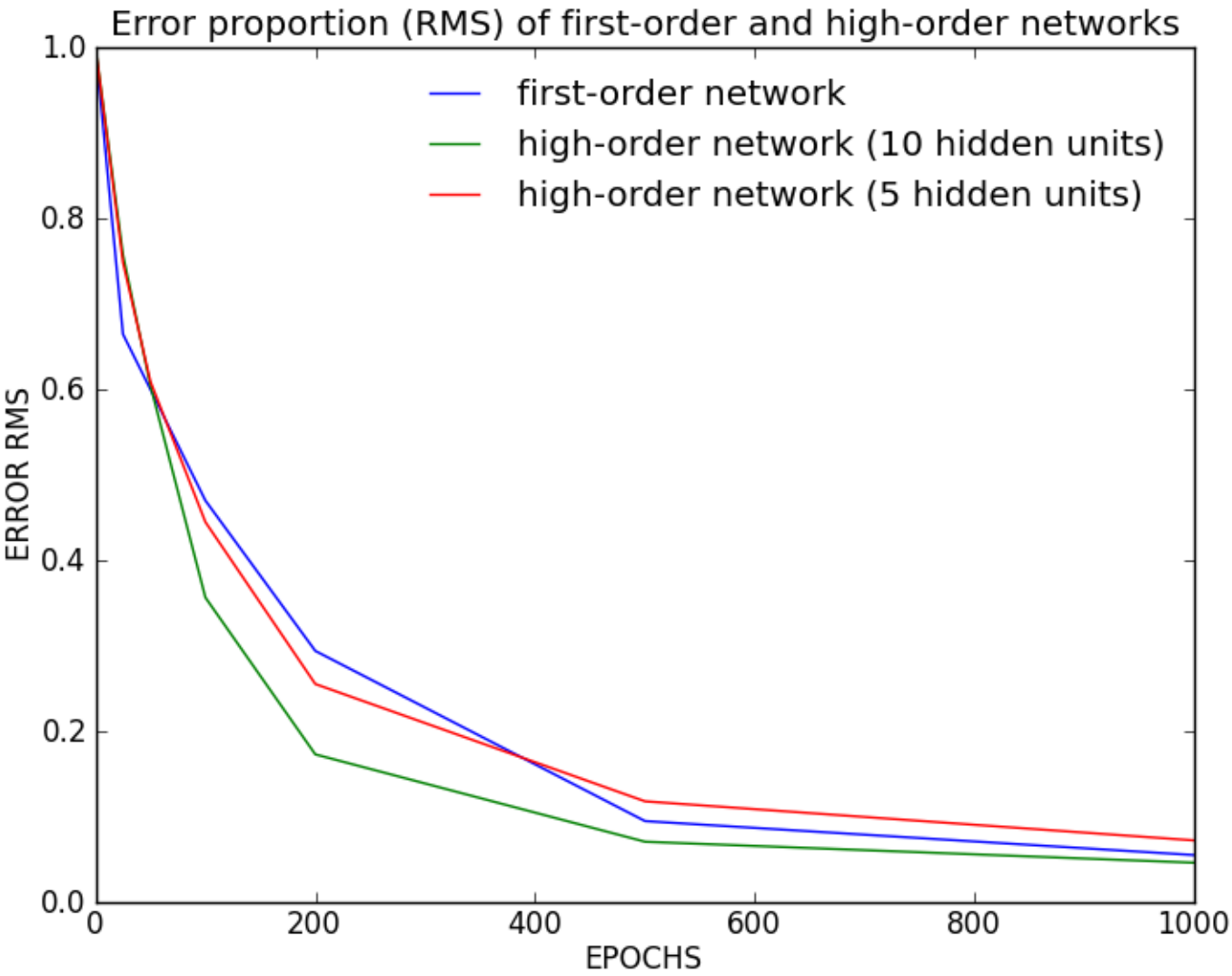
$$\sqrt{\frac{1}{n} \sum_{i=1}^n (o_i - d_i)^2}$$

with  $\begin{cases} n : \text{number of neurons} \\ o : \text{values obtained} \\ d : \text{values desired} \end{cases}$

divisé par l'erreur maximal pour revenir sur [0 ; 1]

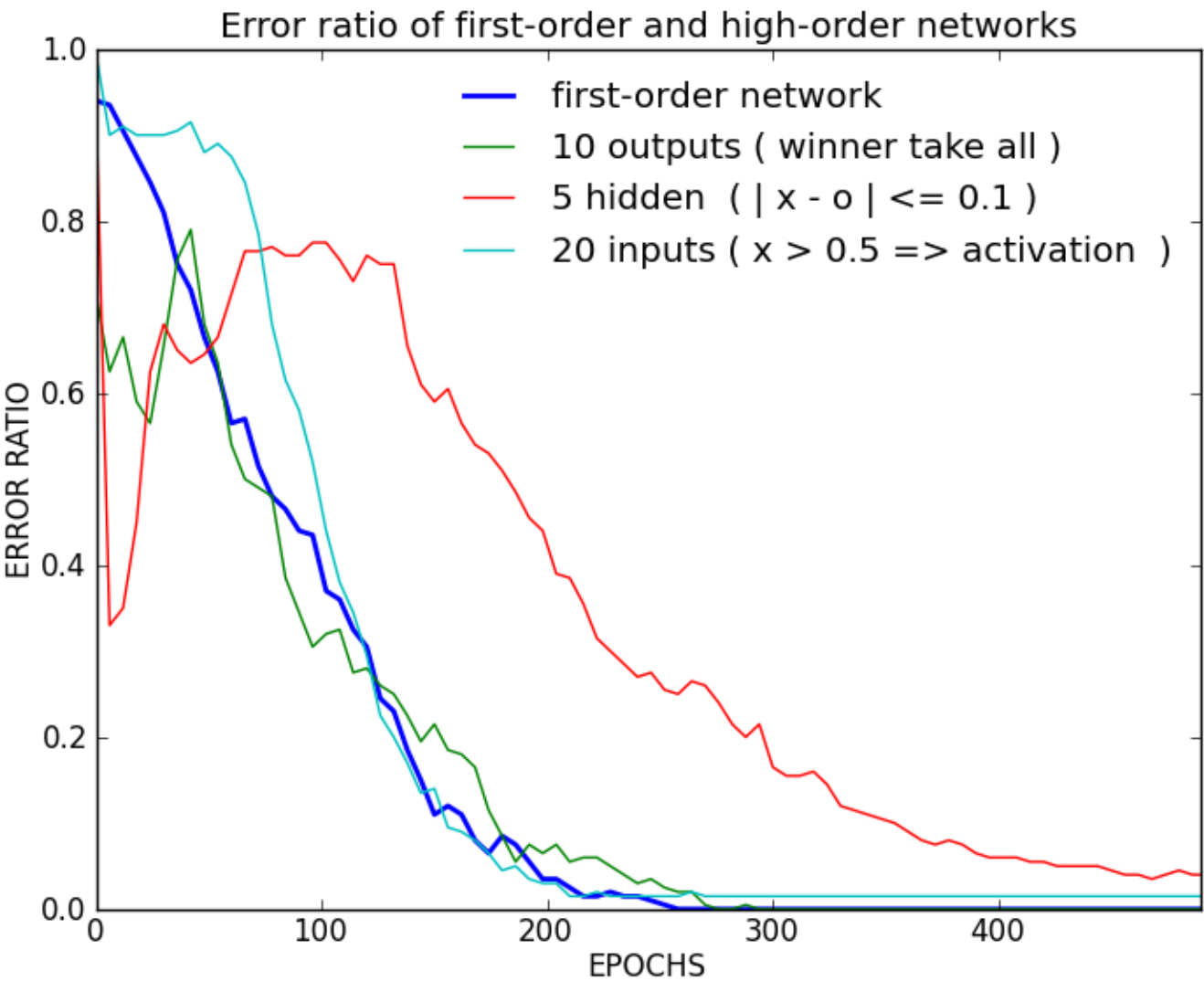
Nos résultats :

Paramètres devinés/supposés			
Température : 1 pour tous	Taux d'apprentissage constant	Rétro propagation online	Poids initialisés entre -0.25 et 0.25
Sorties des neurones sur [0 ; 1]	Sigmoïde sur [0,1] :	$\frac{1}{1+e^{-\theta x}}$	



On retrouve les mêmes courbes.

