

### Compte rendu de TP (reconnaissance de chiffres)

1) Les droites en pointillés sont le squelette de la surface discriminante. La surface discriminante s'articule autour de ces droites (tangentes). La fonction quadratique se rapproche d'une droite, elle est donc très peu adaptée à la répartition spiralée. Le réseau de neurones a une forme de Ying et Yang anguleux, ce discriminateur est moyennement adapté à la répartition spiralée. Le discriminateur avec ddp est une spirale en forme de Ying et Yang symétrique et centrée sur (0;0), il s'agit donc du discriminateur le plus adapté au problème. Le nombre de passes n'influence pas les performances du quadratique ni du RN (logique puisque les mauvaises performances sont liées à la répartition de la population en spirale), tandis que les performances du discriminateur ddp augmentent. Une augmentation du nombre de neurones dans la couche cachée n'influence pas le discriminateur quadratique (puisque les couches cachées n'interviennent pas dans le calcul du discriminateur quadratique) mais augmente les performances du RN qui tend à converger plus rapidement vers la valeur du discriminateur ddp (qui n'est lui-même pas influencé car ne fait pas appel aux couches cachées pour être calculé). Enfin, une augmentation du pas entraîne l'obtention d'un signal beaucoup plus bruité pour les trois discriminateurs. En effet, plus le pas est grand, moins les résultats seront précis. Quand on augmente la base d'apprentissage, on remarque qu'on tend à converger plus vite vers le taux de réussite maximum pour le RN. Le nombre de neurones influence donc les performances du RN. On observe parfois que le discriminateur RN sépare en plusieurs domaines circulaires isolés la population (alors qu'on pensait que la surface discriminante devait être continue). Dans ce cas, pourquoi ne peut-on pas avoir une infinité de domaines qui entourent chaque individu de la population (ce qui contribuerait à avoir un taux de réussite maximal selon nous) ?

2) Plus le nombre de chiffres augmente, plus le taux de réussite diminue, ce qui est logique puisqu'il y a de plus en plus de choix parmi des chiffres qui se ressemblent (par exemple 3 et 8 ou 1 et 7), ce qui génère des erreurs.

3) Lorsqu'on a 2 chiffres,  $\tau_{app}$  converge très vite pour atteindre 1 tandis qu'en augmentant le nombre de chiffres, la convergence est moins rapide, ce qui est logique puisqu'il y a plus de chiffres à apprendre à reconnaître.

4) S'il n'y a qu'un bruit gaussien, alors on n'observe pas de différence significative : les deux discriminateurs sont équivalents puisque leurs taux de réussite sont équivalents. S'il y a un bruit gaussien avec une amplitude inconnue en plus, alors le discriminateur FA gaussien amplitude inconnue est deux fois plus performant que le premier discriminateur. Le discriminateur FA gaussien amplitude inconnue est plus efficace pour des images plus nettes. Pourquoi le premier discriminateur, qui arrive à distinguer des images peu nettes, n'arrive pas à distinguer des images qui semblent plus nettes avec ajout d'amplitude ? Peut-on dire dans ce cas que « qui peut le plus ne peut pas le moins » ?

5) Le RN n'est pas performant car équivalent au discriminateur gaussien SANS amplitude alors qu'on fait le test sur une base de généralisation gaussienne AVEC amplitude. Dans le cas où les deux bases sont générées selon le modèle gaussien avec amplitude, alors le RN arrive à reconnaître les chiffres avec un très bon taux de réussite, cela est dû au fait que l'apprentissage se fait dans les mêmes conditions que la généralisation. Le même réseau neuronal a servi à faire à la fois l'apprentissage et la généralisation alors qu'avant, nous n'avions pas les mêmes conditions car le modèle gaussien SANS amplitude avait servi pour l'apprentissage et le modèle gaussien AVEC amplitude avait servi pour la généralisation. Lorsqu'on augmente le nombre de chiffres, on observe une légère perte de performances, cette baisse est moindre comparée aux expériences précédentes car cette fois-ci, l'apprentissage s'est fait dans des conditions plus strictes. Pour améliorer les performances d'un réseau de neurones, on peut augmenter le nombre de passes mais l'amélioration est très lente (convergence). Sinon, on peut augmenter le nombre de faces cachées de neurones.

Pour une répartition spiralée, le discriminateur le plus performant est le discriminateur avec ddp connu, mais il implique de connaître la densité de probabilité, ce qui n'est pas toujours le cas dans la réalité. Ainsi, l'approximation du réseau de neurones avec apprentissage adéquat (ici, bruit gaussien avec amplitude) nous permet de minimiser l'erreur sans connaître la densité de probabilité. De plus, le réseau de neurones est beaucoup plus performant que le discriminateur quadratique dans ce genre de cas où la population n'est pas linéairement (ou quadratiquement) séparable.