

Introduction au Deep Learning

Chez le même éditeur

Introduction au Machine Learning

Chloé-Agathe Azencott

240 pages

Dunod, 2019

Big Data et Machine Learning

3^e édition

Pirmin Lemberger, Marc Batty, Médéric Morel, Jean-Luc Raffaëlli

272 pages

Dunod, 2019

Machine Learning avec Scikit-Learn

2^e édition

Aurélien Géron

320 pages

Dunod, 2019

Deep Learning avec Keras et TensorFlow

2^e édition

Aurélien Géron

576 pages

Dunod, 2020

Introduction au **Deep** **Learning**

Eugene Charniak

Professeur d'informatique et de sciences cognitives
à l'Université Brown

Traduit de l'anglais par **Anne Bohy**

DUNOD

Traduction de l'ouvrage d'Eugene Charniak :
Introduction to Deep Learning
Copyright © 2018, The Massachusetts Institute of Technology

Illustration de couverture : © kras99 - Adobe Stock

© Dunod, 2021
11 rue Paul Bert, 92240 Malakoff
www.dunod.com
ISBN 978-2-10-082578-3

Le Code de la propriété intellectuelle n'autorisant, aux termes de l'article L. 122-5, 2° et 3° a), d'une part, que les « copies ou reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective » et, d'autre part, que les analyses et les courtes citations dans un but d'exemple et d'illustration, « toute représentation ou reproduction intégrale ou partielle faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause est illicite » (art. L. 122-4).

Cette représentation ou reproduction, par quelque procédé que ce soit, constituerait donc une contrefaçon sanctionnée par les articles L. 335-2 et suivants du Code de la propriété intellectuelle.

TABLE DES MATIÈRES

AVANT-PROPOS	ix
CHAPITRE 1 • RÉSEAUX DE NEURONES À PROPAGATION AVANT	1
1.1 Les perceptrons	3
1.2 Fonctions de perte d'entropie croisée pour les réseaux de neurones	8
1.3 Dérivées et descente de gradient stochastique	13
1.4 Écriture de notre programme	17
1.5 Représentation matricielle des réseaux de neurones	20
1.6 Indépendance des données	23
1.7 Références et lectures supplémentaires	24
Exercices	25
CHAPITRE 2 • TENSORFLOW	27
2.1 Introduction à TensorFlow	27
2.2 Un programme TF	30
2.3 Réseaux de neurones à plusieurs couches	36
2.4 Autres particularités	39
2.5 Références et lectures supplémentaires	45
Exercices	45
CHAPITRE 3 • RÉSEAUX DE NEURONES CONVOLUTIFS	47
3.1 Filtres, pas et marge	47
3.2 Exemple simple de convolution en TF	53
3.3 Convolution à plusieurs niveaux	56
3.4 Présentation détaillée de la convolution	59
3.5 Références et lectures supplémentaires	61
Exercices	63
CHAPITRE 4 • PLONGEMENTS DE MOTS ET RÉSEAUX DE NEURONES RÉCURRENTS	65
4.1 Plongement de mots pour modèles linguistiques	65
4.2 Construction de modèles linguistiques à propagation avant	70

Table des matières

- 4.3 Amélioration des modèles linguistiques à propagation avant 72
 - 4.4 Surajustement 73
 - 4.5 Réseaux récurrents 76
 - 4.6 Longue mémoire à court terme 82
 - 4.7 Références et lectures supplémentaires 85
 - Exercices 86
- CHAPITRE 5 • **APPRENTISSAGE SÉQUENCE À SÉQUENCE** **89**
 - 5.1 Principe du seq2seq 90
 - 5.2 Écriture d'un programme de traduction automatique seq2seq 92
 - 5.3 L'attention dans seq2seq 95
 - 5.4 Seq2Seq multi-longueurs 99
 - 5.5 Exemples de programmation 100
 - 5.6 Références et lectures supplémentaires 103
 - Exercices 104
- CHAPITRE 6 • **APPRENTISSAGE PAR RENFORCEMENT PROFOND** . **105**
 - 6.1 Itération sur les valeurs 106
 - 6.2 Apprentissage sans modèle 109
 - 6.3 Les bases de l'apprentissage par renforcement profond 111
 - 6.4 Méthodes de gradients de politique 115
 - 6.5 Méthodes acteur-critique 121
 - 6.6 Répétition d'expériences 124
 - 6.7 Références et lectures supplémentaires 125
 - Exercices 126
- CHAPITRE 7 • **MODÈLES DE RÉSEaux DE NEURONES NON SUPERVISÉS** **127**
 - 7.1 Encodage de base 127
 - 7.2 Auto-encodage convolutif 130
 - 7.3 Auto-encodage variationnel 134
 - 7.4 Réseaux antagonistes génératifs 142
 - 7.5 Références et lectures supplémentaires 146
 - Exercices 147