Taux d'erreur : ( défini comme le nombre moyen d'échecs sur une epoch)



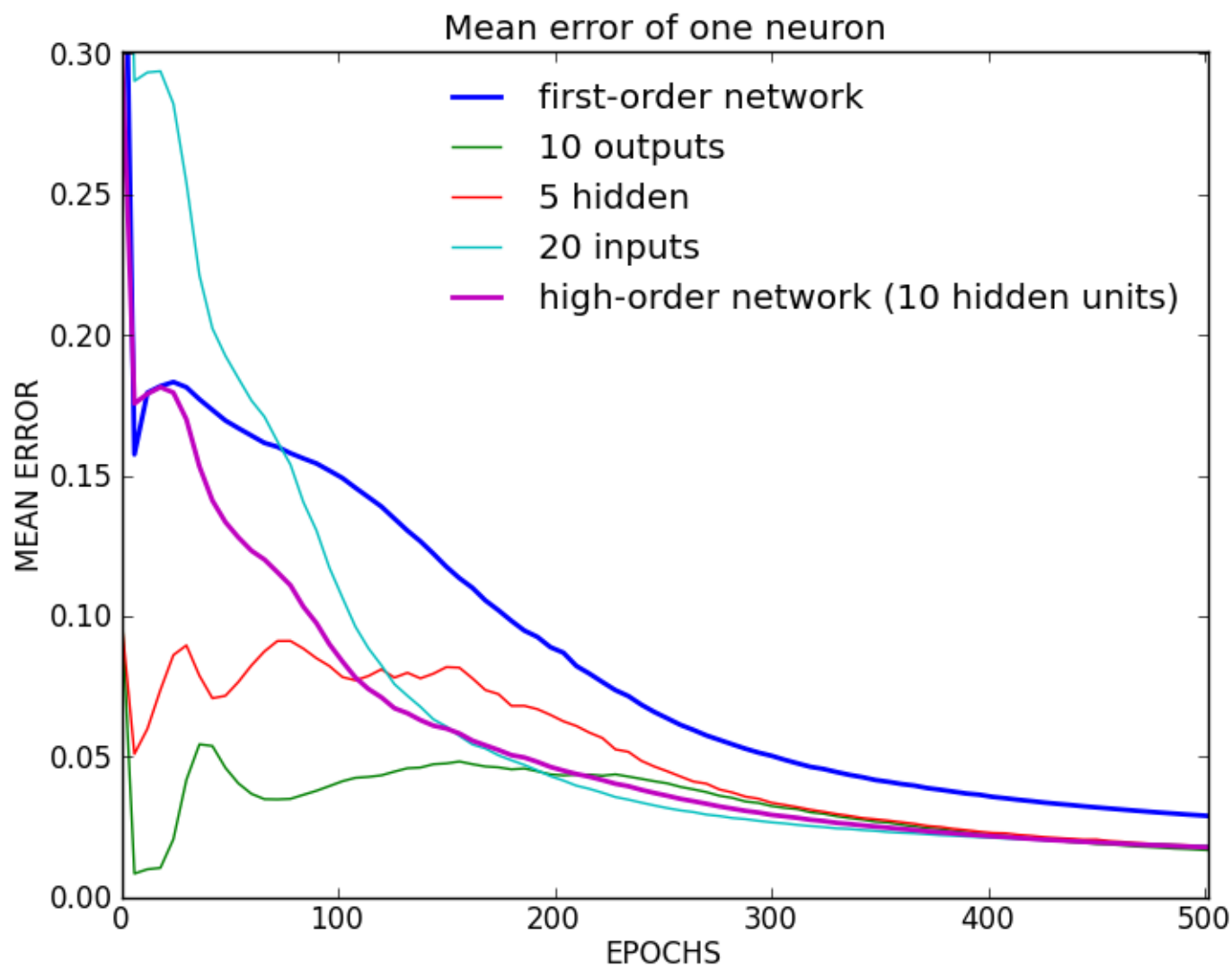
Les 3 courbes représentent en détail le réseau d'ordre supérieur ( avec 10 unités cachées )

Attention à la courbe des 5 unités cachées, puisque le taux d'erreur accordé par neurone joue beaucoup.

On peut voir que le réseau de second-ordre fait moins d'erreur à retrouver la sortie (du premier) que le premier à retrouver la solution, sauf après 150 epochs.

zoomer sur [ 0 ; 500 ]

