

CentraleSupélec
Projet long du cursus Supélec
Encadré par : J. Tomasik et A. Rimmel

Dessine-moi un mouton

Generative Adversarial Network

François Bouvier d'Yvoire
Matthieu Delmas
Romain Poirot
Paul Witz

Etude des réseaux de neurones en perceptrons
avec application au concept des Generative Adversarial Network

Années 2017-2018

Dessine-moi un mouton

Generative Adversarial Network

Résumé

Résumé

Mots-clefs

Mots-clefs

Table des matières

1	Introduction aux réseaux de neurones et premières applications	1
1.1	Outils utilisés pour le projet	1
1.2	Réseaux de neurones et fonctionnement	1
1.2.1	Le neurone	1
1.2.2	Réseau de neurones et perceptron	2
1.2.3	Apprentissage par rétro-propagation	3
1.3	Application au problème du XOR	4
1.4	Application à la base de données MNIST	5
2	Generative Adversarial Networks	6
2.1	Principe	6
2.2	Apprentissage	7
2.3	Résultats	7
2.4	Améliorations	7

Chapitre 1

Introduction aux réseaux de neurones et premières applications

1.1 Outils utilisés pour le projet

Description des outils mis en place pour le projet et de certains choix techniques.

1.2 Réseaux de neurones et fonctionnement

1.2.1 Le neurone

L'unité de base du réseau est le neurone, on peut l'imaginer comme une fonction mathématique. Lui sont attribuées n entrées, chacune affectée d'un poids w_i et une fonction mathématique de \mathbb{R} dans \mathbb{R} . Le rôle du neurone sera de renvoyer le résultat de la fonction, appliquée à la somme pondérée par le poids des entrées. On pourrait ajouter un biais comme paramètre de notre neurone afin d'ajuster notre résultat (choisir quand une fonction seuil renvoie 1 par exemple).

Un exemple simple du réseau de neurones est la séparation d'un plan en deux.

Imaginons un neurone à deux entrées e_1 et e_2 , chacune attribuée d'un poids w_1 et w_2 . On affecte au neurone un biais b et une fonction d'activation seuil (Heavyside par exemple).

Notre neurone renverra : 1 si $(e_1 * w_1 + e_2 * w_2) - b > 0$ et 0 sinon.

Si les e_1 et e_2 représentent les abscisses et ordonnées d'un point du plan, on reconnaît dans l'argument de la fonction d'activation l'équation d'une droite affine. Notre neurone pourra donc distinguer les points du plan selon le côté de la droite où ils se trouvent.

On peut déjà voir qu'une modification des poids entraînera une différente délimitation du

plan. On peut donc imaginer faire « apprendre » au réseau quels points délimiter en modifiant ses poids. Nous reviendrons sur ce concept par la suite.

Cependant, les applications d'un neurone seul sont vite limitées. C'est pourquoi on va s'intéresser à en connecter plusieurs entre eux.

1.2.2 Réseau de neurones et perceptron

On a déjà vu que le neurone se prêtait bien à une séparation binaire des données. On va voir que l'organisation de neurones en réseau permet de meilleures classifications.

L'organisation du réseau se fera au moyen de couches de neurones. Dans les structures les plus basiques, la sortie d'une couche est utilisée comme l'entrée de la couche suivante. On peut imaginer des réseaux plus complexes où la sortie n'est pas réutilisée dans la couche suivante mais plusieurs couches plus loin ou dans des couches antérieures. On appelle la dernière couche, celle qui donne le résultat du réseau de neurones, la couche de sortie ; et la première couche où l'on donne les entrées est appelée couche d'entrée. Les autres couches sont appelées des couches cachées. Cette appellation vient du fait qu'a priori, nous n'avons aucun moyen de voir ou de corriger les comportements des neurones cachés. En effet, avec la seule donnée de la sortie, l'influence des poids des couches cachées sur celle-ci n'est pas évidente.

Le perceptron est un modèle de réseau de neurones auquel on va s'intéresser particulièrement. Il s'agit d'un réseau linéaire où chaque couche est entièrement connectée à la suivante, c'est-à-dire que chaque neurone d'une couche prend en entrée toutes les sorties de la couche précédente. On ne trouve aucune boucle dans le graphe d'un perceptron, c'est donc une propagation vers l'avant. L'utilité d'avoir plusieurs couches se comprend facilement. Si on reprend notre exemple du problème de classification des points, on peut imaginer, par exemple, quatre neurones qui enverront leur sortie sur un neurone à quatre entrées. Chacun des neurones réalisera la séparation du plan en deux selon le principe déjà évoqué précédemment. Le neurone de la couche de sortie pourra réaliser facilement le rôle d'un ET logique. On vient de sélectionner un carré dans le plan. En étendant le raisonnement, on voit qu'un réseau à deux couches permet de sélectionner n'importe quelle zone convexe de l'espace des entrées (ici du plan). De même, un réseau à trois couches pourra sélectionner n'importe quelle zone concave de l'espace des entrées.