ANNEXE : CORRIGÉ DES EXERCICES	149
A.1 Chapitre 1	149
A.2 Chapitre 2	149
A.3 Chapitre 3	150
A.4 Chapitre 4	151
A.5 Chapitre 5	151
A.6 Chapitre 6	152
A.7 Chapitre 7	152
BIBLIOGRAPHIE.....	153
INDEX.....	157

AVANT-PROPOS

Votre auteur, de longue date chercheur en intelligence artificielle, a vu son domaine d'expertise, le traitement du langage naturel, révolutionné par le Deep Learning. Malheureusement, il lui a fallu (il m'a fallu) longtemps pour admettre ce fait. Je peux le justifier en disant que c'est la troisième fois que les réseaux de neurones menacent de tout révolutionner, mais seulement la première fois qu'ils y parviennent. Quoi qu'il en soit, je me suis retrouvé soudainement à la traîne et peinant à rattraper le temps perdu. J'ai donc fait ce que ferait tout professeur qui se respecte : j'ai inclus cette matière dans mon planning d'enseignement, j'ai démarré une formation accélérée en surfant sur le Web, et finalement ce sont mes étudiants qui me l'ont enseignée. (Cette dernière affirmation n'est pas une boutade. Il me faut saluer tout particulièrement l'aide de Siddarth (Sidd) Karamcheti, assistant d'enseignement principal en premier cycle.)

Ceci explique plusieurs caractéristiques essentielles de ce livre. En premier lieu, il est court. J'apprends lentement. Ensuite, il est très orienté projet. De nombreux écrits, tout particulièrement en informatique, sont en perpétuelle tension entre l'organisation propre au sujet et l'organisation des données dans le cadre de projets spécifiques. Séparer les deux est souvent une bonne idée, mais je pense que j'apprends mieux l'informatique en m'asseyant et en écrivant des programmes, c'est pourquoi mon livre reflète largement mon mode d'apprentissage. C'était la façon la plus pratique de tout formaliser, et j'espère que nombreux seront les futurs lecteurs qui trouveront également cela utile.

Ce qui soulève la question de mes futurs lecteurs. Si j'espère que de nombreux adeptes de la programmation trouveront ce livre utile pour la raison même qui m'a poussé à l'écrire, en tant qu'enseignant mes étudiants sont ma première préoccupation, c'est pourquoi ce livre est conçu avant tout comme un support de cours sur le Deep Learning.

Mon enseignement à l'université de Brown s'adresse aux étudiants de premier et deuxième cycles et couvre l'ensemble des sujets développés ici, auxquels s'ajoutent quelques conférences de « culture générale ». Pour obtenir son diplôme, un étudiant doit aussi mener à bien un projet d'une certaine importance. Des connaissances en algèbre linéaire et en analyse multivariée sont requises. Bien que les notions d'algèbre linéaire utilisées ici soient relativement simples, les étudiants m'ont déclaré que sans ces bases ils auraient eu du mal à concevoir des réseaux à plusieurs niveaux et que les tenseurs qu'ils requièrent leur auraient paru difficiles. L'analyse multivariée, cependant, a constitué une plus grande difficulté. Elle n'apparaît explicitement qu'au chapitre 1, lorsque nous construisons entièrement le réseau jusqu'à la rétropropagation et je ne serais pas surpris qu'un complément sur les dérivées partielles s'avère nécessaire. Enfin, il y a un prérequis en probabilités et statistiques. Ceci aide à la

Avant-propos

bonne compréhension de l'exposé et je tiens à encourager les étudiants à suivre un tel cours. Je suppose également que le lecteur a une connaissance rudimentaire de Python. Je n'inclus pas ce sujet dans mon livre, mais mon cours comporte un TD supplémentaire sur les bases de ce langage.

Le fait que votre auteur était en train d'effectuer une formation de rattrapage tout en écrivant ce livre explique également que dans les lectures supplémentaires proposées à la fin de chaque chapitre on puisse trouver, au-delà des références habituelles aux travaux de recherche les plus marquants, de nombreuses références à des sources secondaires — les écrits à caractère pédagogique d'autres personnes. Je n'aurais jamais réussi à apprendre toutes ces choses sans eux.

Providence, Rhode Island, USA
Janvier 2018