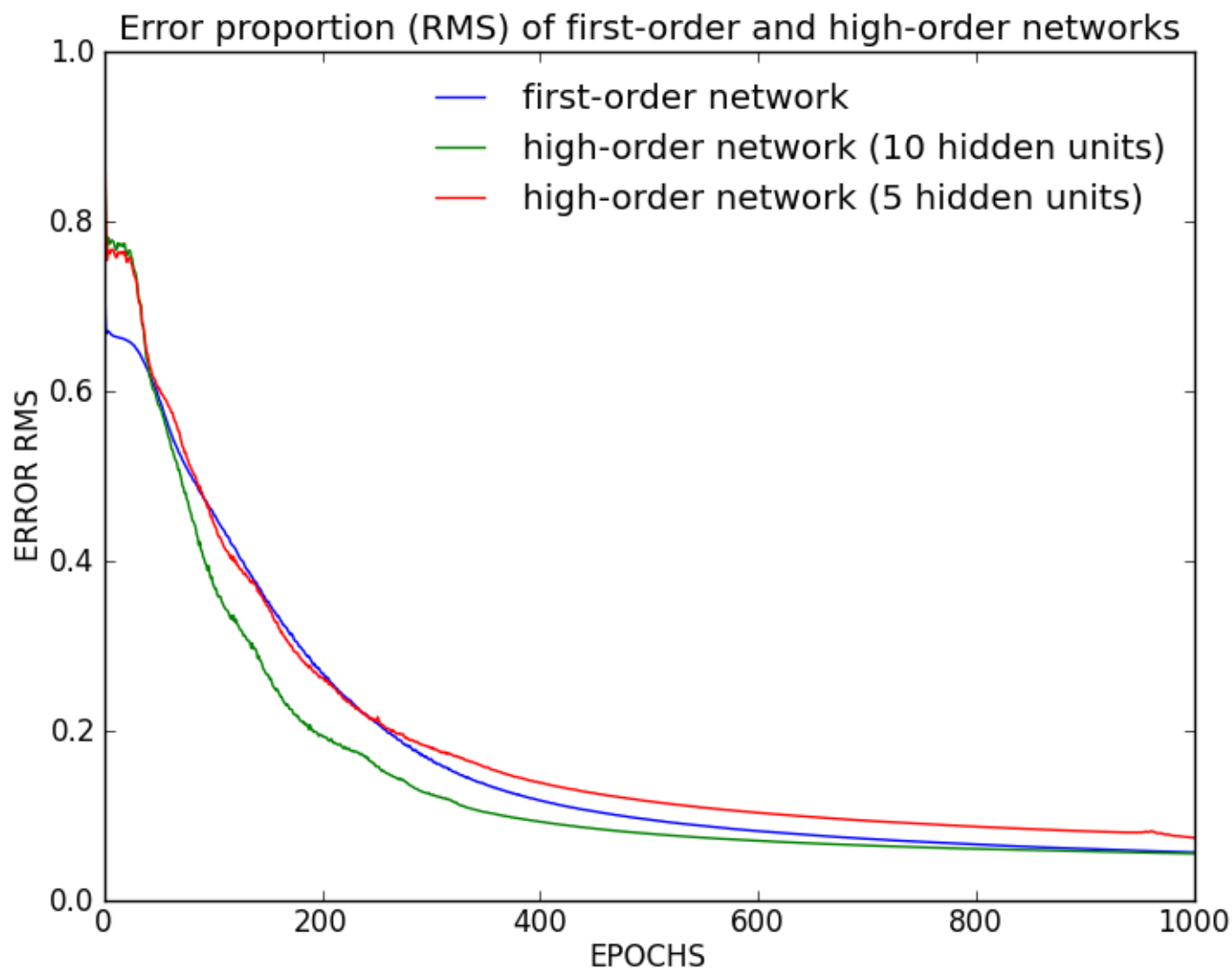
Erreur moyenne pour un neurone



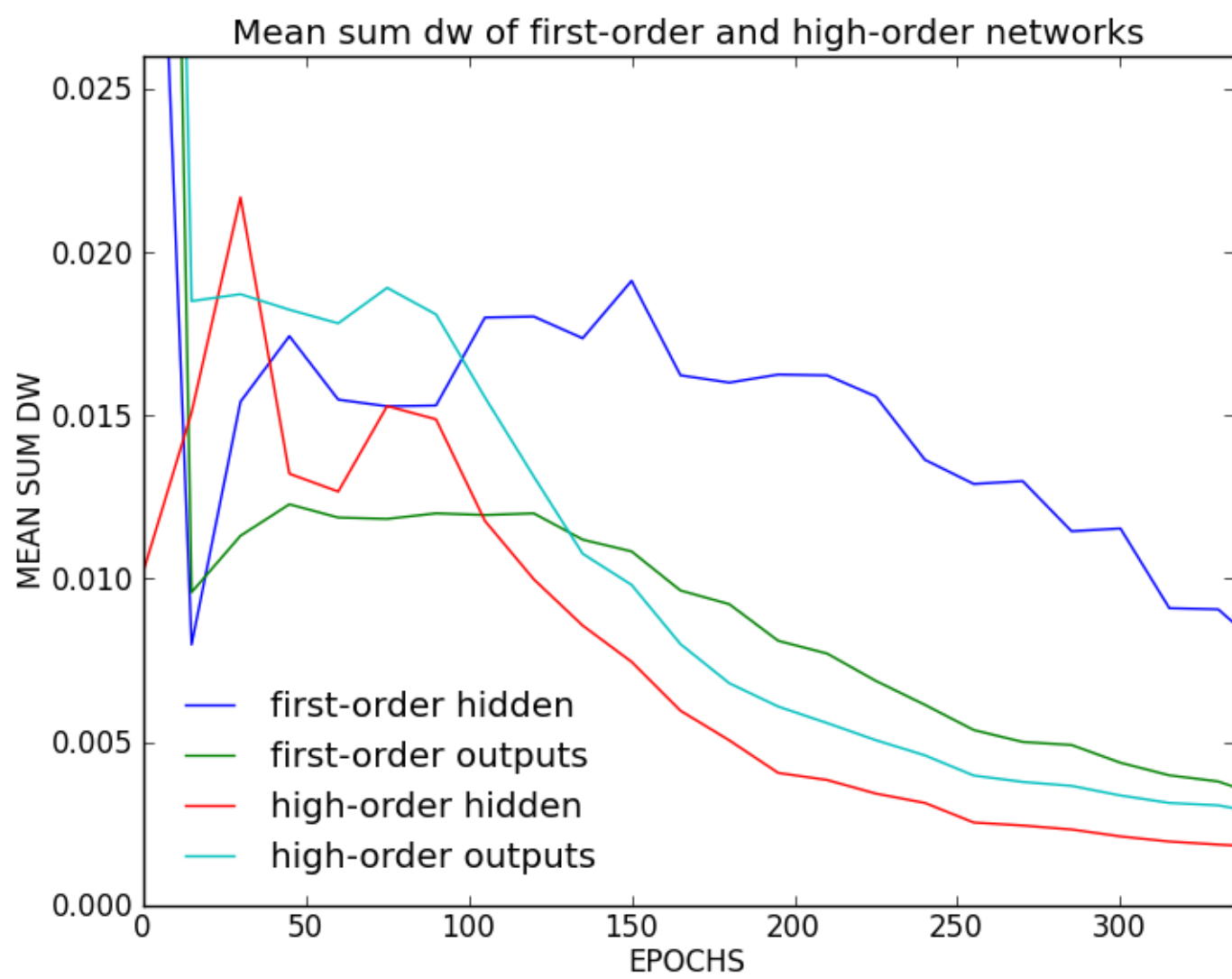
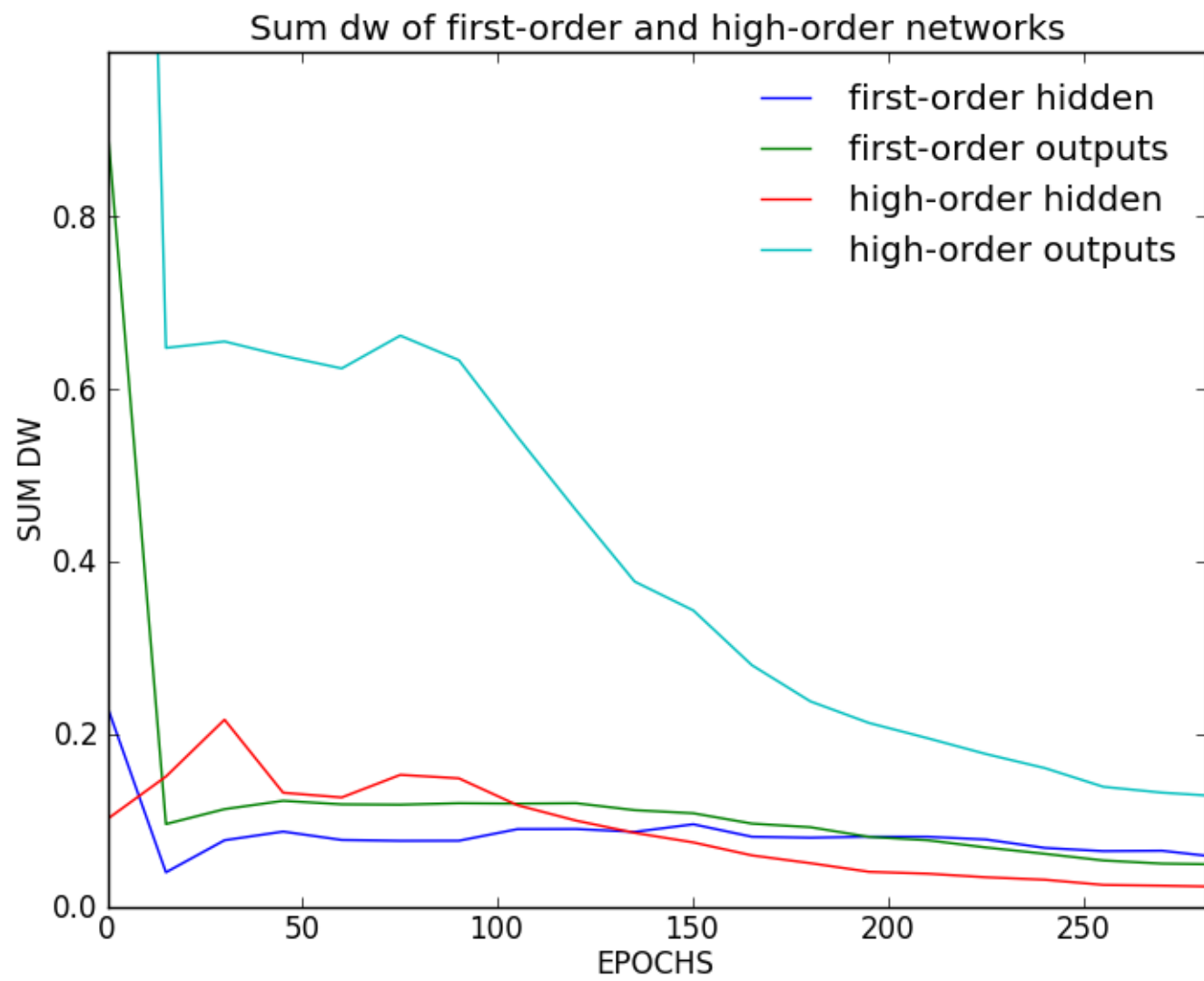
Le réseau supérieur à plus de mal à apprendre la couche caché et la couche de sortie ( même s'il y a peu d'erreur dessus ).