1.2.3 Apprentissage par rétro-propagation

Il existe plusieurs types d'apprentissages du réseau. Les deux grandes catégories sont l'apprentissage supervisé et l'apprentissage non supervisé. Dans le cadre du perceptron, nous utilisons seulement un apprentissage supervisé, la rétro-propagation des erreurs. Un apprentissage supervisé nécessite une base d'apprentissage à enseigner au réseau. Elle est composée d'associations entre entrées et sorties voulues. Le réseau déduira de cette base les autres cas qu'on ne lui aura pas appris.

La rétro-propagation consiste à calculer l'influence de chaque paramètre sur la sortie et à les mettre à jour en fonction de cette influence.

Les paramètres que l'on fait évoluer sont les poids et les biais. La formule de mise à jour est la suivante :

$$W(t+1) = W(t) + \eta \frac{\partial E}{\partial W}$$

avec η le pas de convergence, $\frac{\partial E}{\partial W}$ la matrice de terme général $\frac{\partial E}{\partial W_{i,j}}$.
Pour pouvoir mettre à jour les poids, il faut donc calculer les $\frac{\partial E}{\partial W_{i,j}}$.

A la couche k l'influence des poids est donnée par :

$$\frac{\partial E^p}{\partial W_k} = \frac{\partial F}{\partial W}(W_k, X_{k-1}) \frac{\partial E^p}{\partial X_k}$$

Avec $\frac{\partial F}{\partial W}(W_k, X_{k-1})$ la matrice jacobienne de F par rapport à la variable W_k

Pour pouvoir calculer l'influence des poids de toutes les couches, il faut donc calculer $\frac{\partial E^p}{\partial W_k}$

On peut calculer par récurrence cette valeur pour toutes les couches.

$$\frac{\partial E^p}{\partial X_{k-1}} = \frac{\partial F}{\partial X}(W_k, X_{k-1}) \frac{\partial E^p}{\partial X_k}$$

Avec $\frac{\partial F}{\partial X}(W_k, X_{k-1})$ la matrice jacobienne de F par rapport à la variable X_k . De plus, dans un perceptron on peut noter la sortie de la couche k :

$$Y_k = W_k X_k$$

$$X_k = F(Y_k)$$

On obtient donc ces 3 équations :

$$\begin{aligned}\frac{\partial E^p}{\partial y_k^i} &= f'(x_k^i) \frac{\partial E^p}{\partial x_k^i} \\ \frac{\partial E^p}{\partial w_k^{i,j}} &= x_{k-1}^j \frac{\partial E^p}{\partial y_k^i} \\ \frac{\partial E^p}{\partial x_{k-1}^m} &= \sum_i (w_k^{im} \frac{\partial E^p}{\partial x_k^i})\end{aligned}$$

En forme matricielle, ces équations donnent :

$$\begin{aligned}\frac{\partial E^p}{\partial Y_k} &= \text{Diag}(f'(x_k^i)) \frac{\partial E^p}{\partial x_k^i} \\ \frac{\partial E^p}{\partial W_k} &= X_{k-1}^T \frac{\partial E^p}{\partial Y_k} \\ \frac{\partial E^p}{\partial X_{k-1}} &= W_k^T \frac{\partial E^p}{\partial X_k}\end{aligned}$$

1.3 Application au problème du XOR

Lorsque l'on souhaite travailler sur des algorithmes d'apprentissage par ordinateur, il est recommandé de les essayer sur des problèmes connus afin d'en vérifier les performances.

Le problème du XOR est l'un des plus classiques car il apporte de nombreuses difficultés.

L'objectif du XOR est de séparer le plan complexe en quatre cadrants, $(x > 0, y > 0)$, $(x > 0, y < 0)$, $(x < 0, y > 0)$ et $(x < 0, y < 0)$. Pour l'expérimentation, on restreint le plan à $[-1; 1]^2$. Les sorties attendues par le réseau de neurones sont alors 1 pour les points tel que $x * y > 0$ et -1 pour les points tels que $x * y < 0$.

Le premier intérêt de ce problème est qu'il est non linéaire. Cela se traduit par le fait qu'une droite séparant le plan en 2 ne répond pas du tout au problème.

C'est en se basant sur la résolution du XOR que nous avons construit notre structure de réseau et vérifié la cohérence de notre code. La littérature propose comme réseau le plus simple pour ce problème une couche cachée de 2 neurones, avec 2 entrées (x et y) et 1 sortie dans $[-1, 1]$. Nous avons étudié également quelques autres formes de réseaux pour comparer les résultats.

Notion de résultats La notion de résultats nécessite d'être correctement définie afin de pouvoir être interprétée correctement, en particulier pour la comparaison à d'autres résultats obtenus par nous-même ou par d'autres personnes.

La structure de perceptron sous cette forme classe les objets que l'on donne en entrée. Généralement, le résultat est défini par rapport à un pourcentage de succès dans cette classification. Pour

l'obtenir, on commence par définir une erreur relative, c'est-à-dire une distance entre la sortie cible et la sortie obtenue. Un seuil est alors appliqué afin de définir une sortie booléenne de la classification de l'entrée.

