

1 Introduction

Ce court rapport est une simple analyse de ce qui a été présenté par Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht et Oriol Vinyals dans l'article : Understanding Deep Learning requires rethinking generalization [1].

2 Contributions

A travers plusieurs expérimentations, les auteurs montrent que les approches traditionnelles échouent dans l'explication du phénomène que les réseaux de neurones larges généralisent bien en pratique. Par ailleurs, les expérimentations prouvent que l'état de l'art des réseaux de neurones convolutionnels pour la classification d'images qui ont été entraînés avec des méthodes de descente de gradient stochastique arrivent à apprendre facilement un étiquetage aléatoire des exemples d'apprentissage.

Ce phénomène n'est pas affecté par la régularisation et apparaît même si on applique sur les images originales un bruit complètement aléatoire et non structuré.

De plus, les auteurs montrent que deux réseaux de neurones à simple profondeur arrivent à avoir une expressivité finie d'échantillons une fois le nombre de paramètres dépasse le nombre d'exemples, c'est ce qui se passe dans la pratique.

Les réseaux de neurones profonds ont généralement un nombre de paramètres plus grand

que le nombre d'exemples.

La contribution des auteurs est dans les tests qu'ils ont fait :

2.1 Tests de randomisation

Les auteurs ont entraîné plusieurs architectures de réseaux de neurones sur une copie des données où les étiquettes des classes ont été remplacées par des étiquettes aléatoires. L'erreur obtenue sur l'ensemble d'apprentissage était nulle, et l'erreur sur l'ensemble de test était un peu aléatoire vu qu'il n'y a pas de corrélation entre les étiquettes des deux ensembles. Les conclusions tirées de cette expérience sont :

- La capacité d'un réseau de neurone est suffisante pour mémoriser toutes les données,
- les réseaux de neurones profonds apprennent facilement les étiquettes aléatoires, et
- la randomisation des étiquettes est considérée comme une sorte de transformation appliquée sur les données sauf que tous les autres paramètres du problème restent les mêmes.

2.2 Le rôle de la régularisation explicite

Ici, les auteurs montrent que la régularisation explicite en utilisant par exemple le dropout et le weight-decay peut améliorer les performances de généralisation mais elle n'est pas nécessaire

ni suffisante pour contrôler l'erreur de généralisation. De plus, la régularisation dans les réseaux de neurones profonds joue beaucoup plus le rôle de l'optimisation de la dernière erreur sur l'ensemble de test du modèle.

2.3 Expressivité d'un échantillon fini

Il a été montré aussi que l'utilisation d'un réseau de neurones avec deux couches cachées et la fonction d'activation ReLU avec $p = 2n + d$ paramètres peut apprendre n'importe quel échantillon de taille n en d dimensions.

2.4 Rôle d'une régularisation implicite

Un modèle qui peut bien apprendre les données d'apprentissage n'est pas forcément capable de bien généraliser. Les auteurs montrent que même les méthodes basées sur la fonction gaussienne noyau peuvent bien généraliser sans régularisation sur les petits ensembles. Toutefois, cela n'explique pas pourquoi certaines architectures généralisent mieux que d'autres.

3 Travaux connexes

Plusieurs travaux ont été fait dans le passé pour étudier l'expressivité d'un réseau de neurones et parmi les conclusions marquantes on cite :

- La stabilité d'un algorithme d'apprentissage est indépendante de l'étiquetage de l'ensemble d'apprentissage.
- Même les perceptrons de deux couches cachées seulement ont une expressivité universelle sur les échantillons de taille finie.

- La taille d'un réseau de neurone n'est pas l'acteur principal de son contrôle de capacité.

4 Expérimentations

Tous les tests ont été faits sur les deux bases de données *CIFAR_10* et *ImageNet*.

5 Conclusion

Les expérimentations présentées dans cet article mettent l'accent sur l'importance de comprendre la notion de capacité effective des modèles de réseaux de neurones, qui sont facilement capable de mémoriser tout l'ensemble d'apprentissage. Pour résumer, les raisons pour lesquelles l'optimisation est empiriquement facile doivent être différentes de la vraie cause de généralisation. C'est pour cela qu'il est nécessaire de comprendre les vraies notions qui se cachent derrière la généralisation pour pouvoir bien comprendre l'apprentissage profond.

Références

- [1] Zhang, C., Bengio, S., Hardt, M., Recht, B. and Vinyals, O., 2016. Understanding deep learning requires rethinking generalization. arXiv preprint arXiv :1611.03530.