Les performances de la courbe RMS viennent de la nette diminution de l'erreur sur l'apprentissage des 20 neurones d'entrés.

Courbes continue



Les oscillations sont faibles.  
( moyenne effectuée sur 5 réseaux )

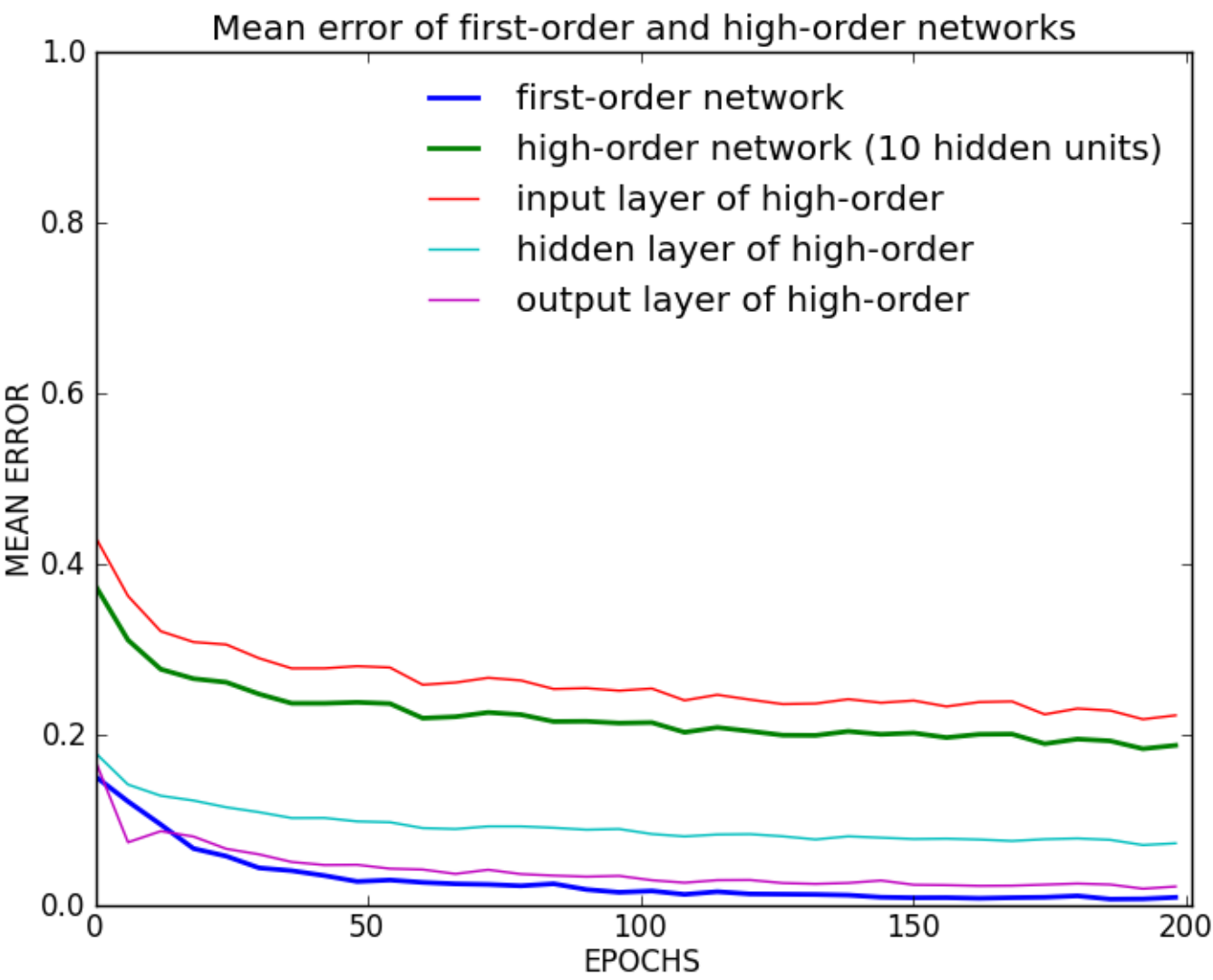


Essaies sur des chiffres manuscrits

Entrées : 16 x 16 pixels  
Apprentissage : 200 epochs x 50 formes (shuffle)  
avec 1600 formes