Dans le cas du XOR on met en place un seuil de 0.5, c'est à dire que, si l'erreur est inférieure à 50%, le réseau a raison. Cela peut s'interpréter comme suit : le réseau donne un résultat qui indique sa confiance dans la sortie. 1 ou 0 si il est certain que la sortie doit être 1 ou 0, 0.5 si il ne peut départager l'un ou l'autre, le seuil consiste à dire que sa réponse est celle en qui il a le plus confiance. On cherche également à évaluer la vitesse d'apprentissage. Ainsi, on calcule le pourcentage de succès du réseau à intervalles réguliers au cours de l'apprentissage. Les réseaux étant soumis à une forte composante aléatoire (l'ordre d'apprentissage, ainsi que l'initialisation des poids), on effectue des apprentissages dans les mêmes conditions plusieurs fois afin d'obtenir des courbes moyennes, et des intervalles de confiances justifiant nos résultats.

Réseau en $2 \rightarrow 2 \rightarrow 1$ Les résultats obtenus au début sur ce réseau extrêmement simple semblaient tout à fait aléatoires et nous ont permis de détecter des erreurs de traduction des équations de rétro-propagation en code Python. Nous avons finalement pu obtenir des résultats satisfaisants, comme le montre la figure ???. Cependant, ce résultat n'était pas obtenu dans l'intégralité des apprentissages, nous fournissant des résultats très différents, comme sur la figure ??. La littérature, et en particulier les rapports des années précédentes [1] et [2], nous ont montré que le XOR n'était effectivement pas juste dans 100% des cas.

Nous avons donc soumis le réseau à de nombreux apprentissages, en faisant varier les paramètres ainsi que la forme du réseau. Voici les résultats les plus intéressants :

Conclusion sur le XOR Instabilité du réseau $2 \rightarrow 2 \rightarrow 1$ et comparaison avec le $2 \rightarrow 4 \rightarrow 1$ et le $2 \rightarrow 2 \rightarrow 2 \rightarrow 1$

Pas d'apprentissage très petit par rapport à la littérature

Influence des fonctions d'activation

1.4 Application à la base de données MNIST

Chapitre 2

Generative Adversarial Networks

La première partie de cette étude nous a permis de maîtriser l'utilisation de réseaux en perceptron et de structurer une architecture logicielle efficace et souple pour l'étude des GAN.

Après avoir obtenu des résultats satisfaisants sur la reconnaissance de motif sur la base MNIST, nous étudions la génération de données à l'aide de réseaux de neurones en nous basant sur le concept de GAN, introduit par []

2.1 Principe

Le but du GAN est de générer des images que l'on ne pourra distinguer de « vraies » images, prises avec un appareil photo. Un GAN est constitué de deux réseaux de neurones : le Générateur (G) et le Discriminateur (D). Le Générateur a pour but de créer les images et le Discriminateur de déterminer si les images qu'on lui donne sont de « vraies » images ou ont été créées par le Générateur. Ces deux réseaux sont mis en compétition : le Générateur a pour but de tromper le Discriminateur tandis que le Discriminateur doit discerner les « fausses » images.

L'apprentissage du Discriminateur se fait à la fois sur des images générées par le Générateur et de « vraies » images, issues d'une banque d'images.

L'apprentissage du Générateur dépend aussi du Discriminateur : lorsqu'il génère une image, on la donne au Discriminateur pour voir si le Générateur a réussi à le tromper. L'apprentissage du Générateur se fait sur la réponse du Discriminateur.

De façon plus formelle, on travaille avec 3 distributions : p_x , la distribution idéale des vraies images, p_{data} , l'échantillon des vraies images et p_{model} la distribution réalisée par les images issues du Générateur. Le but de l'apprentissage est de rapprocher p_{model} de p_x . Comme p_x nous est

inconnu, on va plutôt s'approcher de p_{data} .

Notre algorithme d'apprentissage repose une fois encore sur la descente de gradient et la rétro-propagation des erreurs. Cependant, les fonctions d'erreur sont différentes. En effet, nous ne connaissons pas la distribution exacte et ne pouvons donc pas minimiser une « distance minimale » entre le résultat obtenu et le résultat désiré.

2.2 Apprentissage

L'apprentissage se déroule ainsi :

Nous fournissons au Discriminateur des images x (de la BDD MNIST par exemple) et lui demandons de nous renvoyer un réel entre 0 et 1, qui représente son degré de confiance sur le fait que l'image fournie ait été tirée d'une banque de donnée authentique ($D(x) = 1$) ou du Générateur ($D(x) = 0$).

Le Générateur, quant à lui, devra générer une image à partir d'un vecteur de bruit z . Cette image est ensuite jugée par le Discriminateur : $D(G(z)) = 1$ si le Générateur a dupé le Discriminateur et 0 sinon. Puis, on effectue une rétropropagation à travers les deux réseaux pour les faire apprendre.

Les réseaux utilisés sont ceux que nous avons utilisé auparavant : des perceptrons multicouches. Le Discriminateur, en particulier, reprend la même structure qu'un perceptron utilisé pour reconnaître des chiffres manuscrits.

2.3 Résultats

2.4 Améliorations

Bibliographie

[1] Rapport appartement.

[2] Rapport pinaple.