16\*4 unités caché pour le premier  
16\*4\*2 unités caché pour le second

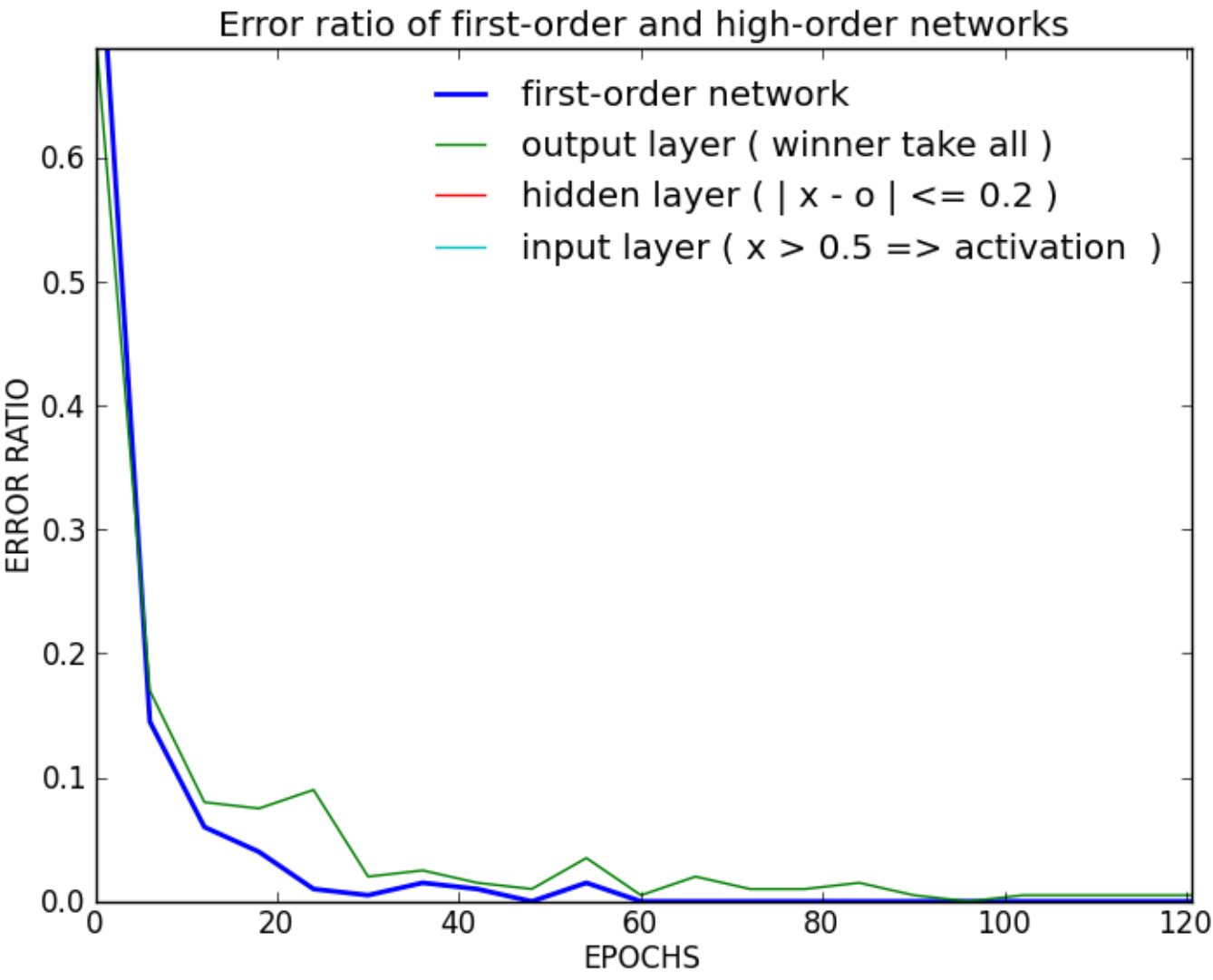
10/5 unités caché pour le second



Tout comme sur les formes simplifiées, l'erreur du second réseau tend vers l'erreur d'entrée qui a beaucoup plus de poids, sauf que cette fois-ci l'entrée est compliqué à apprendre.

C'est pour cela que le réseau de second ordre n'arrive pas à dépasser le premier.

Notons tout de même qu'il arrive à apprendre correctement la sortie du premier réseau.



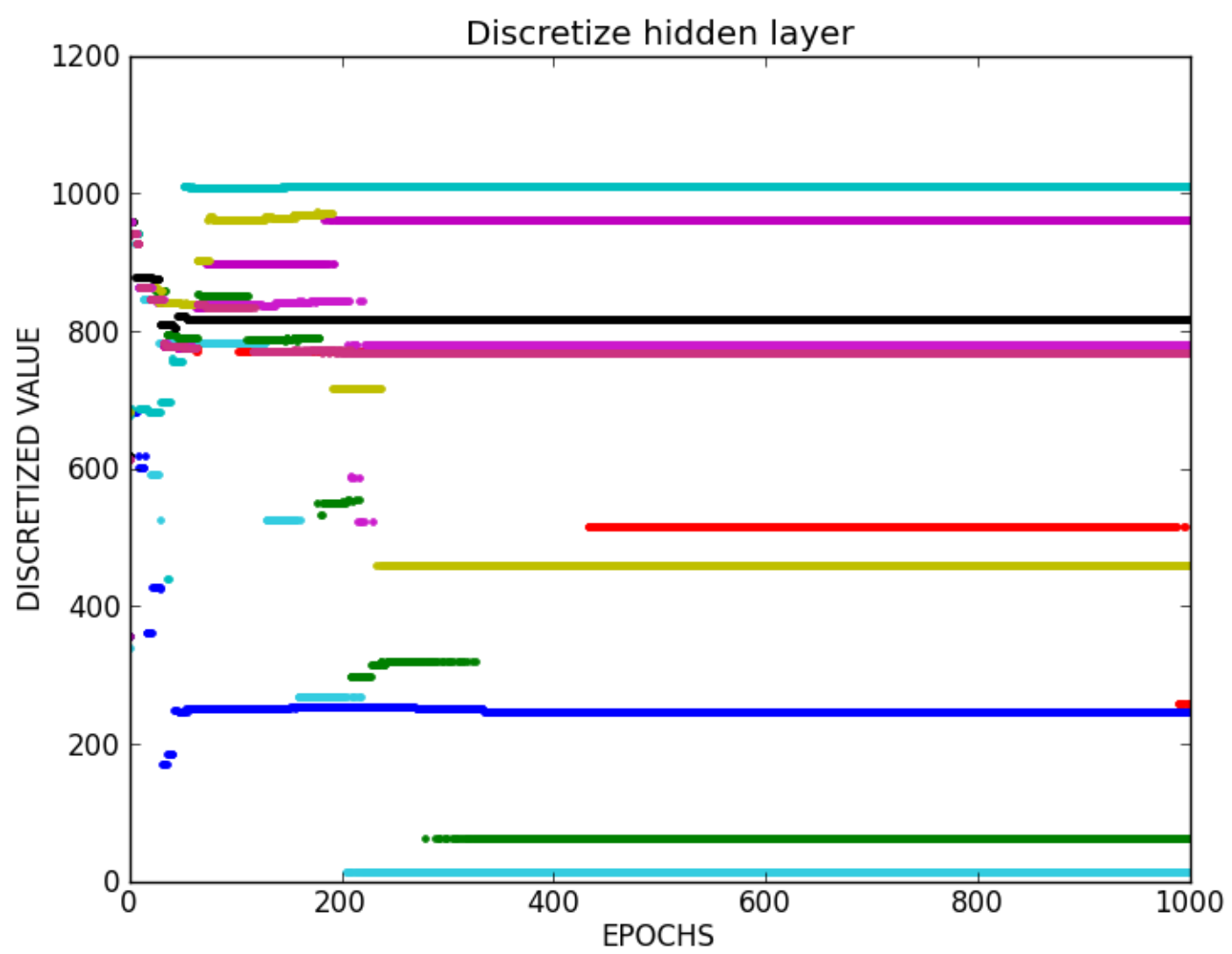
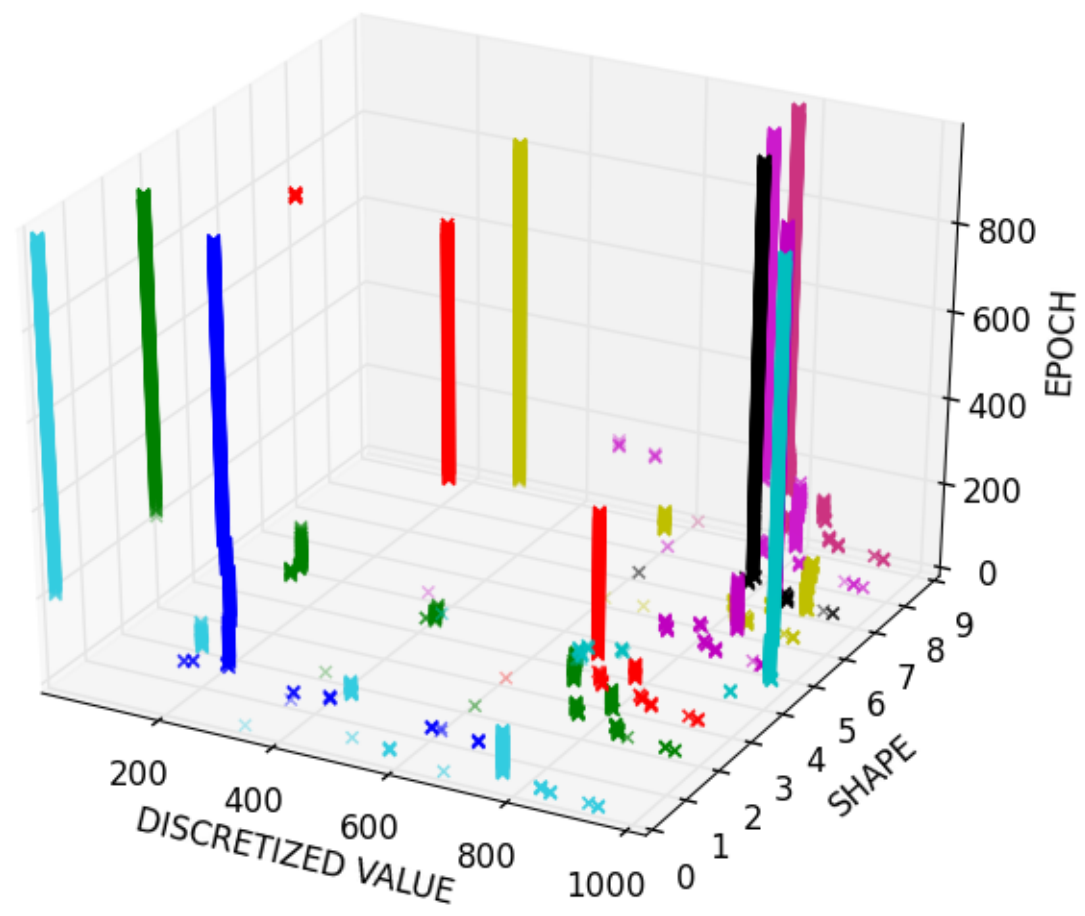
Ces courbes confirment les précédentes.

L'entrée et la couche caché n'est jamais apprise correctement. ( les courbes ne sont pas visibles et restent approximativement très proche de 1. )

Néanmoins le réseau supérieur arrive à apprendre la sortie que le premier va donner ( avec un taux d'erreur plus important que celui-ci ne retrouve la solution )

Si on transforme le second réseau pour qu'il n'apprenne que la couche de sortie, il devient meilleur que le premier au bout de 10 epoch.

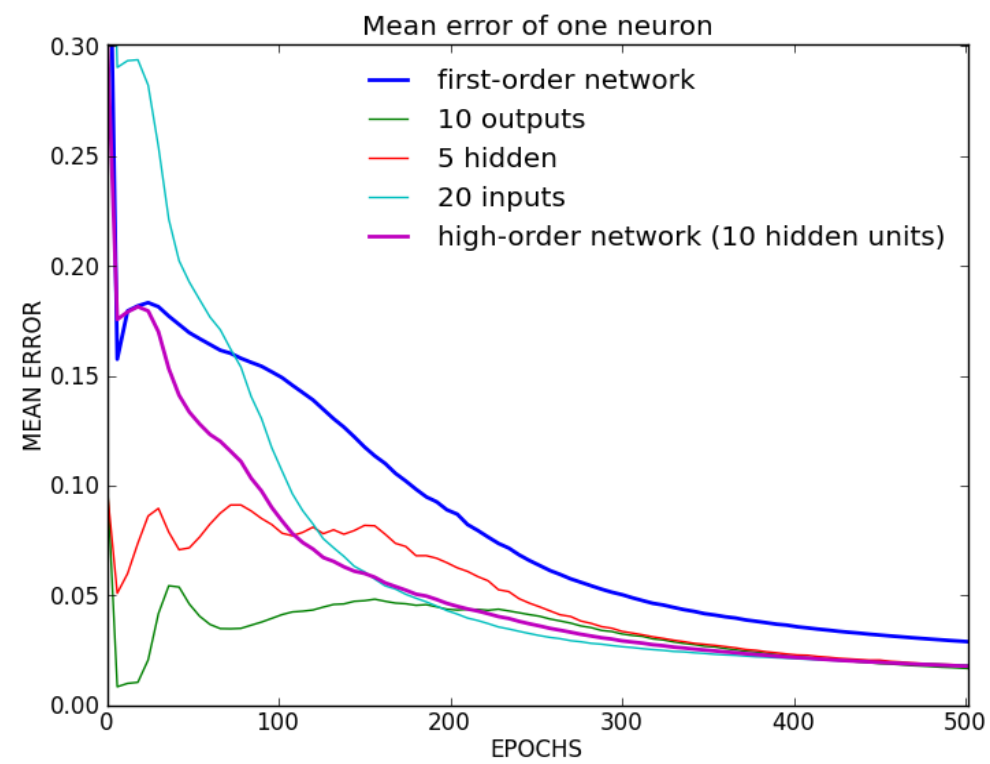
# Stabilisation de la couche cachée ( par discrétisation )



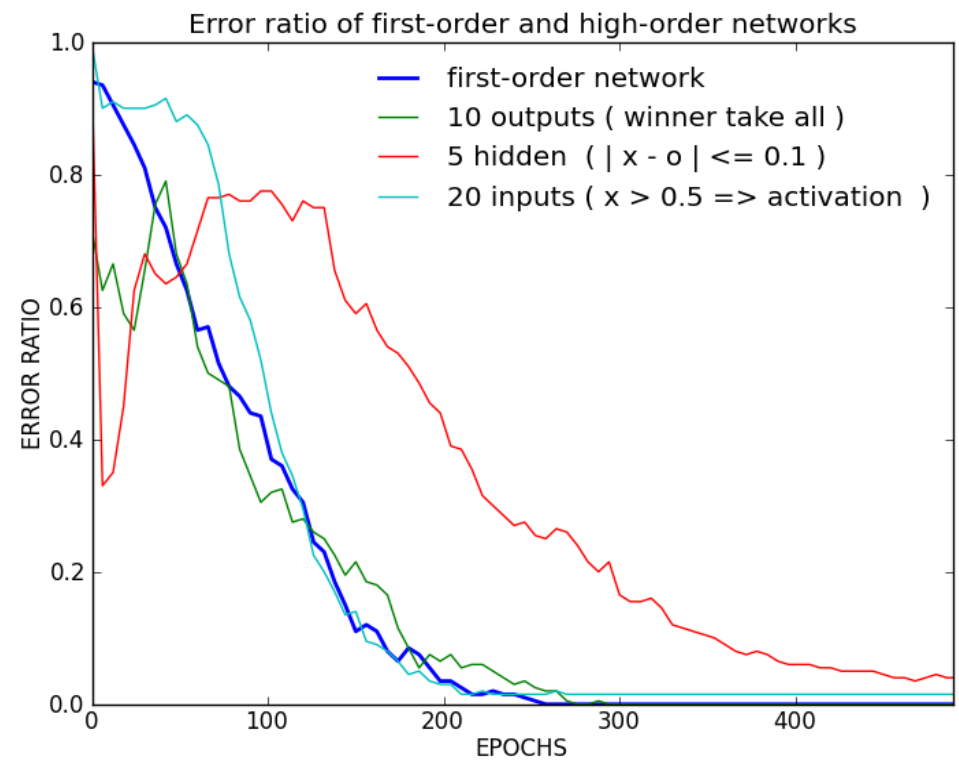
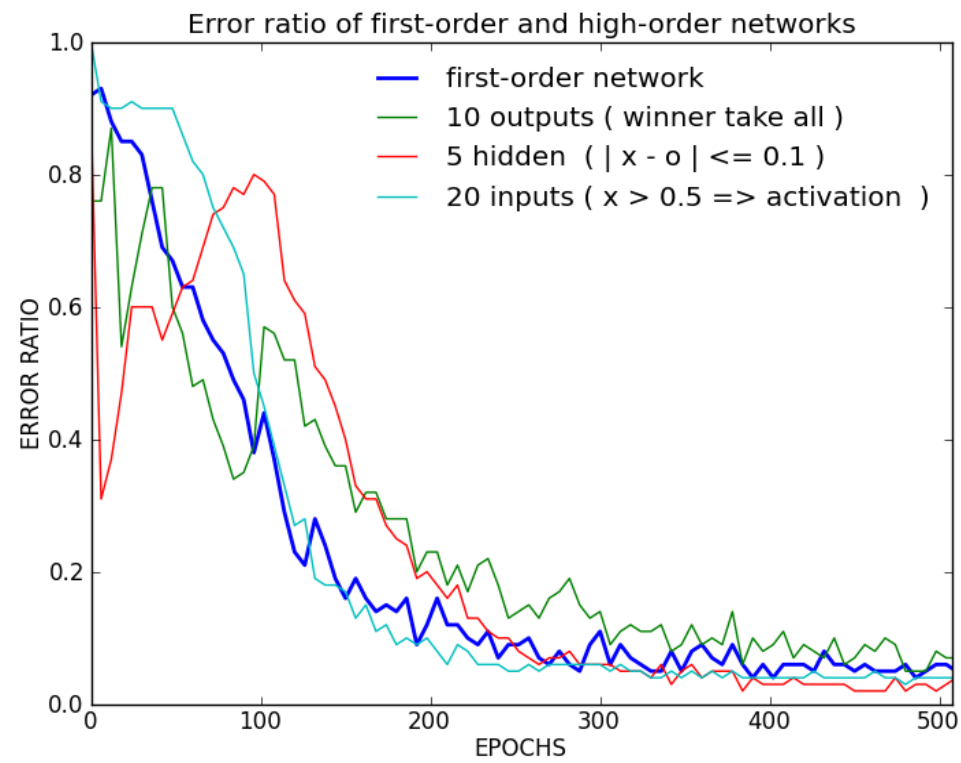


**Bloque apprentissage de la couche caché ( du premier réseau )**

Stop apprentissage à 100 epochs



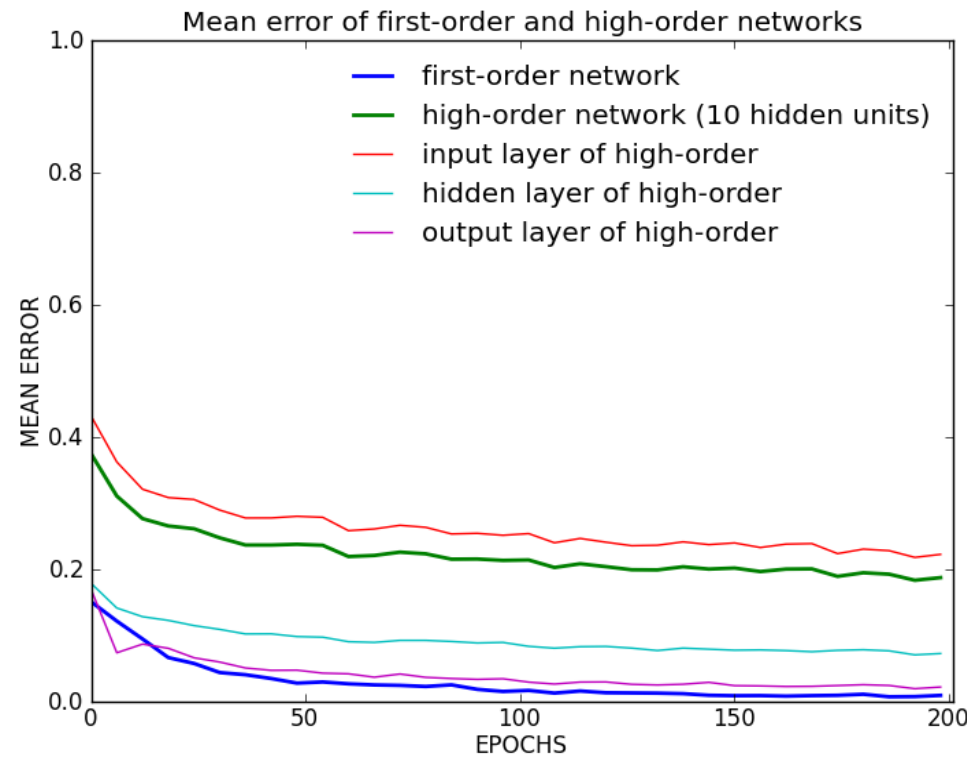
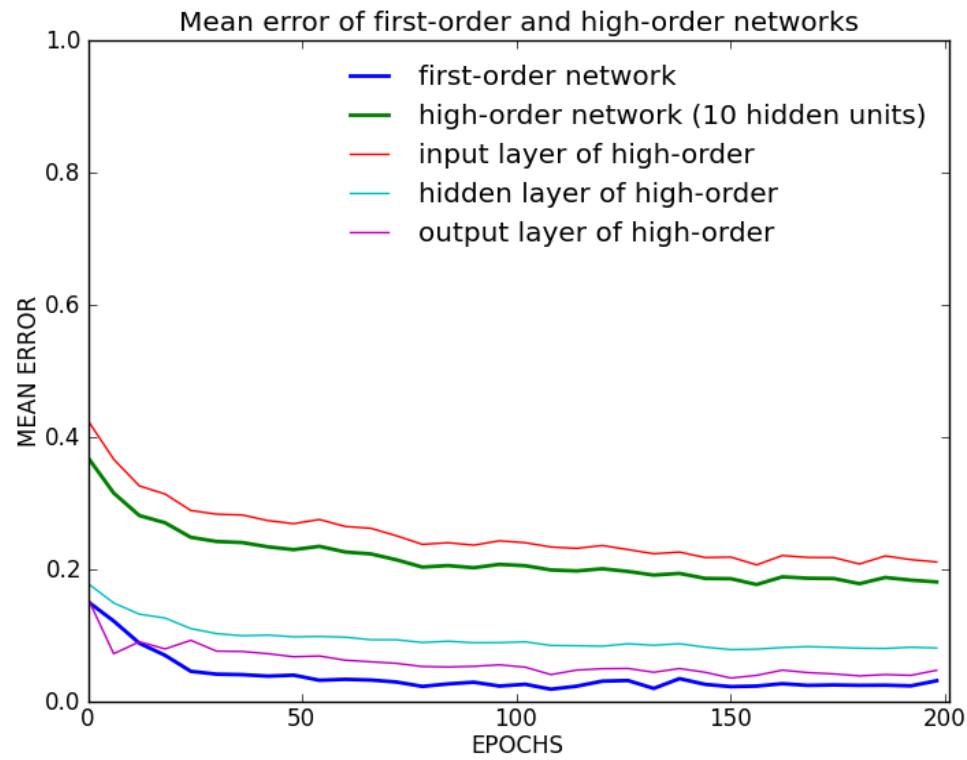
On constate que la seule courbe qui subi une perte de performance est celle de la reproduction des entrées alors que toutes les autres sont améliorées.



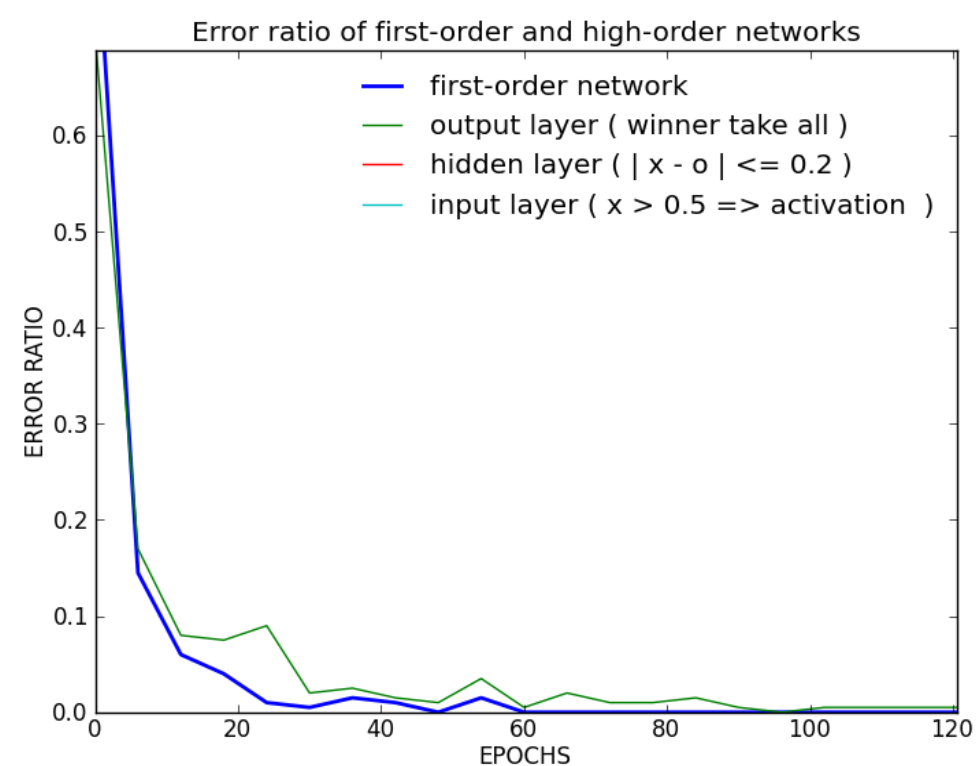
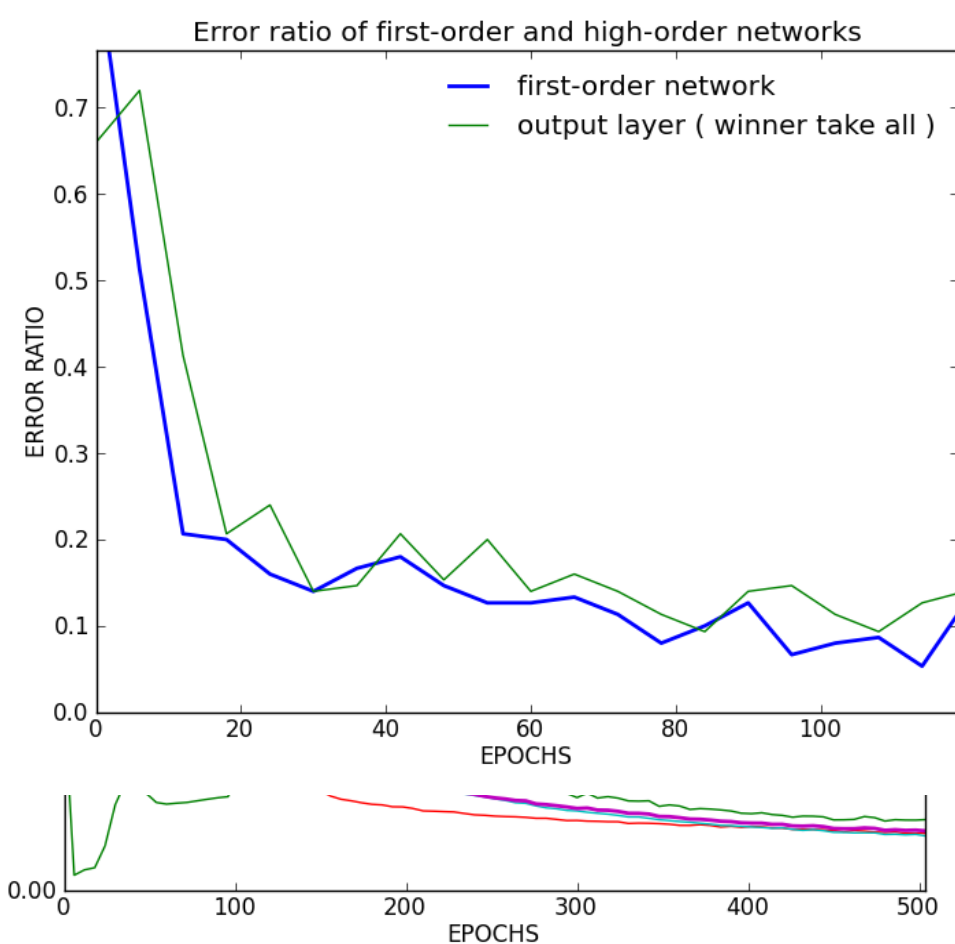
Au niveau des performances pures :

- la reconnaissance n'est pas parfaite ( courbe du premier réseau )
- l'apprentissage des neurones cachés est amélioré ( puisqu'ils ne changent plus )

**Sur des chiffres manuscrits ( blocage à l'epoch 20 )**



L'apprentissage de la couche de sortie est plus complexe.



Les performances pures sont détériorées

Sur des chiffres manuscrits ( blocage à l'epoch 20 ) avec réapprentissage d'une tâche différente ( 10 - x )

